

INTERNATIONAL CONFERENCE ON ADVANCED TECHNOLOGIES, COMPUTER ENGINEERING AND SCIENCE

Alanya, TURKEY

26-28
APRIL
2019



2019
PROCEEDING BOOK

ICATCES 2019 Proceeding Book

International Conference on Advanced Technologies,
Computer Engineering and Science

26-28 Apr 2019 / Alanya, Turkey

Proceeding Book of the International Conference on Advanced Technologies,
Computer Engineering and Science (ICATCES 2019)

Editors

Assoc. Prof. Dr. Oguz FINDIK
Asst. Prof. Dr. Emrullah SONUÇ
Res. Asst. Yusuf Yargı BAYDİLLİ

Published, 2019.

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned. Nothing from this publication may be translated, reproduced, stored in a computerized system or published in any form or in any manner.

<http://icatces.org/>
icatces@karabuk.edu.tr

The individual contributions in this publication and any liabilities arising from them remain the responsibility of the authors.

The publisher is not responsible for possible damages, which could be a result of content derived from this publication.

Honorary Committee

Prof. Dr. Refik Polat, Karabük University, Rector
Prof. Dr. Mehmet AKBABA, Karabük University

Chair

Assoc. Prof. Dr. İlker TÜRKER, Karabük University
Assoc. Prof. Dr. Oğuz FINDIK, Karabük University

Co-Chair

Asst. Prof. Dr. Emrullah SONUÇ, Karabük University

Organization Committee

Prof. Dr. İsmail Rakıp KARAŞ, Karabük University
Assoc. Prof. Dr. İlker TÜRKER, Karabük University
Assoc. Prof. Dr. Oğuz FINDIK, Karabük University
Asst. Prof. Dr. Caner ÖZCAN, Karabük University
Asst. Prof. Dr. Emrullah SONUÇ, Karabük University
Asst. Prof. Dr. Hakan KUTUCU, Karabük University
Res. Asst. Dr. Rafet DURGUT, Karabük University
Res. Asst. Sait DEMİR, Karabük University
Res. Asst. Yusuf Yargı BAYDILLI, Karabük University

Scientific Committee

Prof. Dr. Ali Karcı (İnönü University)
Prof. Dr. Basel Mahafzah (The University of Jordan)
Prof. Dr. Cemil ÖZ (Sakarya Üniversitesi)
Prof. Dr. Derviş KARABOĞA (Erciyes University)
Prof. Dr. Erkan ÜLKER (Konya Technical University)
Prof. Dr. Fatih Vehbi ÇELEBİ (Yıldırım Beyazıt University)
Prof. Dr. Ghulam Ali Mallah (Shah Abdul Latif University)
Prof. Dr. Haldun GÖKTAŞ (Yıldırım Beyazıt University)
Prof. Dr. Halil İbrahim BÜLBÜL (Gazi University)
Prof. Dr. Harun UĞUZ (Konya Technical University)
Prof. Dr. İhsan ULUER (Karabük University)
Prof. Dr. İsmail Rakıp KARAŞ (Karabük University)
Prof. Dr. Mario KOEPPEN (Kyushu Institute Of Technology)
Prof. Dr. Mehmet AKBABA (Karabük University)
Prof. Dr. Mehmet ÖZALP (Karabük University)
Prof. Dr. Mykola S. Nikitchenko (Taras Shevchenko National University of Kyiv)
Prof. Dr. Nurhan KARABOĞA (Erciyes University)
Prof. Dr. Oleksandr I. Provotar (Taras Shevchenko National University of Kyiv)
Prof. Dr. Oleksandr O. Marchenko (Taras Shevchenko National University of Kyiv)
Prof. Dr. Raif BAYIR (Karabük University)
Prof. Dr. Sergiy D. Pogorilyy (Taras Shevchenko National University of Kyiv)
Prof. Dr. Serhii L. Kryvyi (Taras Shevchenko National University of Kyiv)
Prof. Dr. Vasyl M. Tereshchenko (Taras Shevchenko National University of Kyiv)
Prof. Dr. Valentina Emilia BALAS (University Aurel Vlaicu)
Prof. Dr. Yaşar BECERİKLİ (Kocaeli University)
Assoc. Prof. Dr. Abdrakhmanov RUSTAM (Ahmet Yesevi University)
Assoc. Prof. Dr. Adib HABBAL (Karabük University)
Assoc. Prof. Dr. Amirtayev KANAT (Ahmet Yesevi University)
Assoc. Prof. Dr. Bilal Alataş (Fırat University)
Assoc. Prof. Dr. Ergin YILMAZ (Bülent Ecevit University)
Assoc. Prof. Dr. Ivan Izonin (Lviv Polytechnic National University)
Assoc. Prof. Dr. İlhami Muharrem ORAK (Karabük University)
Assoc. Prof. Dr. İlker TÜRKER (Karabük University)

Assoc. Prof. Dr. İsmail BABAOĞLU (Konya Technical University)
Assoc. Prof. Dr. Kemal POLAT (Abant İzzet Baysal University)
Assoc. Prof. Dr. Mesut GÜNDÜZ (Konya Technical University)
Assoc. Prof. Dr. Mustafa Servet KIRAN (Konya Technical University)
Assoc. Prof. Dr. Necaattin Barışçı (Gazi University)
Assoc. Prof. Dr. Oleksii I. Chentsov (Taras Shevchenko National University of Kyiv)
Assoc. Prof. Dr. Rabie A. RAMADAN (Cairo University)
Assoc. Prof. Dr. Taras V. Panchenko (Taras Shevchenko National University of Kyiv)
Assoc. Prof. Dr. Tulep ABDIMUHAN (Ahmet Yesevi University)
Assoc. Prof. Dr. Yuliya Kozina (Odessa National Polytechnic University)
Asst. Prof. Dr. Ahmet BABALIK (Konya Technical University)
Asst. Prof. Dr. Berk Anbaroğlu (Hacettepe University)
Asst. Prof. Dr. Bilal Babayiğit (Erciyes University)
Asst. Prof. Dr. Burhan Konya Technical (Karabük University)
Asst. Prof. Dr. Caner ÖZCAN (Karabük University)
Asst. Prof. Dr. Ebubekir YAŞAR (Gaziosmanpaşa University)
Asst. Prof. Dr. Erkan DUMAN (Fırat University)
Asst. Prof. Dr. Ferhat ATASOY (Karabük University)
Asst. Prof. Dr. Hakkı SOY (Necmeddin Erbakan University)
Asst. Prof. Dr. Hannah INBARAN (Periyar University)
Asst. Prof. Dr. İlker YILDIZ (Abant İzzet Baysal University)
Asst. Prof. Dr. Kasım ÖZACAR (Karabük University)
Asst. Prof. Dr. Mehmet ŞİMŞEK (Düzce University)
Asst. Prof. Dr. Nesrin AYDIN ATASOY (Karabük University)
Asst. Prof. Dr. Nizar BANU (B S Abdur Rahman University)
Asst. Prof. Dr. Nursel YALÇIN (Gazi University)
Asst. Prof. Dr. Oktay AYTAR (Abant İzzet Baysal University)
Asst. Prof. Dr. Omar DAKKAK (Karabük University)
Asst. Prof. Dr. Ömer Kaan BAYKAN (Konya Technical University)
Asst. Prof. Dr. Ömer Muhammet Soysal (Louisiana State University)
Asst. Prof. Dr. Şafak BAYIR (Karabük University)
Asst. Prof. Dr. Şafak KAYIKÇI (Abant İzzet Baysal University)
Asst. Prof. Dr. Yasin ORTAKCI (Karabük University)
Asst. Prof. Dr. Yüksel ÇELİK (Karabük University)
Asst. Prof. Dr. Zafer ALBAYRAK (Karabük University)

Secretary

Res. Asst. Dr. Oğuzhan MENEMENCİOĞLU, Karabük University
Res. Asst. Ayşe Nur ALTINTAŞ, Karabük University
Res. Asst. Berna GÜNEŞ, Karabük University
Res. Asst. Elif KABULLAR, Karabük University
Res. Asst. Furkan SABAZ, Karabük University
Res. Asst. İdris KAHRAMAN, Karabük University
Res. Asst. Mehmet Zahid YILDIRIM, Karabük University
Res. Asst. Sait DEMİR, Karabük University
Res. Asst. Yasemin SANDAL, Karabük University

Welcome Address

It is a pleasure for us to offer you Abstracts Book for the 2nd International Conference on Advanced Technologies, Computer Engineering and Science ICATCES 2019. Our goal was to bring together leading academic scientists, researchers and research scholars to exchange and share their experiences and present their latest research results, ideas, developments, and applications about all aspects of advanced technologies, computer engineering and science. We decided to organize this event with the encouragement of our colleagues in the hope of transforming the event into a symposium series. Our warmest thanks go to all invited speakers, authors, and contributors of ICATCES 2019 for accepting our invitation. We hope that you enjoy the symposium and look forward to meeting you again in one of the forthcoming ICATCES 2020 event.

Contents

Welcome Address	iv
Proceedings	viii
Improving classification performance on microscopic images using Generative Adversarial Networks (GAN)	1
Analyzing of the wound healing by implementing image processing algorithms on FPGA	7
Speckle reduction in images using non-local means filter and variational methods	11
Real time activity recognition using weighted dynamic time warping	20
Remote controlled vehicle for surface and underground object detection	25
Estimation relationship between electricity consumption and urban area from night satellite imagery: A case study for Istanbul	29
Classification of flower species using convolutional neural networks	33
Real-time diseases detection of grape and grape leaves using Faster R-CNN and SSD MobileNet architectures	39
Improved network intrusion detection system using deep learning	45
Sentiment analysis for hotel reviews with recurrent neural network architecture	50
Prediction of absenteeism at work with machine learning algorithms	54
Automated diagnosis of tuberculosis using deep learning techniques	58
A system based on image processing and deep CNN features for classification of defective fruits	63
A survey of joint routing and energy optimization techniques for underwater acoustic sensor networks	71
An application of artificial bee colony algorithm to fatigue life estimation of magnesium alloy	77
An improved crow search algorithm for continuous optimization problems	81
Analysis of electronic countermeasure effects on air defense systems	85
Performance evaluations of meta-heuristic algorithms on solving group elevator control problem	90
Performance comparison between genetic algorithm and ant colony optimization on traveling salesman problem	95
On forecasting ability of a logistic model for prostate cancer under intermittent androgen suppression therapy	102
Comparison of classification algorithms on heart disease data	106
ECG signal classification with neural network ensemble	111
Detection of wart treatment method using machine learning algorithms	115
Detecting student engagement in e-learning environment based on head pose estimation and eye aspect ratio	119
Functional brain network analysis under cognitive task	124
A suggestion for electronic election system based on blockchain	129
Assessing the architectural quality of software projects from an organizational perspective	135
An overview on data-driven prognostic methods of li-ion batteries	139
Game design for rehabilitation of children with disabilities by using depth sensor	142
Prioritizing manual black-box tests using topic modeling	146
Increasing air traffic control efficiency with integrated human machine interface	150
Comparison of classification algorithms in terms of model performance criteria in data mining applications	153
Privacy scoring over professional OSNs: More central users are under higher risk	157

Recent advances and future trends in industrial communication networks: M2M, I-IoT and inter-operability issues	162
Raspberry Pi based personalized encoding ID system inspired by the quick response code software and hardware design	168
The usage statistics of new HTML5 semantic elements in the ClueWeb12 dataset	173
Internet of Things based remote monitoring system design for industrial plants	177
3D scene reconstruction using a swarm of drones	181
Development of a smart helmet for digital data collection and applications on construction sites	185
Android secure camera application	193
A recommendation system for seattle public library using Naïve Bayes Classifier	197
Server based indoor location and navigation using beacon devices	201
A conceptual design for managing internet of things devices in emergency situations	206
Can we fight social media with multimedia learning systems?	212
Increasing the use of digital banking applications for the baby boomer generation by easy interface	219
Detection of P300 ERP waves by different classification methods	223
Investigating personalised applications in MOOCs: the challenge of achieving transparency	227
Unification of IT process models into a simpler framework	231
An IoT based mobile radar system	238
Web and mobile based online joint working platform development for university students	244
Human computer interaction with one-shot learning	249
Local statistical features for multilingual artificial text detection from video images	255
Real time distributed controller for delta robots	260
An overview of captcha systems and bypassing math captcha	265
A review on Nvidia GauGan	269
A simulation based harmony search algorithm for part routing optimization problem	273
Multi class tag prediction on stack overflow dataset	277
A mobile indoor/outdoor augmented reality application for architecture	281
Reassembly of synthetically fractured objects	286
Classification of EEG patterns by using Katz Fractal Dimension	289
Learning semi-supervised nonlinear embeddings for domain-adaptive pattern recognition	295
Tree-seed programming for symbolic regression	300
An automated deep learning approach for bacterial image classification	304
Image processing on electrophoresis image with embedded system	309
Feature selection for text classification based on term frequency and inter-class standard deviation	311
Comparison of machine learning algorithms for franchise approval	315
Face detection using forensic software and deep learning methods on images and video files	320
Removing redundant primes in multiple-valued logic cover by cube algebra	325
Detecting point mutations in next-generation sequencing data	327
IP packet marking and forwarding based on content type with SDN	329
NetCar: A testbed for mobile sensor networks	334
Modeling and simulation of a reconfigurable microstrip antenna for wireless communication and mobile	339
Node weighting method in centrality measure of complex networks	342
Deep learning based web application security	349
Evaluation of student academics performance via machine learning algorithms	355
Criminological evaluation of cyber attacks on information and network security	360
Digital forensics in social media; evaluation in the context of system, informatics, network and cyber security	363
Gyroscope-accelerometer controlled smart disabled wheelchair	367
The optimization of the process of recognition fingerprint through the minutiae technique	371
An automated GIS tool for property valuation	376
Intelligent examination glove system design for use in medical education and user interface application	380
Brain tumor detection via active contours and scale invariant feature transform	384
Spatial preference system for roads maintenance	388
Spatial preferences decision support tool for historical building restoration	390

Satellite images classification in geographic information systems	392
Machine learning and satellite images for agricultural areas determination	394
TDMA scheduling for real-time flows in cluster-based wireless sensor networks	396
Deep learning approaches for traffic flow predictions on signal-controlled intersections	401
Securing TLS from MITM incursion using Diffie-Hellman	407

Proceedings

Improving classification performance on microscopic images using Generative Adversarial Networks (GAN)

Y.Y. BAYDILLI¹, Ü. ATILA¹, K. AKYOL²

¹ Department of Computer Engineering, Karabük University, Karabük/Turkey, yusufbaydilli@karabuk.edu.tr

¹ Department of Computer Engineering, Karabük University, Karabük/Turkey, umitatila@karabuk.edu.tr

² Department of Computer Engineering, Kastamonu University, Kastamonu/Turkey, kakyol@kastamonu.edu.tr

Abstract—A successful modeling can be achieved with a data set containing the correct information. However, achieving accurate information is only provided thanks to balanced distribution of features of samples that lie in the data set. In this study, the Generative Adversarial Networks (GAN) method was used to increase the number of samples belonged to a medical data set (White Blood Cells) that containing a small number of images. Accurate learning of the specific characteristics of both the data and classes allowed the classification success to be maximized.

Keywords—classification, medical data set, deep learning, generative adversarial networks (GAN).

I. INTRODUCTION

DEEP learning has become a very attractive research area especially in recent years due to its advantages. The most important advantage of these methods compared to classical machine learning methods is that they allow an end-to-end training without requiring any pre-processing stage [1]. However, to create a stable end-to-end model, the data set must have some specific properties. The most important of these is that the samples belonging to the different classes of the owned data set should be balanced and quite numerous. Otherwise, training with small data sets will result in memorization of data and failure to learn the correct neuronal activities to classify test data [2]. In this respect, researchers that has small data sets seek the ways to increase the number of data for a robust modeling. One of the methods frequently used for this purpose is “data augmentation” method. This method creates copies of the samples in the data set in different variations. Thus, the architecture can make more inferences from each sample [3].

Although data augmentation methods can increase the success rates on data to be tested, these rates can only increase to a point. Therefore, researchers need to add specific samples that will facilitate the distinction between classes in the data sets. In particular, this is a huge problem for researchers who work with medical images, due to data collecting process suffers from both cost and ethical concerns. So, researchers must obtain maximum inference from the data they have.

Image synthesis emerges as a method that can overcome this problem. Due to it is a highly studied subject in the deep

learning community, there are some quite successful methods (auto-regressive models, auto-encoders), in the literature [4]–[6]. Nevertheless, in 2014, more realistic and sharper images were produced compared to other methods using Generative Adversarial Networks (GAN) which is proposed by I. Goodfellow [7]. From this date onwards, with the modifications and additions performed on this method, it was possible to build architectures that can achieve state-of-the-results [8].

The most important feature that distinguishes medical images from real-world images is that the data carried in each pixel is significant [9]. Therefore, it is expected that the model proposed to work on this data does not cause data loss and reflect the specific characteristics of the training data. In this respect, in the literature, it is possible to find some studies that can overcome the problems mentioned above by the successful application of GAN method in medical images (MRI, CT, Retinal Images) in order to increase the success in tasks such as classification, lesion detection etc. [10]–[14].

White blood cells (WBC) have an important role in the immune system. Therefore, the type, number and intensity of white blood cells provide important information about the patient’s health status. In this respect, the correct classification of the WBC types is of great importance for the hematologist examining the peripheral smear images. When the literature is examined, it is possible to observe some studies that perform WBC classification using deep learning methods [15]–[17]. However, as mentioned earlier, these studies were found to be adversely affected by insufficient and unbalanced data set problems. Therefore, it was observed that researchers applied methods such as data augmentation [15], [16], transfer learning [17], hybrid models [16] to overcome these problems and to increase the performance of classification.

In this study, a different approach to the problem was introduced and the ability of the GAN method on the microscopic world images was questioned. In order to increase WBC classification performance, synthetic data were produced by GAN method. Thus, the generated images were added to the training set and the classification success rate on the test data able to be increased. The main contributions of this study to the literature are:

- It is one of the first studies of blood cell image synthesis in the literature along with synthesis of red blood cell images [18],
- As far as we know, it is the first attempt to synthesize white blood cell images in the literature,
- It has been possible to move the classification performance to the highest point on a data set that has five classes with very little data.

The rest of the article is as follows; In the second part, GAN and used method are explained, and then, in the third part the data set is introduced. In the fourth part, the experimental results of the study are discussed. Finally, in the conclusion part, the main contribution of the study is given.

II. GENERATIVE ADVERSARIAL NETWORKS (GAN)

GAN architectures basically consist of two parts; generator and discriminator. There is a continuous race between these two parts during the training. While the discriminator has the task of learning training data and defining whether the generated images are real or fake, the generator part tries to create more realistic images according to feedback coming from the discriminator. As the training progresses, while the discriminator will be able to evaluate the images better due to it will have more information about images, at the same time, the generator will be forced to create fake images that best reflect the data distribution of the real images in order to fool the discriminator. A standard GAN architecture can be seen in Figure 1.

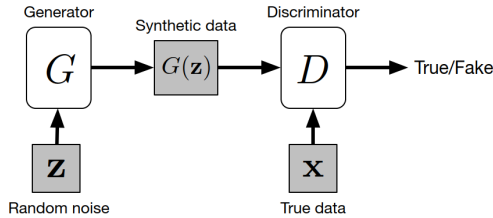


Figure 1: GAN architecture. While the generator tries to generate meaningful data from the noise vector z , the discriminator will try to decide whether the generated image $G(z)$, and the real images X , are true/fake [19].

Both the generator and the discriminator are a neural network. Therefore, the cost values are calculated for the two parts of the model in each epoch using Equation 1 and necessary weight updates are performed.

$$\begin{aligned} \max_D V(D) &= \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \\ &\quad \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \\ \min_G V(G) &= \mathbb{E}_{z \sim p_z(z)} - [\log D(G(z))] \end{aligned} \quad (1)$$

GAN training is a very difficult process and requires patience due to it is unstable. it is therefore reasonable to use all kind of information that can be extracted from the training set to ensure a positive contribution to training process. In their study, Mirza, M. and Osindero, S. showed that more successful

results could be obtained by using the class information of the images included in the training data [20]. In 2017, Odena, A. et al. move this idea one step forward and proposed the AC-GAN model by adding “auxiliary classification” to their models [21]. In this way, instead of just making a distinction, they also provided the discriminator tries to estimate the class of the images by using the training data. Thus, the realism of the images created by the generator was increased (Figure 2).

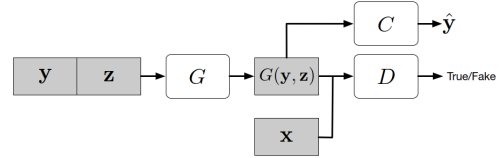







Figure 2: AC-GAN architecture. Besides random noise, the generator also feeds on labels to improve training performance. Classes of the images will also be estimated through the auxiliary classifier C [19].

The GAN architecture to be used in our study is determined as AC-GAN in order to create images that will reflect the class characteristics of each image as the training data set is small-sized. On the other hand, due to that GAN method has some problems such as unstable training and mode collapse (continuous production of the same samples), some techniques proposed by Salimans, T. vd (mini-batch discrimination etc.) have been added to the model [22]. Thus, the architecture with the $\sim 16.8M$ parameters in total is trained for $1,000$ epochs.

III. DATA

One of the frequently encountered obstacles in the classification of WBCs is the unbalanced data set problem. WBCs found in the blood have different density values [23]. Therefore, it is often not possible to obtain enough samples from each class (Basophil, Eosinophil, Lymphocyte, Monocyte, Neutrophil). For this reason, LISC [24] data set was chosen, which has approximately equal number of images in each class, in order to perform the tests. Slides with 128×128 dimensions were taken from the data set and saved as to be each sample will have one WBC. The properties of the data set can be seen in Table 1.

Table 1: LISC data set.

WBC Sub-type	Bas.	Eos.	Lym.	Mon.	Neu.
Sample Images					
No. of Channels	3				
No. of Samples	54	42	59	55	56
Training Set	47	36	46	51	46
Test Set	7	6	13	4	10

A total of 266 samples existed in the data set were divided into two categories as training (85%) and test data (15%). During the GAN training, only the samples in the training set were used. In this way, it is possible to analyze the positive contributions of the generated images to the classification of test data.

IV. EXPERIMENTAL RESULTS

As mentioned earlier, GAN has a very long and unstable training process [25]. The right hyper-parameters that allowing the model can produce stable results are mostly achieved by trial/error, which means extra training time. Therefore, in the tests performed, the samples included in the training and test data were reduced to 32×32 size in order to decrease this period. Loss values calculated by the generator and the discriminator during the training process can be seen in Figure 3.

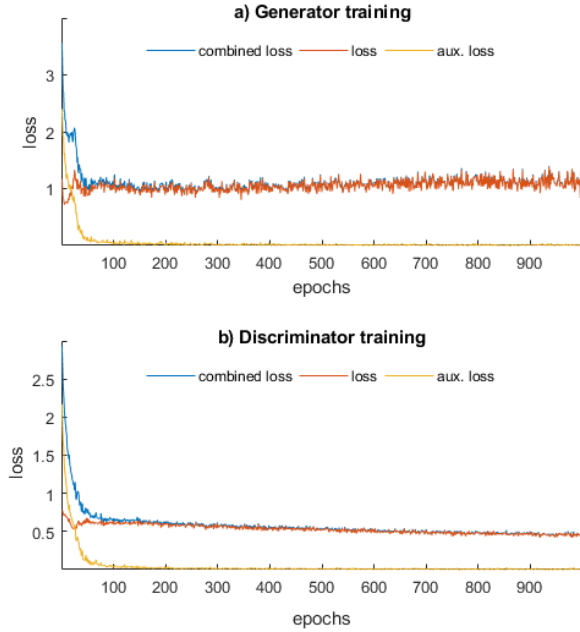


Figure 3: Losses calculated at each epoch in the training process of the generator and discriminator parts. The most important result that can be extracted from these graphs is that the auxiliary classifier has small loss values for both real and fake images, so; the classes of images have been learned.

The generator loss values showed stable characteristics. So, it can be interpreted as the images generated are quite similar to the real images and the data distribution of the real images has been learned correctly. On the other hand, the fact that the auxiliary classifier section, in which the classes are predicted, produces small loss values indicates that the classes were predicted correctly.

In the next stage, samples were produced in every *100 epochs* to examine the success of the training process. The produced samples can be seen in Figure 4.

In the next stage of the study, two data sets were created and tests were carried out with a Convolutional Neural Network (CNN) architecture to examine how the produced images contributed to the classification performance. The first set (data set-1) was created as would have approximately 2,000 samples in each class by using the data augmentation technique (shear, zoom, flipping, etc.). In the second set (data set-2), the samples produced by GAN was added to the original data, thus, the

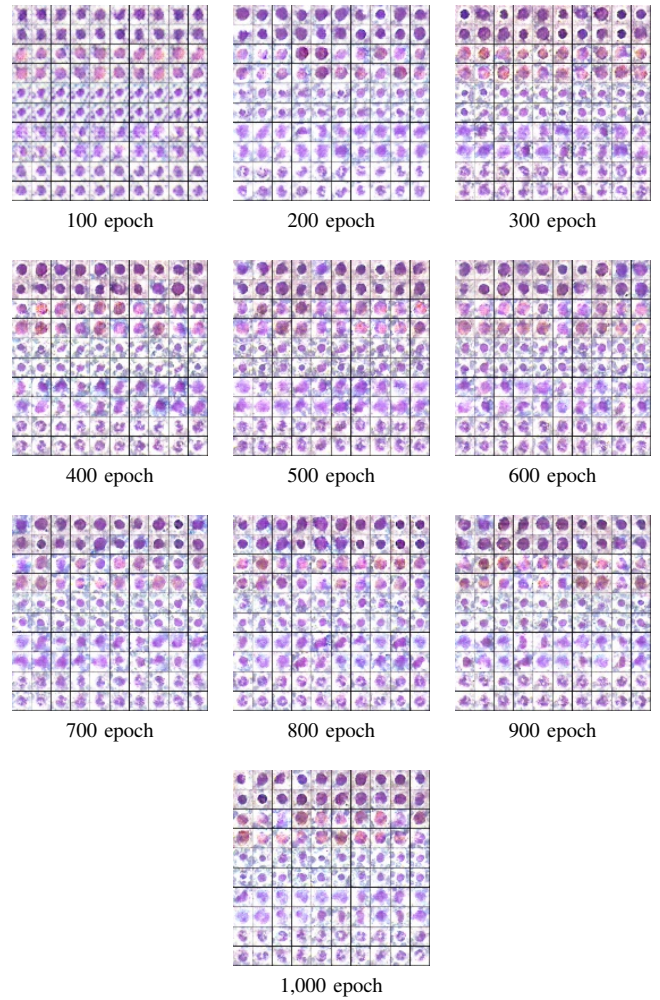


Figure 4: Produced WBC samples. Although the quality of the images is not as good as the real images, it is seen that the samples reflect the characteristics of the classes.

number of samples was tripled for each class. Then, the number of samples was equalized to the data set-1 through the data augmentation technique, again.

The main idea of using the data augmentation technique here was that although the number of samples was increased by GAN, providing to obtain the data size that can form a stable model. Moreover, in this way, contribution of GAN to training data can be evaluated better [26]. The properties of the generated data sets can be seen in Table 2.

The CNN architecture that used to classify data sets is seen in Figure 5. In the training process *10-fold cross validation* method was used to producing robust results on whole set. Besides, we used “network-in-network” [27] and “batch normalization” [28] techniques while building the model. The proposed model was trained for *100 epochs* for both data sets. Results can be observed in Figure 6.

Table 2: Created data sets.

WBC Sub-type Data sets	Basophil	Eosinophil	Lymphocyte	Monocyte	Neutrophil	TOTAL
(Data set-1) Org. + Aug.	47 + 1,944	36 + 1,958	46 + 1,951	51 + 1,948	46 + 1,947	9,974
(Data set-2) Org. + GAN + Aug.	47 + 94 + 1,850	36 + 72 + 1,886	46 + 92 + 1,859	51 + 102 + 1,846	46 + 92 + 1,855	9,974
Test Set	7	6	13	4	10	40

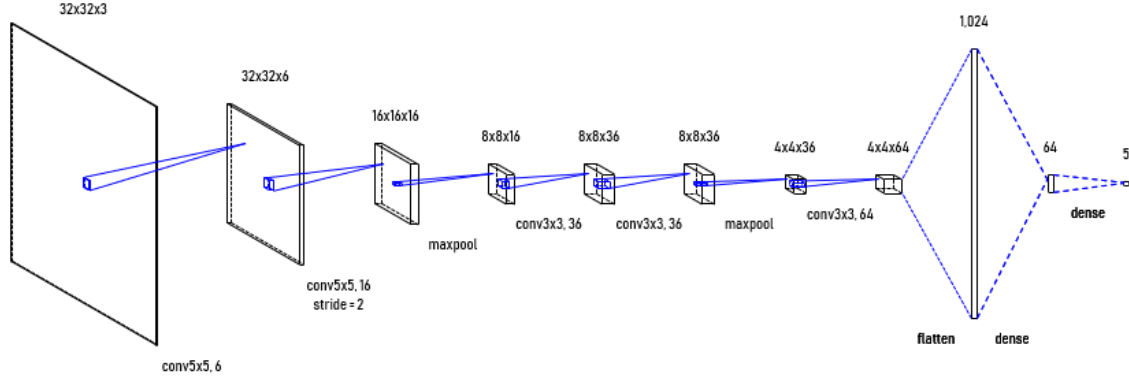
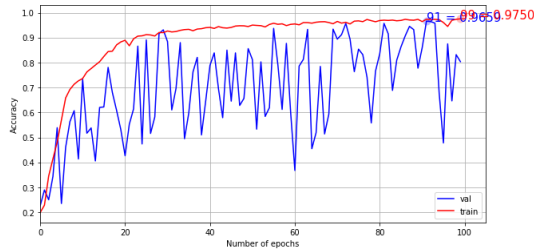
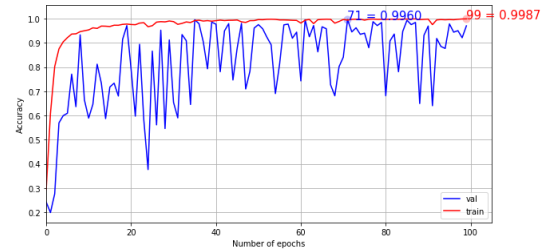


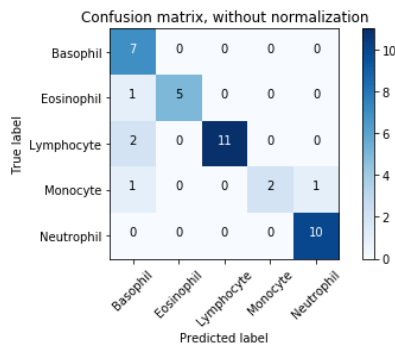
Figure 5: Proposed model. By defining $stride = 2$ rather than max-pool, it was aimed to reduce training parameters while keeping the information loss at minimum. Furthermore, receptive field was increased in all convolutional layers thanks to $padding = same$



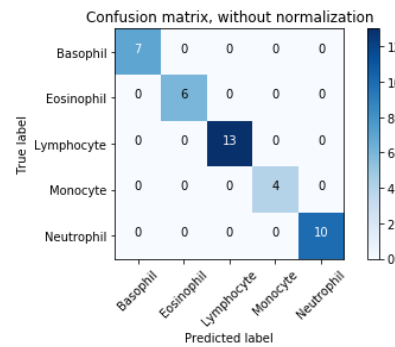
a) accuracy curve for data set-1



b) accuracy curve for data set-2



c) confusion matrix for data set-1



d) confusion matrix for data set-2

Figure 6: (a, b) calculated loss values in each epoch for data set-1 and data set-2, (c, d) confusion matrices for data set-1 and data set-2. Although the model able to reach high training success rates for both data sets, prediction accuracy on data set-1 was measured as 87.5%. However, 100% success rate was observed after training with data set-2.

It would not be wrong to say that both models are successful in learning the training data. However, in the validation phase on data set-1, a more unstable behavior was observed by the model (Figure 6-a). Since 10-fold cross validation method is used, this type of behavior is expected, but validation on data set-2 can be considered as more stable (Figure 6-b).

After the training was finalized, when the model performed classification on the test data that the model had never seen during the training stage, it was seen that the prediction value for the data set-1 was 87.5% (Figure 6-c). Thus, it is possible to interpret that the model memorized the training data and could not learn the features that could distinguish between classes.

Using data set-2, it was seen that the model carried the prediction success on the test data up to “perfect” point (Figure 6-d). Thanks to GAN, characteristics of the data and the distribution of meaningful information that lies in the pixels were able to be learned correctly by the model. In this way, a contribution of 12.5% was made to the classification success rate by adding unique data to training.

V. CONCLUSION

The GAN method has been one of the most popular and deep learning methods since 2014. With this method, one can produce synthetic images that similar to original samples by performing an unsupervised training process on images. Moreover, performing a semi-supervised training using the class information is possible, too. Hence, the model is forced not only to produce images, but also to create samples that reflect class characteristics.

Besides, using this type of training model, the insufficient and unbalanced data set problems can be solved. So, it is predicted that the model to be trained will get more inference than images, and learn more accurately the exact parameters that will provide successful classification. In this way, success rates can be carried to higher levels.

In order to confirm this prediction, WBC sub-types were classified by using peripheral smear images. For this purpose, two data sets were created, while the first data set was formed using only traditional data enhancement techniques, the second data set was improved by both GAN and data augmentation technique. At the end, it was observed that the classification success can be increased to 100% by only increasing up to three times the number of original data in the training set. Thus, this method has been shown to be a good alternative for solving problems for researchers who are struggling with the insufficient and unbalanced data set problem and especially working on medical images.

REFERENCES

- [1] D. Shen, “Deep learning in medical imaging analysis,” in *19th International Conference on Medical Image Computing & Computer Assisted Intervention*, Athens, Greece, 2016. [Online]. Available: https://cs.adelaide.edu.au/~dlmia2/DG_talk.pdf
- [2] S. K. Zhou, H. Greenspan, and D. Shen, *Deep Learning for Medical Image Analysis*, 1st ed. London; San Diego: Academic Press, Feb. 2017.
- [3] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, “Convolutional neural networks for medical image analysis: Full training or fine tuning?” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016. [Online]. Available: <http://arxiv.org/abs/1706.00712>
- [4] A. van Oord, N. Kalchbrenner, and K. Kavukcuoglu, “Pixel recurrent neural networks,” in *33rd International Conference on Machine Learning (ICML 2016)*, vol. 48. New York City, NY, US: Proceedings of Machine Learning Research, 2016, pp. 1747–1756.
- [5] A. van den Oord, N. Kalchbrenner, L. Espeholt, K. Kavukcuoglu, O. Vinyals, and A. Graves, “Conditional image generation with PixelCNN decoders,” in *Advances in Neural Information Processing Systems 29 (NIPS 2016)*. Barcelona, Spain: Curran Associates, Inc., 2016, pp. 4790–4798. [Online]. Available: <http://papers.nips.cc/paper/6527-conditional-image-generation-with-pixelcnn-decoders.pdf>
- [6] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, “Autoencoding beyond pixels using a learned similarity metric,” in *Proceedings of the 33rd International Conference on International Conference on Machine Learning*, ser. ICML’16, vol. 48. New York, NY, US: JMLR.org, 2016, pp. 1558–1566, event-place: New York, NY, USA. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3045390.3045555>
- [7] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *arXiv:1406.2661 [cs, stat]*, Jun. 2014, arXiv: 1406.2661. [Online]. Available: <http://arxiv.org/abs/1406.2661>
- [8] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” in *6th International Conference on Learning Representations*, Vancouver, BC, Canada, 2018, arXiv: 1511.06434. [Online]. Available: <http://arxiv.org/abs/1511.06434>
- [9] J.-G. Lee, S. Jun, Y.-W. Cho, H. Lee, G. B. Kim, J. B. Seo, and N. Kim, “Deep learning in medical imaging: General overview,” *Korean Journal of Radiology*, vol. 18, no. 4, pp. 570–584, 2017. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5447633/>
- [10] P. Costa, A. Galdran, M. I. Meyer, A. M. Mendonça, and A. Campilho, “Adversarial synthesis of retinal images from vessel trees,” in *Image Analysis and Recognition*, ser. Lecture Notes in Computer Science. Montreal, QC, Canada: Springer International Publishing, 2017, pp. 516–523.
- [11] A. Chatsias, T. Joyce, R. Dharmakumar, and S. A. Tsiftaris, “Adversarial image synthesis for unpaired multi-modal cardiac data,” in *Simulation and Synthesis in Medical Imaging*, ser. Lecture Notes in Computer Science. Québec City, QC, Canada: Springer International Publishing, 2017, pp. 3–13.
- [12] J. M. Wolterink, A. M. Dinkla, M. H. F. Savenije, P. R. Seevinck, C. A. T. van den Berg, and I. Išgum, “Deep MR to CT synthesis using unpaired data,” in *Simulation and Synthesis in Medical Imaging*, ser. Lecture Notes in Computer Science. Québec City, QC, Canada: Springer International Publishing, 2017, pp. 14–23.
- [13] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, “GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification,” *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231218310749>
- [14] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, “Medical image synthesis with context-aware generative adversarial networks,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, ser. Lecture Notes in Computer Science. Quebec City, QC, Canada: Springer International Publishing, 2017, pp. 417–425.
- [15] J. W. Choi, Y. Ku, B. W. Yoo, J.-A. Kim, D. S. Lee, Y. J. Chai, H.-J. Kong, and H. C. Kim, “White blood cell differential count of maturation stages in bone marrow smear using dual-stage convolutional neural networks,” *PLoS ONE*, vol. 12, no. 12, p. e0189259, Dec. 2017. [Online]. Available: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0189259>
- [16] J. Zhao, M. Zhang, Z. Zhou, J. Chu, and F. Cao, “Automatic detection and classification of leukocytes using convolutional neural networks,” *Medical & Biological Engineering & Computing*, vol. 55, no. 8, pp. 1287–1301, Aug. 2017. [Online]. Available: <https://link.springer.com/article/10.1007/s11517-016-1590-x>
- [17] A. I. Shahin, Y. Guo, K. M. Amin, and A. A. Sharawi, “White blood cells identification system based on convolutional deep neural learning

- networks,” *Computer Methods and Programs in Biomedicine*, Nov. 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S016926071730411X>
- [18] O. Bailo, D. Ham, and Y. M. Shin, “Red blood cell image generation for data augmentation using conditional generative adversarial networks,” *arXiv:1901.06219 [cs]*, Jan. 2019, arXiv: 1901.06219. [Online]. Available: <http://arxiv.org/abs/1901.06219>
 - [19] H. Huang, P. S. Yu, and C. Wang, “An introduction to image synthesis with generative adversarial nets,” *arXiv:1803.04469 [cs]*, Mar. 2018, arXiv: 1803.04469. [Online]. Available: <http://arxiv.org/abs/1803.04469>
 - [20] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” *arXiv:1411.1784 [cs, stat]*, Nov. 2014, arXiv: 1411.1784. [Online]. Available: <http://arxiv.org/abs/1411.1784>
 - [21] A. Odena, C. Olah, and J. Shlens, “Conditional image synthesis with auxiliary classifier GANs,” in *Proceedings of the 34th International Conference on Machine Learning*, ser. ICML’17, vol. 70. Sydney, NSW, Australia: JMLR.org, 2017, pp. 2642–2651, event-place: Sydney, NSW, Australia. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3305890.3305954>
 - [22] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, and X. Chen, “Improved techniques for training GANs,” in *Advances in Neural Information Processing Systems 29*. Curran Associates, Inc., 2016, pp. 2234–2242. [Online]. Available: <http://papers.nips.cc/paper/6125-improved-techniques-for-training-gans.pdf>
 - [23] A. Adewoyin and B. Nwogoh, “Peripheral blood film - A review,” *Annals of Ibadan Postgraduate Medicine*, vol. 12, no. 2, pp. 71–79, Dec. 2014. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4415389/>
 - [24] S. H. Rezatofighi and H. Soltanian-Zadeh, “Automatic recognition of five types of white blood cells in peripheral blood,” *Computerized Medical Imaging and Graphics*, vol. 35, no. 4, pp. 333–343, Jun. 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0895611111000048>
 - [25] X. Wu, K. Xu, and P. Hall, “A survey of image synthesis and editing with generative adversarial networks,” *Tsinghua Science and Technology*, vol. 22, no. 6, pp. 660–674, Dec. 2017. [Online]. Available: <http://ieeexplore.ieee.org/document/8195348/>
 - [26] L. Perez and J. Wang, “The effectiveness of data augmentation in image classification using deep learning,” *arXiv:1712.04621 [cs]*, Dec. 2017. [Online]. Available: <http://arxiv.org/abs/1712.04621>
 - [27] M. Lin, Q. Chen, and S. Yan, “Network in network,” *arXiv:1312.4400 [cs]*, Dec. 2013. [Online]. Available: <http://arxiv.org/abs/1312.4400>
 - [28] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *32nd International Conference on Machine Learning*, vol. 37, Lille, France, Feb. 2015, pp. 448–456. [Online]. Available: <http://arxiv.org/abs/1502.03167>

ANALYZING OF THE WOUND HEALING BY IMPLEMENTING IMAGE PROCESSING ALGORITHMS ON FPGA

C. DOGAN¹, A. OZFIRAT¹ and M. PALANDOKEN¹

¹ Izmir Katip Celebi University, Izmir/Turkey, ccemal.dogann@gmail.com, anilozfirat@gmail.com
merihpalandoken@ikc.edu.tr

Abstract - Field-Programmable Gate Arrays (FPGA) have significant attributes, such as, parallel computation and reconfigurable logic arrays. These attributes have crucial role in the image processing algorithms. In the paper, the area of the wounded region in a sample image has been calculated. The sample wound image has been transformed into the vector form in Matlab platform and sent via Universal Asynchronous Receiver-Transmitter (UART) to the FPGA kit. The resulting image after the image processing algorithms have been sent via UART to the PC to calculate the area of the wounded region in Matlab platform.

Keywords – Image Processing, FPGA, UART, Serial Communication, Morphological Operations.

I. INTRODUCTION

Image processing algorithms have been used in numerous applications, for instances, intelligent transportation systems, remote sensing, moving object tracking and biomedical imaging techniques. The main algorithms which lie behind the image processing algorithms are multiplication and summation, for example, masking and convolution. Implementing image processing algorithms in hardware platforms whose microprocessor has ability to do calculations in parallel manner provide huge advantage to the user. In comparison to the ordinary microcontrollers, FPGAs have capability to the parallel computation. Therefore FPGAs have been widely used in digital image processing, especially real-time digital image processing applications [1].

In the literature, image processing algorithms which implemented in FPGA kits are available, however, most of these applications are not intended for a specific event, but for general image processing algorithms which implemented in FPGA kits. In 2017, Boudjelal and his friends have implemented edge detection algorithms in FPGA kit. They have embedded the sample image which is in .coe format in the Read-Only Memory (ROM) that is implemented by Xilinx Core Generator. In order to see the resulting image after the edge detection algorithm, they have used Matlab platform [2]. Abdul Manan has been implemented median filter algorithm and morphological operations, namely dilation and erosion, in FPGA kit. He has been used UART communication protocol, in order to made the communication between PC and FPGA kit in 2011 [3]. In 2015, Jin Zhao has designed video/image processing algorithms, for instance, lane departure warning

system, traffic sign detection system using speeded up robust features (SURF), in the Simulink platform. After that, those algorithms have been implemented in FPGA kit by the help of Matlab HDL Coder [4]. Divya and his friends have been implemented coin counting application in FPGA by using brightness manipulation, operating threshold and contrast stretching algorithms [5]. In 2000, Anthony Edward Nelson has implemented median filtering and closing operation in a sample image in both Matlab and FPGA platforms and has compared the resulting images [6].

II. IMPLEMENTATION DETAILS AND METHODS

In this section, image processing algorithms that have been used in the proposed work have been discussed. In addition, resulting images after each image processing algorithms have been presented.

A. Intensity Level Slicing

Intensity transformation functions, as its name indicated, are the transformation of the pixel's intensity levels into the desired output intensity levels. Depending on the application different types of intensity transformation functions can be chosen. In this paper, one of the piecewise-linear transformation functions, namely the intensity level slicing algorithm has been used. Intensity level slicing is an algorithm that highlights a particular range of intensities in a digital image [7].

In the proposed work, the wounded region in the sample image is highlighted by using intensity level slicing algorithms to eliminate other gray levels from the wounded region itself.

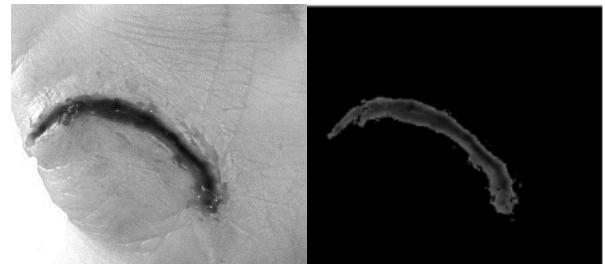


Figure 1: (a) Gray scale input image (b) Resulting image after the intensity level slicing algorithm in Matlab

B. Binarization

Binarization is process of converting gray scale images into a binary image whose pixel values are only 1 or 0. It can be classified as global binarization methods, for instance, Otsu method, Kittler method, and local binarization methods, for example, Niblack Method, Adaptive Method. While converting gray scale images in global binarization methods single threshold value is used, however, in local binarization method the threshold values are calculated locally and applied in the corresponding region [8].



Figure 2: Resulting image after the binarization algorithm in Matlab

After intensity level slicing algorithms have been applied on the sample image, binarization has been implemented. Holes in the binarized image, that is shown in Figure 2, makes miscalculation of the wounded region, therefore filling holes algorithm, which is closing operation, have been implemented and discussed in morphological operation section.

C. Morphological Operation

If geometric structures in binary or grayscale images are interested, morphological operations fundamentally based on set theory are used [9]. Filtering and morphological operations are similar to each other. Morphological operations are the same logical operation using a mask over an image as using a structuring element over an image. In filtering operations, multiply-accumulate operation is used, however, in morphological operations, results are depending on set theory. Structuring elements are small sets used in morphological operations and depending on the application it is chosen by user.

0	0	0	1	0	0	0
0	0	1	1	1	0	0
0	1	1	1	1	1	0
1	1	1	1	1	1	1
0	1	1	1	1	1	0
0	0	1	1	1	0	0
0	0	0	1	0	0	0

Figure 3: Diamond structure element

1. Dilation Operation

Dilation operation is one of the basic morphological operations which dilates the image boundaries. This operation is widely used for filling the holes in an image, and that

procedure is highly related with the structuring element itself. The dilation operation of an image A by structuring element B is defined as

$$A \oplus B = \{ z \in E | (B^s)_z A \neq \emptyset \} \quad (1)$$

where B^s denotes the symmetric of B, that is,

$$B^s = \{ x \in E | -x \in B \} \quad (2)$$



Figure 4: Resulting image after the dilation operation in Matlab

Dilation operation has been applied with the 7x7 diamond structuring element to the binarized image is shown in Figure 4. The object boundaries, which is wounded region in this case, has been dilated.

2. Erosion Operation

Erosion operation, in contrast of dilation operation is one of the basic morphological operation which erodes the image boundaries. Erosion makes the objects in a binary image smaller or thins, and it can be identified as morphological filtering operation [7]. The erosion operation of an image A by structuring element B is defined as

$$A \ominus B = \{ z \in E | B_z A \} \quad (3)$$

where B_z is the translation of B by the vector z,

$$B_z = \{ b + z | b \in B \} \quad z \in E \quad (4)$$

3. Closing Operation

Closing operation is combination of dilation and erosion operations and fills the holes in a image. Applying dilation operation to an image results in filling the holes. But this application, at the same time, dilates the boundaries of the interested object. In order to prevent the dilation on the interested object, erosion operation has to be applied to the resulting image after the dilation operation. The closing of an image A by structuring element B is defined as

$$A \cdot B = (A \oplus B) \ominus B \quad (5)$$

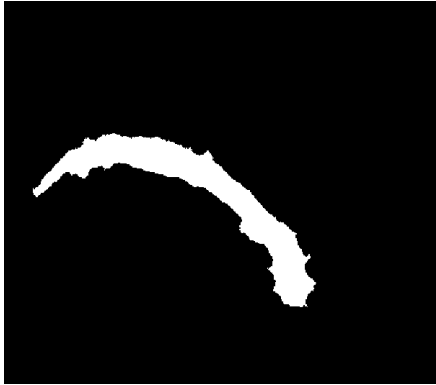


Figure 5: Resulting image after the closing operation in Matlab

Erosion operation has been applied to the dilated image which means the closing operation has been successfully implemented on the binarized image.

III. RESULTS

In the paper, Arty Z7-20 FPGA kit has been used to implement image processing algorithm. It has 650 MHz dual-core Cortex-A9 processor, 512 MB DDR3 RAM, 53200 Look-up Tables (LUTs) and 106400 Flip-Flops.

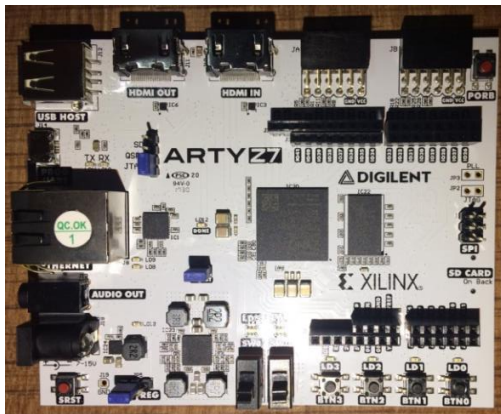


Figure 6: Arty Z7-20 FPGA kit

To implement the image processing algorithms on FPGA kit, Vivado Design Suite 2018.1 software platform has been used. The designed schematic which is shown in Figure 9 has been tested via simulation process as it is shown in Figure 7.

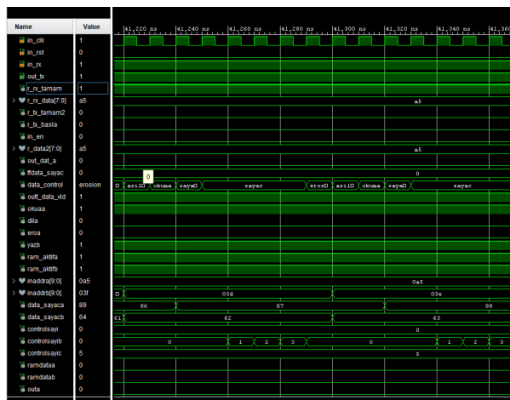


Figure 7: Simulation result of the project

Resource	Utilization	Available	Utilization %
LUT	590	20800	2.84
LUTRAM	18	9600	0.19
FF	488	41600	1.17
IO	4	210	1.90

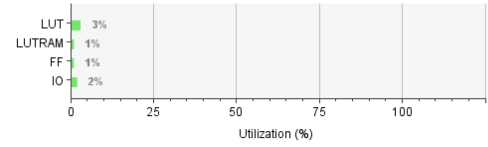


Figure 8: Utilization report of the proposed work

- The sample image which is taken as in Red-Green-Blue(RGB) format transformed into grayscale format and vectorized in Matlab platform.
- UART is implemented to make the serial communication between PC and FPGA board.
- Vectorized image has been sent to FPGA kit via UART Receiver module.
- The wounded region in the sample image has been highlighted by using intensity level slicing algorithms.
- Binarization algorithms have been applied to the sample image.
- Closing operation which is a morphological operation has been applied with 7x7 diamond structuring element to the binarized image.
- Resulting pixel values has been sent by UART transmitter module to Matlab to calculate the area of the wounded region.

In order to calculate the area of the wounded region in a sample image it is required to know the real distance between any two coordinates in the image. With the information above, the real distance between two pixels, and the area of one pixel square have been calculated. The area of the wounded region in a sample image is calculated by simple mathematics.

IV. CONCLUSION

In this paper, the area of the wounded region in a sample image has been calculated successfully by implementing image processing algorithms in FPGA kit.

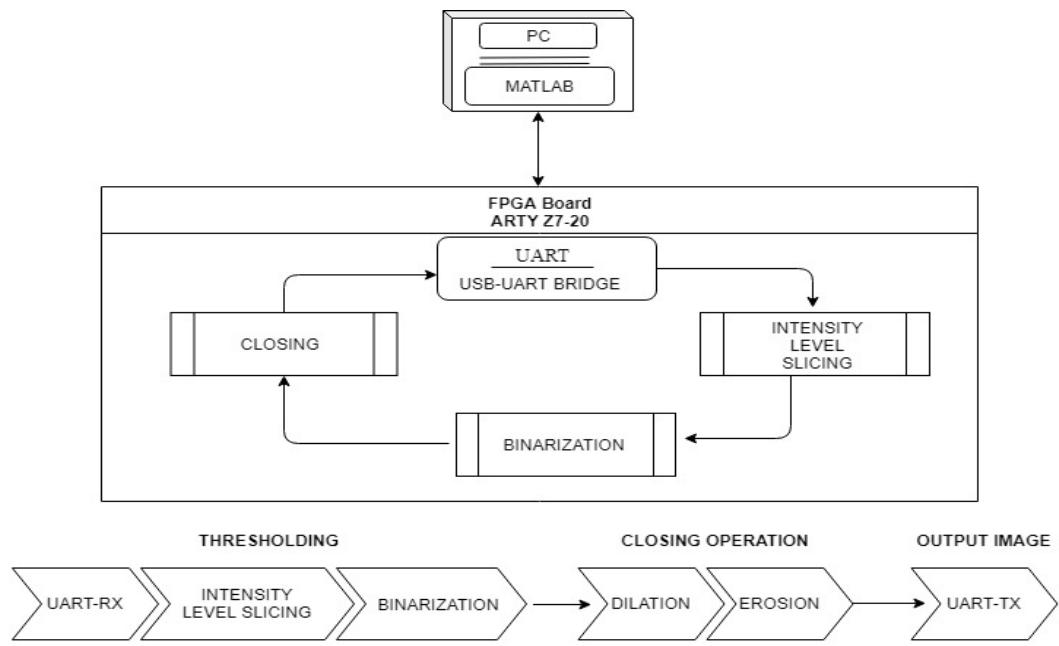


Figure 9: Block diagram of the project on FPGA

In the sample image, the real distance between the reference points are 6 cm which corresponds to 392 pixel. By using this knowledge one pixelsquare has been calculated as $2.342e-4 \text{ cm}^2$. After the closing operation all the ones have been counted and the area of the wounded region has been calculated as 2.42 cm^2 .

- [9] R. M. Haralick, S.R. Sternberg, X. Zhuang, "Image Analysis Using Mathematical Morphology," IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. PAMI-9, NO. 4, JULY 1987.

ACKNOWLEDGMENT

This project has been supported by the Scientific and Technological Research Council of Turkey (TUBITAK), Grand No: 1139B411801392.

REFERENCES

- [1] J. Xiong, Q. M. J. Wu, "An Investigation of FPGA Implementation for Image Processing," 2010 International Conference on Communications, Circuits and Systems (ICCCAS), July 2010.
- [2] F. A. Ferhat, L. A. Mohamed, O. Kerdjidi, K. Messaoudi, A. Boudjelal and S. Seddiki, "Implementation of SOBEL, PREWITT, ROBERTS Edge Detection on FPGA," The 2013 World Congress in Computer Science, At Las Vegas, Nevada, USA, July 2013.
- [3] A. Manan, "Implementation of Image Processing Algorithm on FPGA," AKGEC JOURNAL OF TECHNOLOGY, Vol. 2, No. 1, 2011.
- [4] J. Zhao, "Video/Image Processing on FPGA (Thesis)," Degree of master of science, Dept. Elec. and Comp. Eng., Worcester Polytechnic Institute, Worcester, USA, April 2015.
- [5] R. Dhanabal, S. K. Sahoo, V. Bharathi, Bh.S.R. P. Varma, D. Kalyan, V. Divya, "FPGA based image processing unit", 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO).
- [6] A. E. Nelson, "Implementation of image processing algorithms on FPGA hardware (Thesis)," Degree of master of science, Dept. Elec. Eng., Nashville, TN, May 2000.
- [7] R.C. Gonzalez, R.E. Woods, Digital Image Processing, USA, 2008.
- [8] P. Grdjet, N. K.Garg, "Binarization Techniques used for Grey Scale Images," International Journal of Computer Applications (0975 – 8887) Volume 71– No.1, June 2013.

Speckle Reduction in Images Using Non-Local Means Filter and Variational Methods

Ş. G. KIVANÇ¹, B. ŞEN², F. NAR³

¹ Department of Computer Engineering, Ankara Yıldırım Beyazıt University, Ankara/Turkey,
gkivanc@ybu.edu.tr

² Department of Computer Engineering, Ankara Yıldırım Beyazıt University, Ankara/Turkey, bsen@ybu.edu.tr

³ Department of Computer Engineering, Konya Food and Agriculture University, Konya/Turkey,
faith.nar@gidatarim.edu.tr

Abstract - In this study a novel approach which combines the advantages of Total Variation based Sparsity Driven Despeckling (SDD-QL) and Nonlocal Means is proposed in order to improve the despeckling quality. Both SDD-QL and Nonlocal Means have advantages and disadvantages on image despeckling. SDD-QL performs really well on homogeneous areas and quite fast but deteriorates the texture areas on the image. Nonlocal means preserves the textures and structures in the image but it is really slow. SDD-QL and Nonlocal Means are combined in a single cost function using a texture map as a weighting variable. Texture map is a binary matrix that is extracted from the image. Each value in the texture map indicates whether a pixel consists a texture or not. Nonlocal means based regularization term is applied on texture areas on the image and SDD-QL is applied on homogeneous areas. With combination of the two methods using the information from the texture map, a better speckle reduction performance is achieved without deterioration of texture areas. Speckle reduction performance of the proposed method is showed by using images with artificial speckle noise.

Keywords – NLTV, Speckle, Despeckle, SDD-QL, Nonlocal Means

1. INTRODUCTION

Remote Sensing is an increasingly growing and important topic. Remote Sensing and imaging can be used in military and commercial cases. Optical radars and Synthetic Aperture Radars (SAR) can be given as examples of remote sensing. These two types of radars are especially popular in remote sensing. Optical radars can get high quality images of surface of the earth but since it operates with optic lens it needs sunlight and good weather conditions to operate. SAR systems on the other hand operates on high powered electromagnetic pulses. System emits these pulses onto the surface of the earth and receives the backscattered signals. These signals are then converted to digital values and modeled as images. Since SAR systems works with electromagnetic pulses, it can operate on all types of weather conditions and without sunlight [1].

Synthetic Aperture Radar (SAR) has been widely used for

the last three decades because of its unique imaging capability. It can be used for target detection, change detection, crop detection in agriculture, vehicle detection for military, etc. Electromagnetic pulse system used in SAR systems gives SAR a big advantage over optical radars. But backscattering mechanism on SAR causes a noise called speckle. Speckle is a signal depended granular noise which degrades the quality of SAR images [2]. Speckle can depend on various factors such as features of the radar, surface of the earth, etc. Images with speckle noise becomes hard to work on. Removing the speckle from SAR images increase the quality of the image and makes it easier to work on SAR images with tasks such as target and change detection, classification and edge detection. Removing the speckle noise from a SAR image is called despeckling [2]. Despeckling is becoming a popular search field. In the past decades a lot of despeckling approaches have been offered to increase the quality of SAR images. Increasing number of radars that have been launched in the past decade indicates that SAR systems will become even more popular in the future.

One of the oldest approaches is called Lee Filter [3]. In Lee Filter, the value of a pixel is calculated over a fixed size window around the pixel with the local means and variance. The value of a pixel is changed with the one from the result of Lee filter. Another approach is called Frost filtering [4]. Frost is an adaptive filter. Frost uses locally observed mean and standard deviation to smooth homogeneous areas while preserving edge structure. Kuan filter [5] uses a similar approach to Lee filter. It uses nonstationary local mean and variance to smooth the image and adapt the changes in image statistics. Sigma filter [6] is proposed by Lee in the early 1980's. Sigma filter replaces the value of a pixel with the average of neighboring pixels that have gray level values within two noise standard deviations according to concerned pixel. Improved sigma filter is proposed to overcome the shortcomings of the original sigma filter [7]. Bilateral Filtering [8] smooths images by using nonlinear combination of nearby image values while preserving edges. The applications that applies bilateral filtering to SAR images can be found in [9, 10]. Mean, Median, Geometric, Wiener,

Bayesian, Local Sigma filters can also be used in SAR image despeckling [11, 12, 13, 14, 15, 16]. The study in [17] compares complex Wavelet Coefficient Shrinkage (WCS) and several standard speckle filters like Lee, Frost, Kuan, Geometric, Gamma, Kalman, etc. A discrete wavelet transformation (DWT) based on oversampled filter banks and a discrete wavelet transformation based on two dual real wavelet trees is applied in [18]. After the good results achieved by wavelet transformation methods, in the last years there have been an increase in the studies that based on (MAP) Bayesian approach. The study [19] converts multiplicative noise to additive noise by taking logarithm. Then study continues to model the radar cross section (RCS) using heavy-tailed Rayleigh density function. Another study [20] proposes a despeckling algorithm based on undecimated wavelet decomposition and maximum a posteriori (MAP) estimation without applying a logarithmic transformation. The study at [21] proposes a wavelet shrinkage method. Bayesian wavelet shrinkage factor is derived to estimate the noise-free wavelet coefficients. To preserve the edges, an edge detector is applied to original image. There are some studies that use Markov Random Field (MRF). The proposed study at [22] suggests to use Gauss MRF for texture areas and an adaptive neighborhood system for edge preservation in the image. Another popular filtering technique is called Non Local filtering. The most popular one is called Non Local Means (NLM) [23]. NLM algorithm calculates the value of a pixel by taking the weighted average of all the pixels that are similar to the target pixel in a search window of fixed size centered around target pixel. If a pixel is more similar to the target pixel then the contribution of that pixel to the averaging process increases. There have been a lot of research about NLM in the recent years [24, 25, 26]. Another variation of NLM which is called block-matching and 3-D filtering (BM3D) is proposed in [27]. BM3D filter uses NL principle to collect similar image patches, then a wavelet decomposition is computed of the resulting 3D blocks. The similar patches will likely have a very sparse representation in the wavelet domain. This allows an effective separation of noisy and noise-free coefficients. Sparse representation is another way to despeckle images [28, 29, 30]. A learned dictionary from the noisy image used for despeckling in [31]. Nonlocal sparse models for image despeckling are studied in [32, 33, 34, 35]. Another popular algorithm used in speckle reduction is Anisotropic Diffusion (AD) [36]. Anisotropic diffusion is an extremely popular algorithm in image processing society. The aim of AD is to reduce noise in the image without losing information such as edges from the original image. Another type of AD specialized at reducing speckle at SAR images is called speckle reducing anisotropic diffusion (SRAD) [37]. Another popular denoising approach in image processing society is called ROF model or Total Variation (TV) [38]. This algorithm minimizes a cost function with a data fidelity term that have a prior to smooth the image while preserving edges. Several approaches have been proposed to overcome the weaknesses of the original TV model [39, 40, 41].



Figure 1: Noisy Barbara image on top, extracted texture map from Barbara on bottom. White represents texture and black represents homogeneous areas

In this study, advantages of two of the most popular denoising methods Total Variation (TV) [38] based Sparsity Driven Despeckling with Quadratic Linear term (SDD-QL) [42] and Nonlocal means (NLM) depended regularization term [23] are combined in a single cost function to increase the despeckling quality and preserve the texture structures in images. SDD-QL converges really fast and performs high quality despeckling on homogenous regions but deteriorates the texture areas in an image. SDD-QL is based on Sparsity Driven Despeckling approach (SDD) [43]. SDD is used to reduce the processing time and Quadratic Linear regularization term [42] is used to increase the speckle reduction quality. Although Nonlocal Means algorithm performs speckle reduction while keeping texture areas intact processing time of NL Means increases as the size of the image increases. That is why it is hard to apply it in real time applications.

The proposed method in this study is called Nonlocal Total Variation (NLTV). NLTV applies SDD-QL [42] to a pixel if that pixel contains no texture. If the pixel is classified as a texture pixel than NL Means depended regularization term is applied. A semi-automatic system is developed to extract texture information about each pixel from a given image. Entropy filtering is applied to the input image. Entropy

filtering can detect the subtle variations in a given neighborhood. The result from entropy filtering is then scaled into [0,1] range. After rescaling, binarization is applied with a certain threshold value. In order to achieve the best texture map, a certain threshold value needs to be applied. This value varies from image to image. That is why in this study, texture map extraction process is semi-automatic. The resultant binary matrix is called texture map. Texture map indicates whether a pixel is classified as a texture pixel or homogeneous pixel. The cost function of NLTV is minimized using preconditioned conjugate gradients (PCG) method.

The results obtained from SDD-QL, Nonlocal Means and NLTV by applying these algorithms on three different test images with speckle noise are compared. SSIM, PSNR and SNR values are used to determine the quality of speckle reduction. In this study actual SAR images with speckle noise are not used. Instead speckle noise is artificially added to the three classic test images named Barbara, Lena and House. All of the tests are performed on these images.

In Chapter 2, a detailed explanation, mathematical models are given for SDD-QL and Nonlocal Means method. In Chapter 3 the proposed method NLTV is explained. Mathematical model is given for NLTV. In the final chapter, the results from SDD-QL, NL Means and NLTV are compared.

2. METHODS

In this section, the explanation, mathematical models of SDD-QL and NL Means will be showed. Advantages and disadvantages are also going to be discussed.

2.1 Sparsity-Driven Despeckling with Quadratic Linear Approximated Regularization Term (SDD-QL)

Variational methods gives excellent results on different image processing tasks. One of the most popular variational method is called ROF model or Total Variation [38]. One of the approaches of total variation is called sparsity-driven despeckling (SDD) [43]. Quadratic Linear approximated SDD is called SDD-QL. SDD-QL reduce the processing time and increase the accuracy of the method [42]. The cost function of SDD-QL is given below in Equation (1):

$$J(f) = \frac{1}{2N} \sum_{p=1}^N (f_p - g_p)^2 + \lambda |(\partial f)_p| \quad (1)$$

where g is the input speckled image, f is the desired despeckled image. λ is a smoothing factor. N is the number of pixels and ∂ is the derivative operator. In the cost function data fidelity term ensures that the result image stays similar to

the input image and regularization term implies penalty on changes in image gradients in ℓ_1 -norm manner. Although ℓ_1 -norm preserves details, it is difficult to minimize it because ℓ_1 -norm is not differentiable. SDD proposed a quadratic approximation to non-differentiable ℓ_1 -norm in Equation (2).

$$|z| \simeq (|\hat{z}| + \epsilon)^{-1} z^2 \quad (2)$$

\hat{z} is a proxy constant for z and ϵ is a small constant. Author [42] also propose to use a linear approximation to improve the quadratic approximation as given in Equation (3).

$$|z| \simeq (1 - \alpha)(|\hat{z}| + \epsilon)^{-1} z^2 + \alpha \text{sgn}(\hat{z})z \quad (3)$$

$0 \leq \alpha \leq 1$, $\text{sgn}(\cdot)$ is signum function and $\text{sgn}(\hat{z})$ is the linear approximation of \hat{z} . Equation (3) is a combination of linear and quadratic approximations.

If the $|(\partial f)_p|$ defined as $|(\partial_x f)_p|$ and $|(\partial_y f)_p|$ and the cost function in Equation (1) with QL approximation can be written as follows.

$$J^{(n)}(f) = \frac{1}{2N} \sum_{p=1}^N (f_p - g_p)^2 + (f_p - \hat{f}_p)^2 + \lambda [(1 - \alpha)(w_{x,p}(\partial_x f)_p^2 + w_{y,p}(\partial_y f)_p^2) + \alpha(s_{x,p}(\partial_x f)_p + s_{y,p}(\partial_y f)_p)] \quad (4)$$

where n is the iteration number. \hat{f} is a proxy constant for f_p . $(f_p - \hat{f}_p)^2$ is a new regularization term [42]. It helps f_p to stay close to \hat{f} . Equation (4) can be written in matrix-vector form as below:

$$J^{(n)}(f) = \frac{1}{2N} \left((v_f - v_g)^T (v_f - v_g) + (v_f - v_{\hat{f}})^T (v_f - v_{\hat{f}}) + \lambda [(1 - \alpha)(v_f^T C_x^T W_x C_x v_f + v_f^T C_y^T W_y C_y v_f) + \alpha(s_x^T C_x v_f + s_y^T C_y v_f)] \right) \quad (5)$$

where v_f , v_g , $v_{\hat{f}}$, s_x , s_y are the vector forms of f_p , g_p , \hat{f}_p , $s_{x,p}$, $s_{y,p}$. W_x and W_y are the diagonal matrix form of $w_{x,p}$, $w_{y,p}$. C_x , C_y are defined as forward difference gradient operators.

Equation (5) is a convex function so it can be differentiated. One needs to take the derivative of Equation (5) with respect to v_f and equalize to zero to find its minimum. This operation leads to a linear system as follows in Equation (6).

$$Av_f^{(n+1)} = b \quad (6)$$

$A = 2I + \lambda(1 - \alpha)(C_x^T W_x C_x + C_y^T W_y C_y)$ and $b = v_g + v_f - \lambda(\alpha/2)(C_x^T s_x + C_y^T s_y)$. I is the identity matrix.

To solve the linear equation, author [42] is used preconditioned conjugate gradient (PCG) with incomplete Cholesky preconditioner (ICP). We will also be using PCG and ICP to solve the linear equation of the proposed NLTV method. The results of experiments of SDD-QL will be given in Results section.

2.2 Non Local Means (NLM)

Nonlocal means reduce the noise in a semi-local manner instead of locally. Nonlocal means calculates value of a pixel by the weighted average of all the pixels in a given search neighborhood [23]. NL Means algorithm is defined as below:

$$NL[u](x) = \frac{1}{C(x)} \int_{\Omega} e^{-\frac{(G_a * |u(x+) - u(y+)|^2)(0)}{h^2}} u(y) dy, \quad (7)$$

where $x \in \Omega$ and $C(x) = \int_{\Omega} e^{-\frac{(G_a * |u(x+) - u(y+)|^2)(0)}{h^2}} dz$ is a normalizing constant. G_a is Gaussian kernel to improve the importance of current pixel in averaging process. h is a smoothing factor. This formula is basically states that the denoised value of pixel x is equal to the weighted average of all pixels in a search window [23]. The formula at Equation (7) can be written as below:

$$NL[v](i) = \sum_{j \in I} w(i, j) v(j) \quad (8)$$

where $w(i, j)$ represents the weights that depends on the similarity between pixels i and j . $w(i, j)$ must satisfy the following conditions, $0 \leq w(i, j) \leq 1$, $\sum_j w(i, j) = 1$. Similarity between two pixels depends on the similarity between square neighborhood of fixed size and centered at each pixel. Weights given in Equation (8) can be expanded as below in Equation (9):

$$w(i, j) = \frac{1}{Z(i)} e^{-\frac{\|v(N_i) - v(N_j)\|_{2, \alpha}^2}{h^2}} \quad (9)$$

The normalizing constant $Z(i)$ is as below in Equation (10):

$$Z(i) = \sum_j e^{-\frac{\|v(N_i) - v(N_j)\|_{2, \alpha}^2}{h^2}} \quad (10)$$

Nonlocal Means compares the similarities between geometrical configuration between two neighborhood filters. This allows NL Means to have a more robust comparison rather simply comparing gray intensity levels between two pixels [23]. As the similarity between two pixel increases, $w(i, j)$ gets closer to 1. Original NL Means [23] uses Euclidean distance to measure the similarity between two pixels. In this study mean squared error (MSE) is used instead of Euclidean distance.

3. NON LOCAL TOTAL VARIATION (NLTV)

In this study, advantages of SDD-QL and NL Means are combined in a single cost function. The aim is to preserve the texture areas with NL Means but also reduce the processing time with the help of SDD-QL. The proposed cost function for NLTV is as below:

$$J(f) = \frac{1}{2N} \sum_{p=1}^N T_p (f_p - g_p)^2 + \gamma (f_p - \hat{f}_p)^2 + (1 - T_p) (f_p - h_p)^2 + \lambda T_p (w_{x,p} (\partial_x f)_p^2 + w_{y,p} (\partial_y f)_p^2) \quad (11)$$

where f is the desired despeckled image and g is the input image with speckle noise. T is the binary texture map matrix and T_p indicates whether the pixel at location p contains texture or not. The term $\gamma (f_p - \hat{f}_p)^2$ is same as SDD-QL [42]. h is the despeckled image from Nonlocal Means algorithm. The term $(f_p - h_p)^2$ ensures that the image stays closer to despeckled image from NL Means algorithm. If a pixel is a texture pixel the term $(1 - T_p) (f_p - h_p)^2$ contributes to the result. If not then the term $\lambda T_p (w_{x,p} (\partial_x f)_p^2 + w_{y,p} (\partial_y f)_p^2)$ contributes to the result. We can write the formula in Equation (11) in matrix form as below:

Table 1: Despeckling results from experiments with SDD-QL, NL Means, NLTV methods on Barbara, Lena and House images

	Barbara			House			Lena		
	PSNR	SNR	SSIM	PSNR	SNR	SSIM	PSNR	SNR	SSIM
SDD-QL	21.6602	16.796	0.65275	28.693	23.9831	0.85391	24.1707	18.5142	0.76307
NLM	23.5156	18.6514	0.72233	26.822	22.1121	0.82291	27.2257	21.5692	0.73998
NLTV	23.5391	18.675	0.74399	27.5335	22.8236	0.84233	24.3304	18.6739	0.7683

$$\begin{aligned}
 J^{(n)}(v_f) = & \frac{1}{2N} (v_f - v_g)^T T (v_f - v_g) \\
 & + \gamma (v_f - \widehat{v}_f)^T (v_f - \widehat{v}_f) \\
 & + (v_f - v_h)^T (I \\
 & - T) (v_f - v_h)^2 \\
 & + \lambda (v_f^T D_x^T W_x T D_x v_f \\
 & + v_f^T D_y^T W_y T D_y v_f)
 \end{aligned} \quad (12)$$

where $v_f, v_g, \widehat{v}_f, v_h$ are the vector forms of $f_p, g_p, \widehat{f}_p, h_p$ respectively. I is the identity matrix. T is the diagonal matrix form of the texture map T_p . W_x, W_y are the diagonal matrix form of $w_{x,p}, w_{y,p}$ respectively. D_x, D_y are the diagonal matrices for derivative operations. One needs to take the derivative of the Equation (12) with respect to v_f and equalized it to zero to find the minimum point of the cost function. After the derivative and equalization operations a linear system is achieved as below in Equation (13):

$$A v_f^{(n+1)} = b \quad (13)$$

where $A = (1 + \gamma) + \lambda(D_x^T W_x T D_x + D_y^T W_y T D_y)$ and $b = T v_g + \gamma \widehat{v}_f + (I - T) v_h$.

Because of the despeckled image from NL Means is used in the NLTV cost function, it may seem that it is arbitrary to run NLTV algorithm. In order to improve the process time of NLTV, NL Means algorithm is only applied on texture pixels and not on all the speckled pixels. By applying this method, process time of NLTV is reduced by 2.6 times compared to NL Means.

4. DISCUSSION

In this section, the advantages of using NLTV method will be showed in more detail. Both SDD-QL and NL Means method solely relies on one smoothing parameter. This restricts the freedom in both methods. Increasing smoothing parameter to a high value over-smooths the target input image. In Equation (12), only quadratic regularization term relies on the smoothing parameter λ . NL Means based regularization term does not need λ , since all the required information is already stored in h_p .



SDD-QL



NL Means



NLTV

Figure 2: Despeckling results from SDD-QL, NL Means, NLTV



Figure 3: Despeckling results for different λ and h values for a) SDD-QL, b) NL Means and c) NLTV

In Error! Reference source not found., artificially speckled Barbara image and texture map that is extracted from the original noise-free Barbara image is shown. The despeckling results for different λ and h values for SDD-QL, NL Means and NLTV are shown in Figure 3. Row a) shows the results from SDD-QL. At $\lambda=50$, texture areas are already distorted. At $\lambda = 100$ even though better despeckling result is achieved on homogeneous areas, all the texture information is lost. At $\lambda = 300$, the image is over-smoothed. On row b) the results from NL Means can be seen. At $h = 35$, a good despeckling performance is achieved with texture information is still intact. At $h = 55$ and $h = 100$ the image is over-smoothed. Even though the smoothing parameter is increased to really high levels, the texture in the image can still be seen. The

texture preserving feature of NL Means can clearly be seen in Figure 3 row b). Row c) shows the results from NLTV method. At $\lambda = 55$, a good despeckling performance is achieved. Since texture areas on the image are no longer dependent on the smoothing parameter, λ can be increased without any constraints. At $\lambda = 100$ and $\lambda = 500$ over-smoothing of the homogeneous areas and preserving of the texture areas is achieved. By combining the advantages of SDD-QL and NL Means a better texture preserving despeckling method is achieved.

Another advantage of NLTV over NL Means is the execution time of the method. By applying NL Means to only texture areas in the input image, the execution time of NLTV method is decreased.

The execution times of the methods used in this study is given

in Table 2. Fastest execution time belongs to the SDD-QL method. The comparison of execution times between NL Means and NLTV is worth noting. On Barbara image NL Means took 273.35 seconds to despeckle. NLTV on the other hand only took 104.456 seconds to despeckle the image. The difference between the execution times is caused by applying NL Means method only on texture pixels for NLTV. The execution time of the NLTV method depends on the ratio of texture pixels to homogeneous pixels. If the input image contains large portions of texture areas, NLTV gets a similar execution time to NL Means since it mostly behaves like NL Means. If input image contains large portions of homogeneous areas, then NLTV behaves like SDD-QL.

5. RESULTS AND ANALYSIS

In this section the results from the experiments of SDD-QL, NL Means and NLTV will be given. Speckle noise is artificially added to Lena, House and Barbara images with a variance of 0.07. Each algorithm is compared with each other according to experiment results by using Structure Similarity

Table 2: The execution times of the methods. (seconds)

	Barbara	House	Lena
SDD-QL	1.193	1.453	1.578
NL means	273.35	638.224	274.352
NLTV	104.456	156.067	76.383

Index (SSIM) to decide the despeckling quality.

The result according to SSIM scores for each algorithm can be seen in Table 1. Algorithm with the highest SSIM score is bold highlighted. The numerical results show that SDD-QL get a low SSIM score on Barbara image which contains a lot of texture areas. NLTV outperforms NL Means on Barbara image. Despite having a low SSIM on Barbara, SDD-QL achieves 0.85391 SSIM score on House image which is the highest score among the three algorithms. On the last test image Lena, NLTV again achieves the highest SSIM. According to results in Table 1, NLTV outperforms NL Means and SDD-QL on separate images. Processing time of NLTV is slower than SDD-QL but faster than NL Means. SDD-QL is the fastest algorithm among the three algorithms followed by NLTV then followed by NL Means.

In Figure 2, the despeckling results from SDD-QL, NL Means and NLTV can be seen. Despeckling quality is determined according to the SSIM score. So in order to select the images with the highest SSIM score, many different variations have been tried out. Among many despeckled images, the one with the highest SSIM score is selected. This is the reason why the result from SDD-QL is still appears to be speckled. SDD-QL method can achieve high smoothing performance, but during

smoothing a lot of texture information will be lost. In Figure 2, NL Means seems to achieve high despeckling performance. But if the carpet behind the Barbara is examined carefully, small artifacts caused by NL Means can be seen. NLTV on the other hand can achieve high level of smoothing in the homogeneous areas and still preserve the texture areas in the image.

6. CONCLUSION

In this study, texture preserving feature of NL Means is combined with fast SDD-QL in a single cost function. By only applying NL Means to the texture pixels and applying SDD-QL to homogeneous pixels, a better despeckling performance with texture preserving feature is achieved.

REFERENCES

- [1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek ve K. P. Papathanassiou, «A tutorial on synthetic aperture radar,» *IEEE Geoscience and Remote Sensing Magazine*, cilt 1, no. 1, pp. 6-43, 2013.
- [2] F. Argenti, A. Lapini, T. Bianchi and L. Alparone, "A Tutorial on Speckle Reduction in Synthetic Aperture Radar Images," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 3, pp. 6-35, 2013.
- [3] J.-S. Lee, «Digital Image Enhancement and Noise Filtering by Use of Local Statistics,» *IEEE Transactions on Pattern Analysis & Machine Intelligence*, cilt 2, no. 2, pp. 165-168, 1980.
- [4] V. S. Frost, J. . A. Stiles, K. S. Shanmugan ve J. . C. Holtzman, «A Model for Radar Images and Its Application to Adaptive Digital Filtering of Multiplicative Noise,» *IEEE Transactions on Pattern Analysis & Machine Intelligence*, cilt 4, no. 2, pp. 157-166, 1982.
- [5] D. T. Kuan, A. A. Sawchuk, T. C. Strand and P. Chavel, "Adaptive Noise Smoothing Filter for Images with Signal-Dependent Noise," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vols. PAMI-7, no. 2, pp. 165-177, 1985.
- [6] J.-S. Lee, "A simple speckle smoothing algorithm for synthetic aperture radar images," *IEEE Transactions on Systems, Man, and Cybernetics*, Vols. SMC-13, no. 1, pp. 85-89, 1983.
- [7] J.-S. Lee, J.-H. Wen, T. . L. Ainsworth, K.-S. Chen ve A. . J. Chen, «Improved Sigma Filter for Speckle Filtering of SAR Imagery,» *IEEE Transactions on Geoscience and Remote Sensing*, cilt 47, no. 1, pp. 202-213, 2008.
- [8] C. Tomasi ve R. Manduchi, «Bilateral filtering for gray and color images,» %1 içinde *Sixth International Conference on Computer Vision*, Bombay, 1998.
- [9] W. Zhang, F. Liu ve L. Jiao, «SAR image despeckling via bilateral filtering,» *Electronics Letters*, cilt 45, no.

- 15, pp. 781-783, 2009.
- [10] G.-T. Li, C.-L. Wang, P.-P. Huang ve W.-D. Yu, «SAR Image Despeckling Using a Space-Domain Filter With Alterable Window,» *IEEE Geoscience and Remote Sensing Letters*, cilt 10, no. 2, pp. 263-267, 2012.
- [11] R. C. Gonzales and R. E. Woods, *Digital Image Processing*, Pearson, 2007.
- [12] T. R. Crimmins, "Geometric filter for reducing speckle," in *29th Annual Technical Symposium*, San Diego, 1985.
- [13] K. Kondo, Y. Ichioka ve T. Suzuki, «Image restoration by Wiener filtering in the presence of signal-dependent noise,» *Applied Optics*, cilt 16, no. 9, pp. 2554-2558, 1977.
- [14] A. Lopes, H. Laur and E. Nezry, "Statistical Distribution And Texture In Multilook And Complex Sar Images," in *10th Annual International Symposium on Geoscience and Remote Sensing*, College Park, Maryland, 1990.
- [15] E. M. Eliason ve A. S. McEwen, «Adaptive box filters for removal of random noise from digital images,» *Photogrammetric Engineering and Remote Sensing*, cilt 56, no. 4, pp. 453-458, 1990.
- [16] S. Sarkka, *Bayesian filtering and smoothing*, Cambridge University Press, 2013.
- [17] L. Gagnon, "Speckle filtering of SAR images: a comparative study between complex-wavelet-based and standard filters," in *Optical Science, Engineering and Instrumentation '97*, San Diego, 1997.
- [18] R. Sveinsson ve J. A. Benediktsson, «Almost translation invariant wavelet transformations for speckle reduction of SAR images,» *IEEE Transactions on Geoscience and Remote Sensing*, cilt 41, no. 10, pp. 2404 - 2408, 2003.
- [19] A. Achim, E. E. Kuruoglu ve J. Zerubia, «SAR image filtering based on the heavy-tailed Rayleigh model,» *IEEE Transactions on Image Processing*, cilt 15, no. 9, pp. 2686 - 2693, 2006.
- [20] F. Argenti, T. Bianchi ve L. Alparone, «Multiresolution MAP Despeckling of SAR Images Based on Locally Adaptive Generalized Gaussian pdf Modeling,» *IEEE Transactions on Image Processing*, cilt 15, no. 11, pp. 3385 - 3399, 2006.
- [21] M. Dai, C. Peng, A. K. Chan ve D. Loguinov, «Bayesian wavelet shrinkage with edge detection for SAR image despeckling,» *IEEE Transactions on Geoscience and Remote Sensing*, cilt 42, no. 8, pp. 1642 - 1648, 2004.
- [22] M. Walessa ve M. Datcu, «Model-based despeckling and information extraction from SAR images,» *IEEE Transactions on Geoscience and Remote Sensing*, cilt 38, no. 5, pp. 2258 - 2269, 2000.
- [23] A. Buades, B. Coll ve J.-M. Morel, «A non-local algorithm for image denoising,» %1 içinde *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, 2005.
- [24] T. Brox ve D. Cremers, «Iterated Nonlocal Means for Texture Restoration,» %1 içinde *Scale Space and Variational Methods in Computer Vision. SSVM 2007 Lecture Notes in Computer Science*, vol 4485. , Springer, Berlin, Heidelberg, 2007, pp. 13-24.
- [25] T. Brox, O. Kleinschmidt ve D. Cremers, «Efficient Nonlocal Means for Denoising of Textural Patterns,» *IEEE Transactions on Image Processing* , cilt 17, no. 7, pp. 1083 - 1092, 2008.
- [26] A. Buades, B. Coll ve J. M. Morel, «Nonlocal Image and Movie Denoising,» *International Journal of Computer Vision*, cilt 76, no. 2, p. 123-139, 2008.
- [27] K. Dabov, A. Foi, V. Katkovnik ve K. Egiazarian, «Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering,» *IEEE Transactions on Image Processing*, cilt 16, no. 8, pp. 2080 - 2095, 2007.
- [28] M. Elad ve M. Aharon, «Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries,» *IEEE Transactions on Image Processing*, cilt 15, no. 12, pp. 3736 - 3745, 2006.
- [29] J. Mairal, M. Elad ve G. Sapiro, «Sparse Representation for Color Image Restoration,» *IEEE Transactions on Image Processing*, cilt 17, no. 1, pp. 53 - 69, 2007.
- [30] J. Mairal, F. Bach, J. Ponce, G. Sapiro and A. Zisserman, "Non-local sparse models for image restoration," in *2009 IEEE 12th International Conference on Computer Vision*, Kyoto, 2009.
- [31] S. Foucher, "SAR Image Filtering Via Learned Dictionaries and Sparse Representations," in *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*, Boston, 2008.
- [32] J. Jiang, L. Jiang and N. Sang, "Non-local sparse models for SAR image despeckling," in *2012 International Conference on Computer Vision in Remote Sensing*, Xiamen, 2012.
- [33] M. Amirmazlaghani ve H. Amindavar, «A Novel Sparse Method for Despeckling SAR Images,» *IEEE Transactions on Geoscience and Remote Sensing*, cilt 50, no. 12, pp. 5024 - 5032, 2012.
- [34] X. Zhang, T. Bai, H. Meng ve J. Chen, «Compressive Sensing-Based ISAR Imaging via the Combination of the Sparsity and Nonlocal Total Variation,» *IEEE Geoscience and Remote Sensing Letters*, cilt 11, no. 5, pp. 990 - 994, 2013.
- [35] B. Xu, Y. Cui, Z. Li ve J. Yang, «An Iterative SAR Image Filtering Method Using Nonlocal Sparse Model,» *IEEE Geoscience and Remote Sensing Letters* , cilt 12, no. 8, pp. 1635 - 1639, 2015.
- [36] A. Perona ve J. Malik, «Scale-space and edge detection using anisotropic diffusion,» *IEEE Transactions on Pattern Analysis and Machine Intelligence*, cilt 12, no. 7, pp. 629 - 639, 1990.
- [37] Y. Yu ve S. T. Acton, «Speckle reducing anisotropic diffusion,» *IEEE Transactions on Image Processing*, cilt 11, no. 11, pp. 1260 - 1270, 2002.
- [38] L. I. Rudin, S. Osher ve E. Fatemi, «Nonlinear total

variation based noise removal algorithms,» *Physica D: Nonlinear Phenomena*, cilt 60, no. 1-4, pp. 259-268, 1992.

- [39] A. Chambolle ve P. L. Lions, «Image recovery via total variation minimization and related problems,» *Numerische Mathematik*, cilt 76, no. 2, pp. 167-188, 1997.
- [40] S. Teboul, L. Blanc-Feraud, G. Aubert ve M. Barlaud, «Variational approach for edge-preserving regularization using coupled PDEs,» *IEEE Transactions on Image Processing*, cilt 7, no. 3, pp. 387 - 397, 1998.
- [41] A. Chambolle, V. Caselles, D. Cremers, M. Novaga ve T. Pock, «An introduction to total variation for image analysis,» *In: Theoretical Foundations and Numerical Methods for Sparse Recovery*, pp. 263-340, 2010.
- [42] F. Nar, «SAR image despeckling using quadratic-linear approximated ℓ_1 -norm,» *Electronics Letters*, cilt 54, no. 6, pp. 387 - 389, 2018.
- [43] C. Özcan, B. Sen ve F. Nar, «Sparsity-Driven Despeckling for SAR Images,» *IEEE Geoscience and Remote Sensing Letters*, cilt 13, no. 1, pp. 115-119, 2015.

Real time activity recognition using weighted dynamic time warping

Rafet DURGUT¹ and Oğuz FINDIK²

¹ Karabuk University, Karabuk/Turkey, rafetdurgut@karabuk.edu.tr

² Karabuk University, Karabuk/Turkey, oguzfindik@karabuk.edu.tr

Abstract: In this study, real-time action recognition system has been developed using Microsoft Kinect v2 depth sensor. Key frames on local joint motion trajectory are used as discriminative feature vector in the developed system. Using the Weighted Dynamic Time Warping method, the cost value between the key frames of reference action and the key frames of the test action is obtained. The cost value is used to perform classification. The success of the method has been tested on the data set. In this system, there is no time constant for reference and test actions. Furthermore, the system no need to an extra posture for start and finish.

Keywords: Real time activity recognition, weighted dynamic time warping, key frames on local joint motion trajectory

I. INTRODUCTION

HUMAN computer interaction (HCI) is a discipline based on studies that discuss the understanding and classification human behaviors using computer technology. It basically aims to positively contribute to human life. It has many work areas such as gesture recognition, human motion recognition, hand motion recognition etc. There are various application areas for each of these. Interpreting and making inferences from human behaviors can be vital for many applications. Computer technologies cover many hardware and devices such as personal computers, smart phones, microprocessors etc. Control or analysis of the equipment based on human demands and behaviors have made them a key component for HCI.

Human activity recognition (HAR) is one of the study areas of HCI. It includes studies on recognition and classification of human activities with the help of equipment. Studies conducted on this area include important subjects such as making the things easier for people's daily lives, keeping them away from dangers and monitoring their developmental processes. Studies conducted on HAR can be basically divided into two groups. The first one of these focuses on studies on recognition and classification and the second one focuses on studies on application and analysis. According to the several studies conducted on recognition and classification, applicability and efficiency of application and analysis practices have increased.

Real time depth sensing systems have an important place in HAR applications. They also improve the quality of work because of the high resolution and high fps they

offer to the users. Kinect v2, a depth sensing system, processes depth map to obtain 60 frames skeleton information in one second [1]. Within each frame, there are (X, Y, Z) positions on the 3D coordinate plane of the 25 joints on human body. These 25 joints of the human body can be used in HAR tasks. HAR algorithms do recognition and classification using these joint points.

Human body representation is one of the important steps in activity recognition and classification applications. This representation determines the features to be used. Using the joint location streamed by the hardware directly causes some problems. For example; pre-processing is necessary for people with different heights who perform the motion and for human postures at different positions. For this reason, there are studies on the determination of distinguishing features in activity recognition applications. There are studies in the literature that use 3D locations of joints and location changes [2-8]. The angles the joints have with each other or with reference points are also used in motion recognition applications [9-14]. Joint motion and velocity can be pointed out as another feature for applications used for motion recognition [15-18].

Another important criterion for real-time applications is time warps. Dynamic Time Warping (DTW) is one of the time-warp-sensitive synchronization methods. There are many studies in this area in the literature [19-22]. When the classic DTW is used, the effect of all body joints on the result is the same. Instead, weighted dynamic time warping (WDTW) has emerged by weighting each joint and contributing to the result in proportion with its weight. In the study conducted, WDTW has been shown to perform a more accurate classification than classical DTW [23]. Activity recognition can work with high accuracy for offline applications when WDTW is used on joint changes. However, it is not sufficient in terms of the period of the study for real-time applications. Therefore, it is necessary to use distinguishing features that can be represented with less data. In this study, key frames features on joint motion trajectory were used together with WDTW method. Furthermore, a real time activity recognition application was developed.

II. KEY FRAMES ON LOCAL JOINT MOTION TRAJECTORY

Joints location changes during a human activity. Activity recognition applications utilize various measures

of these joints. During activity recognition, one-to-one changes of joints can be controlled while comparing the motion performed by the user with the reference motion performed by the expert. In this case, the number of the comparisons will be the same as the number of frames for each joint. Instead of comparing for each frame, this can be done using key frames on Local Motion Joint Trajectory (LJMT), for relatively significant frames. As the comparison made using these key frames will enable comparing with fewer data on fewer frames, it can work with higher performance compared to frame-by-frame comparison. One of the most important criteria for real-time applications is the speed of operation.

LJMT is another distinguishing feature for skeleton-based activity recognition applications. The key components of the activity can be represented by key frames on LJMT. With key frames, the number of elements in the data set that make up the activity can also be significantly reduced. Through this advantage, it can be used for real time systems.

The most basic features of an activity are body joint coordinates. Other features used for activity recognition are obtained from this information. F^p gives the body joint coordinates. t gives the frame number and n gives the total number of frames of the motion. $P_{i,t}^x$ gives the position of the i equation on x coordinate at time t . F^d gives the value of the change between the previous frame and the next frame of a joint.

$$F^p(t_0:t_n) = \langle P_{i,t}^x, P_{i,t}^y, P_{i,t}^z \rangle \quad (1)$$

$$F^d = F^p(t_{i+1}) - F^p(t_i) \quad (2)$$

Key frames are obtained using F^d time series. If the total resultant change exceeds a certain threshold value, that frame is marked as key frame. a gives the starting moment of the lower joint change, and b gives the ending point of the lower joint change. In that case;

$$F_{i,t}^{bp} = \sum_{j=a}^b F_{i,j}^d \quad (3)$$

$$F_{i,t}^{bpp} = \begin{cases} t, & F_{i,t}^{bp} > F^T \\ 0, & otherwise \end{cases} \quad (4)$$

$F_{i,t}^{bpp}$ gives the key frames value for the joint i . $F_{i,t}^{bp}$ gives the size value in keyframes. Joints can be compared based on these two features. $F_{i,t}^{bp}$ gives the total differentiation and $F_{i,t}^{bpp}$ gives the time for key frame. F^T gives the threshold value where the real change amount is controlled. If the size of total change is not sufficient, then the frame in question is not defined as key frame. Key frames for one joint reference and test motion (x, y, z coordinates) are shown in Figure 1.

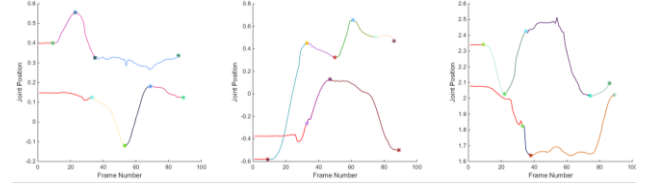


Fig 1. Detecting Key frames on Joint Motion Trajectory

III. DYNAMIC TIME WARPING

DTW is a time-warp-sensitive algorithm to can find the best match between the given reference and the test sequences. The distance between each element in the reference sequence and each element in the test sequence calculated by (5).

$$d(r_i, t_i) = \text{Dist}(r_i, t_i) \quad (5)$$

The comparison is made on the D matrix of dimension $n \times n$, which is the distance matrix in (6).

$$D = \sum_{i=1}^n d(r_i, t_i) \quad (6)$$

r reference sequence is also used as test sequence t . $d(r_i, t_i)$ gives the distance between two elements. Euclidean distance can be used. A minimum path length between sequences r and t is calculated using Bellman's principle in total cost computation [24].

$$D_{min}(r_i, t_i) = \min \begin{cases} D_{min}(r_{i-1}, t_{i-1}) \\ D_{min}(r_{i-1}, t_i) \\ D_{min}(r_i, t_{i-1}) \end{cases} + d(r_i, t_i) \quad (7)$$

When the test sequence is compared with DTW with all reference sequences, the minimum cost value of $D_{min}(r_n, t_n)$ is obtained for each series. This process is repeated depending on the number of features in the activity recognition applications. Then, the comparison is made for each feature and the total cost is obtained. Since the coefficients of each feature are equal to each other in this process, their effects on results are also equal. In practice, the changes in activity of each body joint are not equal. Weighted DTW was developed on this basis [23].

IV. WEIGHTED DYNAMIC TIME WARPING

WDTW is an algorithm that is used for time-warp-sensitive matching of multidimensional sequences. It aims to solve the problem in which all dimensions in classical DTW have equal significance. Each feature vector has its own weight value. This value can be calculated by various methods.

$$W_j^g = \frac{1 - e^{-\beta D_j^g}}{\sum_k 1 - e^{-\beta D_k^g}} \quad (8)$$

W_j^g gives the weight g of joint j inside the activity class. D_j^g gives the total change of the joint. β is the model

parameter used for setting up the weights. Simpler calculations can be used if one wishes to reduce the number of parameters.

$$W_j^g = \frac{D_j^g}{\sum_k D_k^g} \quad (9)$$

It is necessary to update the new distance function $d(r_i, t_i)$ to be made using the weight values. The total cost must be calculated for each gesture. The test motion is assigned to the gesture class which has the minimum cost.

$$d^g(r_i, t_i) = \text{Dist}^g(r_i, t_i) * W_j^g \quad (10)$$

V. REAL TIME ACTIVITY RECOGNITION

As shown in Fig. 2, the proposed system consists of 3 blocks, namely, Hardware Classification and Reference

blocks. Each block has a basic function. In the hardware module, the skeletal joint information of the user performing its activity in front of the hardware. The reference block contains information about the previously recorded activities. The key frames information of the activities recorded in the Classification block is presented. In the Classification block, comparison and classification are performed based on the information obtained from the other two blocks. This block provides user with performance data such as the name of the activity performed, the similarity value and erroneous body joint information. Hardware and Classification modules run in real time. The reference block runs only when the reference database is updated.

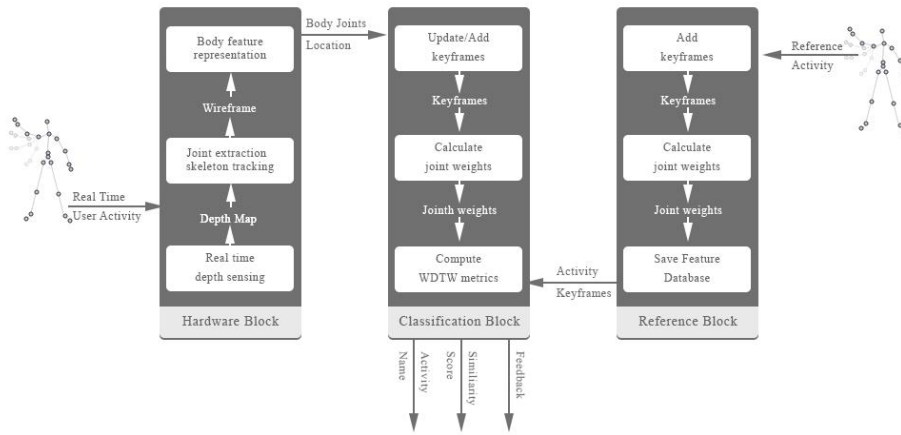


Fig 2. System overview

Hardware block is the section where the user first interacts with the system. In this section, the RGBD image of the environment and the user's skeleton wireframe image are transferred to the computer environment. Data is continuously streamed through the skeleton tracking algorithm (STA) in the hardware. The skeleton image of the person is displayed on the computer screen using the data received. In this study, Microsoft Kinect v2 was used as the hardware sensor. The test environment is shown in Fig 3.



Fig 3. Test environment

The reference block is the section where the motions performed by the expert are recorded and the key frames of these motions are determined. The key frames calculated for each joint that the Classification block needs are produced in this section. The reference joint weights required for WDTW are also calculated in this section.

In the classification block, the key frames that dynamically generated are compared with the key frames

that derived from the reference block. The comparison is made in a time-warp-sensitive manner. During comparison, reference joint weights and test joint weights are considered as Eq.9. When the recognized motion is performed, the user is notified. An example recognition

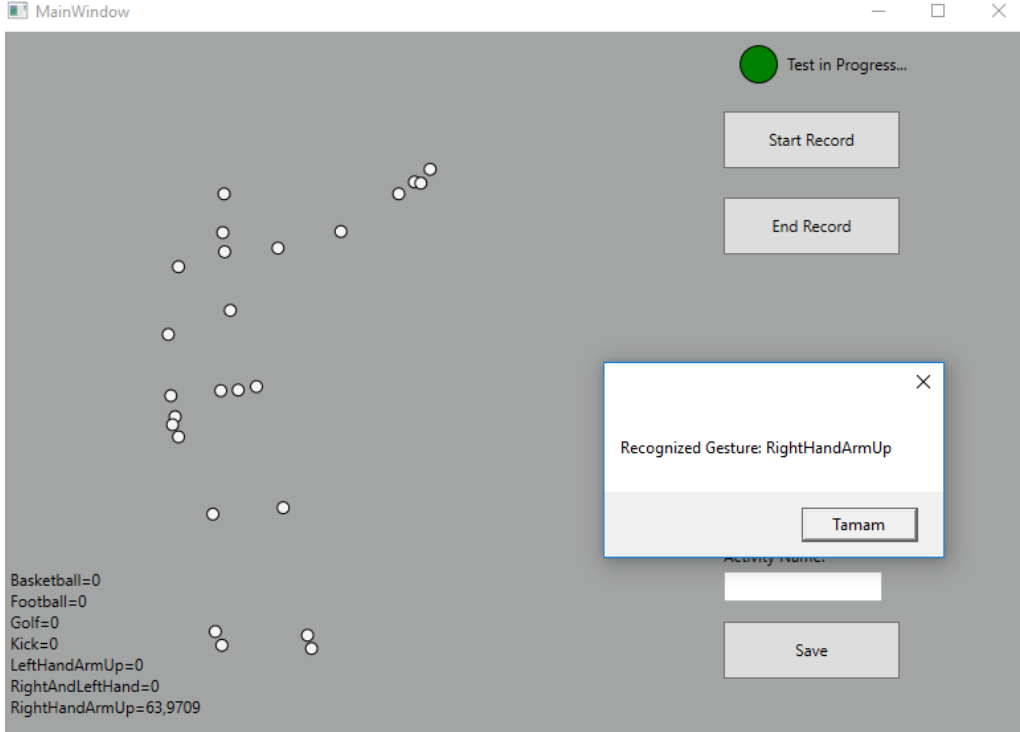


Fig 4. Application Interface

VI. RESULTS

In order to determine the accuracy of the developed system, it has to be run on available data sets. For this purpose, the accuracy rate of the method used on the datasets is measured. There are 8 different motions in the data set used to test the system's accuracy and 28 time series for each motion [22]. The confusion matrix achieved by the developed method over this motion data set is given in Table 1. The success rates of the study using the methods in the literature with the same data set are given in Table 2.

Table 1. Confusion matrix

	1	2	3	4	5	6	7	8
1	100	0	0	0	0	0	0	0
2	0	100	0	0	0	0	0	0
3	0	0	100	0	0	0	0	0
4	0	0	0	100	0	0	0	0
5	0	0	0	0	100	0	0	0
6	0	0	0	0	0	100	0	0
7	0	0	0	0	0	0	100	0
8	0	0	0	0	0	0	0	100

moment is shown in Fig. 4. Sensitivity can be determined according to the activity set to be used. The name and the similarity rate of the recognized motion and the erroneous joints are detected as a result of the algorithm.

Table 2. Success rates

Method	Success
Classical DTW	84.41% [22]
State Of Art	86.56%
Weighted DTW	97.13% [22]
WDTW With Key frames	100%

VII. CONCLUSION

The system, which is developed based on the results obtained, has a smaller size and a distinctive feature vector. Thanks to this feature, it runs in real time. It is robust against time warps according to the comparison algorithm. It can work on people with different body sizes. The motions in the reference database can be of different duration. During real-time operation, it does not need an external command or posture for the start and end of activities. Thanks to these features, it can be used in real-time activity recognition applications.

Acknowledgment

This study was supported by the Scientific Research Coordination Unit of Karabuk University under Grant KBÜ-BAP-16/1-DR-170

REFERENCES

- [1] "Microsoft Kinect", <https://developer.microsoft.com/en-us/windows/kinect/develop> ,(Access Date: 12/08/2016)
- [2] L. Xia, C.C. Chen, J. Aggarwal, View invariant human action recognition using histograms of 3d joints, *CVPRW, IEEE* (2012) 20–27.
- [3] X. Yang, C. Zhang, Y. Tian, Recognizing actions using depth motion maps- based histograms of oriented gradients, 2012.
- [4] X. Yang, Y. Tian, Eigenjoints-based action recognition using naïve-bayes- nearest-neighbor, in: *CVPRW, IEEE*, 2012, pp. 14–19.
- [5] A. Yao, J. Gall, G. Fanelli, L. Van Gool, Does human action recognition benefit from pose estimation?, in: *BMVC*, 2011.
- [6] Rosa-pujaz, A., Barbancho, I., Tard, L. J., & Barbancho, A. M. (2016). Fast-gesture recognition and classification using Kinect : an application for a virtual reality drumkit, 8137–8164. <http://doi.org/10.1007/s11042-015-2729-8>
- [7] Wu, M., Chen, T., Chen, K., & Fu, L. (2016). Daily Activity Recognition Using the Informative Features from Skeletal and Depth Data, 1628–1633.
- [8] Cippitelli, E., Gasparrini, S., Gambi, E., & Spinsante, S. (2016). A Human Activity Recognition System Using Skeleton Data from RGBD Sensors. *Computational Intelligence and Neuroscience*, 2016. <http://doi.org/10.1155/2016/4351435>
- [9] M.Z. Uddin, N.D. Thang, J.T. Kim, T.S. Kim, Human activity recognition using body joint-angle features and hidden markov model, *ETRI J.* 33 (2011) 569– 579.
- [10] V. Bloom, D. Makris, V. Argyriou, G3d: a gaming action dataset and real time action recognition evaluation framework, in: *CVPRW*, 2012, pp. 7–12.
- [11] J. Sung, C. Ponce, B. Selman, A. Saxena, Human activity detection from RGBD images. *PAIR*, 2011.
- [12] Raptis, M., Kirovski, D., & Hoppe, H. (2011). Real-Time Classification of Dance Gestures from Skeleton Animation. *ACMSIGGRAPH Symposium on Computer & Animation*.
- [13] Alwani, A., Salih, A., & Youssef, C. (2016). recognition using modified spherical harmonics, 0, 1–10. <http://doi.org/10.1016/j.patrec.2016.05.032>
- [14] Kumar, A., Kumar, A., Kumar Singh, S., & Kala, R. (2016). Human Activity Recognition in Real-Times Environments using Skeleton Joints. *International Journal of Interactive Multimedia and Artificial Intelligence*, 3(7), 61. <http://doi.org/10.9781/ijimai.2016.379>
- [15] K. Yun, J. Honorio, D. Chattopadhyay, T.L. Berg, D. Samaras, Two-person interaction detection using body-pose features and multiple instance learning, *CVPRW, IEEE* (2012) 28–35.
- [16] S.Z. Masood, C. Ellis, A. Nagaraja, M.F. Tappen, J. LaViola, R. Sukthankar, Measuring and reducing observational latency when recognizing actions, *ICCV Workshops, IEEE* (2011) 422–429.
- [17] Kim, H., Lee, S., Kim, Y., Lee, S., Lee, D., Ju, J., & Myung, H. (2016). Weighted joint-based human behavior recognition algorithm using only depth information for low-cost intelligent video-surveillance system. *Expert Systems with Applications*, 45, 131–141. <http://doi.org/10.1016/j.eswa.2015.09.035>
- [18] Chen, X., & Koskela, M. (2015). Skeleton-based action recognition with extreme learning machines. *Neurocomputing*, 149(Part A), 387–396. <http://doi.org/10.1016/j.neucom.2013.10.046>
- [19] Sempena, S., Maulidevi, N. U., Sc, M., & Aryan, P. R. (2011). Human Action Recognition Using Dynamic Time Warping, (July).
- [20] Rekha, J., Bhattacharya, J., and Majumder, S. (2011). Shape, texture and local movement hand gesture features for indian sign language recognition. In *Trendz in Information Sciences and Computing (TISC)*, 2011 3rd International Conference on, pages 30 –35.
- [21] Reyes, M., Dominguez, G., and Escalera, S. (2011). Feature weighting in dynamic time warping for gesture recognition in depth data. In *Computer Vision Workshops (ICCV Workshops)*, 2011 IEEE International Conference on, pages 1182 –1188.
- [22] Wenjun, T., Chengdong, W., Shuying, Z., and Li, J. (2010). Dynamic hand gesture recognition using motion trajectories and key frames. In *Advanced Computer Control (ICACC)*, 2010 2nd International Conference on, volume 3, pages 163 –167.
- [23] Celebi, S., Aydin, A. S., Temiz, T. T., & Arici, T. (2013, February). Gesture Recognition using Skeleton Data with Weighted Dynamic Time Warping. In *VISAPP* (1) (pp. 620-625).
- [24] Bellman, R. (1954). The theory of dynamic programming. *Bull. Amer. Math. Soc*, 60(6):503–515.

Remote Controlled Vehicle for Surface and Underground Object Detection

Bora UĞURLU¹, Utku BAYRAM¹, Vildan BAYRAM¹, İsmail KADAYIF¹

¹ Canakkale Onsekiz Mart University, Department of Computer Engineering, boraugurlu@comu.edu.tr

¹ Canakkale Onsekiz Mart University, Department of Computer Engineering, ubayram@comu.edu.tr

¹ Canakkale Onsekiz Mart University, Department of Computer Engineering, vildan@comu.edu.tr

¹ Canakkale Onsekiz Mart University, Department of Computer Engineering, kadayif@comu.edu.tr

Abstract - GPR- Ground Penetrating Radar is a tool which is used for monitoring shallow zones of a surface. It is possible to monitor the underlying surface without any excavation by examining radar waves. With that tool, it is easy to detect the buried water pipes, rebars, landmines, and historical buildings. In this study, we have developed a remotely controlled vehicle which can serve as a GPR. This vehicle is controlled by an operator. It sends the collected data to the user via a wireless network. Our vehicle contains different kind of sensors, so researchers from various disciplines can use it to effectively conduct their research.

Keywords – Mobile robot, M2M, Object detection, Wireless.

I. INTRODUCTION AND BACKGROUND

Robots are a machine that is controlled by a computer to automate one or more tasks. These machines are not only controlled externally, but also internally by integrating microcontrollers into it. The first digital and programmable robot was invented by George Devol in 1954. Its name was Unimate [1].

Based on their usage and functionality, there are many kind of robots such as, robotic manipulator, wheeled mobile robots, underwater robots, flying robots, robot vision, artificial intelligence and industrial automation [2].

Wheeled mobile robots can move around freely and do their job in different physical environments. AGV- Automatic guided robot is a good example of this. These robots are intensely used in industry and military fields for numerous purposes. For example, they can perform cleaning and maintenance in hazardous and risky industry and military fields, as well as search victims in harsh conditions such as an earthquake and flood. Sensors on these robots provide a wide range of data to be processed [3].

Rapid technology developments have resulted in many consequences in our daily lives. Human beings have many interactions between many smart devices. The internet, which is also known networks of networks, plays a crucial role. Many applications are used as a service on the internet. Cloud computing is a new shape of internet technologies. In this concept, distributed computers put their processing power and other resources together to serve to other devices [4]. There are

many services for the user specific needs such as Amazon Web Services, Google Map and so on. To operate a mobile robot, there is no need to be at the same place with it. Wireless networks and/or Bluetooth allow us to control them remotely. The only limitation is the range of physical distance. To overcome this problem, services on the cloud are used. Another concept that is related to this is known as M2M- Machine-to-Machine communication. In this concept, more than one device can transfer data between them without user intervention.

Web based and cloud based techniques are used to control robots remotely [5]. MQTT- Message Queue Telemetry Transport protocol is used to control the robot via the internet [6]. This protocol reduces the transmission time. It is also possible to control a sensor network with this protocol [7].

Mobile robots have been developed to be used in many different fields, for example collecting data from challenging terrain conditions. Also, Curiosity, which was sent to Mars by NASA in 2011, is one of the many examples of mobile robots [8]. Temperature, moisture, pressure and much more data can be collected easily by mounting necessary sensors onto a mobile robot. GPR is also one of the devices that can be mounted.

Water pipes can be detected by a GPR. It is only specialized for underground water pipes, so its mounted device is called as a PPR-Pipe Penetrating Radar [9]. In this manner, detecting complicated underground water pipe installations is easy and less time consuming.

Another application of a GPR is using the device in dangerous places. Landmine territories are typical places of that kind. GPR mounted and remotely controlled vehicles can detect landmines. However, this kind of task is very dangerous for humans. For this purpose, Fukuda and Yabushita et al. developed a system [10-11]. To improve the efficiency of open coal mine, GPR mounted vehicle is developed and used by Ralston and Strange[12].

As described above, there are many different GPR mounted vehicle for different purposes. In this study, we have developed a mobile robot for detecting underground objects. The rest of the paper organized as follows. The next section presents the architecture of our system. The implementation issues are discussed in Section III. Section IV concludes our study.

II. ARCHITECTURE

Our mobile robot has four main components: a personal computer, a wireless network, an Xbee modem and a mobile vehicle. The general architectural view of our system is given Figure 1.

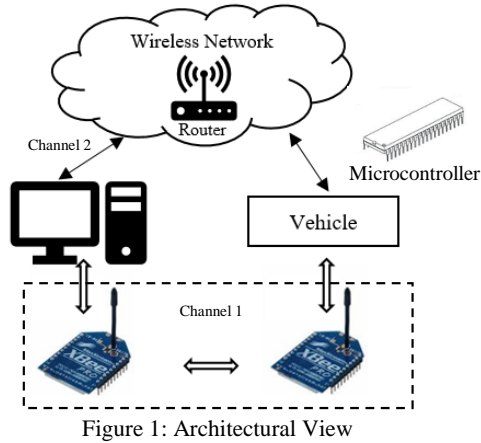


Figure 1: Architectural View

The personal computer is used to communicate with the microcontroller, which controls movements of the mobile vehicle. The computer has two communication channels. While the first one communicates with microcontroller and Xbee modem, the second one receives data collected by sensors over the wireless network. In our setup, the data collected from various sensors and the GPR are transmitted to the computer over the wireless network rather than the cloud. The vehicle has also two 360-degree cameras, and its own batteries. These batteries can be charged with solar energy. The vehicle itself is shown in Figure 2.



Figure 2: Our vehicle with a GPR installed

Received images from cameras can be seen in Figure 3. These images are transferred into the computer via router.



Figure 3: Image from vehicle camera

III. IMPLEMENTATION

The overview of our hardware design is depicted in Figure 4. It contains five main components: a micro controller, a RS-232 serial communication unit with a Wi-Fi modem, a steering wheel motor and a differential motor driver unit.

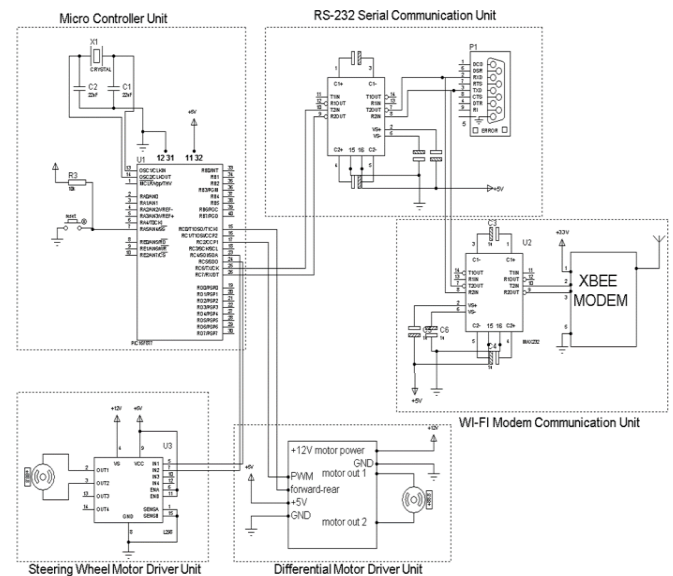


Figure 4: Hardware Design

A program code can be uploaded from the computer to the microcontroller over the Xbee modem. To do so, we use an RS-232 serial communication unit as an interface between the microcontroller and the Xbee modem as shown in Figure 4. The microcontroller controls the motor drivers and collects the sensor data and then makes an evaluation. PIC16F877 is chosen for this purpose. It has 33 input-out pins and can use PWM modulation. It is capable of serial commutation and can send/receive digital and analog signals.

Our Wi-Fi modem communication unit has an Xbee modem, which uses the IEEE 802.15.4 standard with 2.4 Ghz. Its operational voltage and current values are 3.3V and 50mA, respectively. Thus, our vehicle can move around without losing the network connection as far as it stays with this range. Before the communication between the vehicle and computer starts, the modems need to be programmed. The speed of the Xbee modem at computer side is 9600 bps and XB-24B Zigbee Coordinator as seen Figure 5.

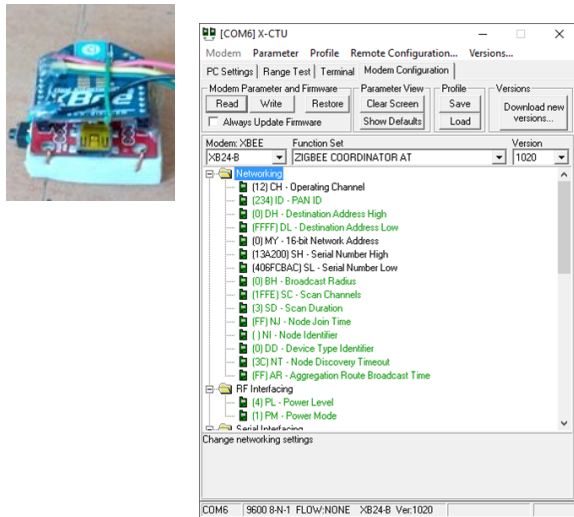


Figure 5: Xbee modem configuration

To control our vehicle, we need another control unit, which is called as a steering wheel motor unit. This unit can be controlled both manually and automatically. A linear motor controlled by the microcontroller help the vehicle direction. Load capacity of the motor is 400N. This value is acceptable for determining the front wheels' positions. The maximum current rate on load is 1.25A.

A differential motor generates wheel movement. This motor requires 24V and 750W power. Although maximum current rate on load is 31A, there is only 7A on average. Motor driver is not only capable of meeting these requirements, but also has PWM support.

In our system, the differential motor is the one which consumes the largest amount of energy. To satisfy this need, two 12V gel batteries are used. The gel batteries do not harm itself when discharged completely. These batteries are connected in parallel. These batteries are charged by a solar panel. This panel supplies 40W. As a result, the duty time is doubled, which is approximately 6 hours.

IV. CONCLUSION

In the literature, there are many implementations of mobile robots that have various sensors mounted. Bluetooth and wireless networks are widely used approaches for controlling them. These networks have range constraints. Cloud services

and protocols can be used to overcome these constraints.

In this study, we have designed a mobile vehicle for detecting both ground and underground objects. It can only be controlled remotely and take sample pictures through its camera system. It has a GPR to detect underground objects.

When GPRs in market are considered, they mostly need an operator physical help. This makes researchers' life more complicated. With our designed system, this concern can be eliminated largely. In this way, researchers focus on other issues of their studies.

In the beginning, our vehicle was designed for geosciences, but it can be easily adapted to other disciplines by adding different type of sensors to it based on needs of the study. Our design has two main limitations. The first is range limitation in term of range. The vehicle can move only within the Xbee modem range, which is approximately 2 km. Inconvenient physical factors can reduce this range easily. The second is router range while transmitting the data. When these limitations are considered, wireless solutions may not be convenient for some different disciplines.

This work was supported by Çanakkale Onsekiz Mart University the Scientific Research Coordination Unit, Project number: 1412.

REFERENCES

- [1] Malone B. (26 Sep 2011). George Devol. A Life Devoted to Invention, and Robots. *IEEE Spectrum*, Available: <https://spectrum.ieee.org/automaton/robotics/industrial-robots/george-devol-a-life-devoted-to-invention-and-robots>.
- [2] *Robotics: Engineering Series*, Appin Knowledge Solutions, Infinity Science Press, 2007.
- [3] Du, Z., He, L., Chen, Y. 2017. "Robot Cloud: Bridging the power of robotics and cloud computing", *Future Generation Computer Systems*, 74, pp.337-348.
- [4] Microsoft, IBM. (17 Feb 2019). Available: <https://azure.microsoft.com/en-in/overview/what-is-cloud-computing/>, <https://www.ibm.com/cloud-computing/learn-more/what-is-cloud-computing/>.
- [5] Sundaram, A., Gupta, M., Rathod, V. 2015. "Remote Surveillance Robot System-A Robust Framework using Cloud", *Nanoelectronic and Information Systems (iNIS) IEEE International Symposium on*, Indore, India, pp.213-218.
- [6] Aroon, N. 2016. "Study of using MQTT Cloud Platform for Remotely Control Robot and GPS Tracking", *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Chiang Mai, Thailand.
- [7] Ahmed, S., Topalov, A., Shakev, N. 2017. "A Robotized Wireless Sensor Network Based on MQTT Cloud Computing", *International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM)*, Donastia-San Sebastian, Spain.
- [8] NASA. (23 May 2011). Available: https://www.nasa.gov/mission_pages/msl/multimedia/gallery/pia14156.html.
- [9] Ékes, C., Neducza, B. 2012. "Robot Mounted GPR for Pipe Inspection", *14th International Conference on Ground Penetrating Radar, (GPR)*, June 4-8, Shanghai, China.
- [10] Yobushita, H., Kanehama, M., Hirata, Y., Kasuge, K. 2005. "3D Ground Adaptive Synthetic Aperture Radar for Landmine Detection", *International Conference on Intelligent Robots and Systems (IROS)*, 2-6 Aug., Edmonton, Alta, Canada.

- [11] Prado, J., Marques, L. 2015. "Multi-sensor and multi-platform data fusion for buried objects detection and localization", *International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, 8-10 April, Vila Real, Portugal.
- [12] Ralston, J.C., Strange, A.D. 2016. "3D Robotic Imaging of Coal Seams using Ground Penetrating Radar Technology", *International Conference on Ground Penetrating Radar (GPR)*, 13-16 June, Hong Kong, China.

Estimation Relationship Between Electricity Consumption And Urban Area From Night Satellite Imagery: A Case Study For Istanbul

E. Yücer¹, İ. Kahraman², İ. R. Karas³

¹ Karabük University, Karabük / Turkey, emreyucer@karabuk.edu.tr

² Karabük University, Karabük / Turkey, idriskahraman@karabuk.edu.tr

³ Karabük University, Karabük / Turkey, ismail.karas@karabuk.edu.tr

Abstract - Satellite imagery of night-time lights provided by the us air force defense meteorological satellite program (dmisp), using the operational linescan system (ols), has been used to estimate the spatial distribution of electricity consumption throughout istanbul. There are very high correlation between state electricity consumption and night-time lights. it can be also extract urban area from night time satellite image. Population and urban area have high correlation with night lights, and also change in population and urban area will be estimated from images between 2002-2013.

Keywords: Night Satellite Imagery, Classification, Correlation

I.INTRODUCTION

Urbanization in the world is increasing rapidly and according to data of United Nations Population Organization, it is estimated that by 2050 the world population will be 12 billion and 80% of this population will live in cities. The rapid increase in population is mainly due to the large metropolitan areas such as Istanbul, Ankara, Izmir, Kocaeli, Bursa etc. such as industrialization and job opportunities are concentrated towards the provinces. After allowing immigrants the electricity consumption intensifies on these cities.

The wave of migration towards metropolitan cities causes the physical structure of the city to change as well as the population. Unplanned and distorted construction with rapid population growth causes many environmental problems, from the deterioration of the socio-economic structure of the city to the misuse of natural resources. Energy resources are decreasing, demand and cost of energy are increasing. This affects the cost of the product. In addition, it uses climate resources more efficiently and positively with global warming and drought. As resources used in electricity production are limited and their prices are increasing, new alternatives for temperature control in the greenhouse system are being investigated. Therefore, cities with variable and dynamic structure should be monitored and followed up regularly.

Letu and others, presented a methodology for estimating electric power consumption from saturated nighttime DMSP/OLS imagery using a stable light correction. They performed an area correction and developed a correction method of saturation light by a cubic regression equation. Correlation rises after the correction of saturation light. Electric power consumption and the CO2 emission have depended on statistical data. Their study shows that the electric power consumption can be estimated with high precision from the stable light [1].

Levin and Duke used night imagery in Israel and West Sharia to examine economic and demographic differences and compared these results with population data (Arab and Jewish) [10]. Identifying the dynamics of urban expansion [11], studies in areas such as revealing changes in urban areas [12]. In recent studies, many scientists have used low-resolution MODIS or DMSP-OLS images to quickly identify expansion in rapidly growing cities [2,13, 14].

II.DATA SOURCE

The basic data source used in the determination of urban areas is satellite images. In parallel with the developments in satellite technology, the quality and variety of images have increased. DMSP-OLS (Defense Meteorological Satellite Program- Operational Linescan System) at night. It is a data set created by Air Force Meteorological Satellite Program and has been widely used in recent years. The temporal series of images is available from 1992 to 2013 and is available to obtain from <http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>. Each pixel at night images has a resolution of 30x30 arcseconds (or 0.86 km² in equator) and the images cover the longitude between +180 and -180 of the Earth and the area between -65 and +75 latitudes. Night images were created as a composite presentation of two satellite images taken during cloudless nights. The data consists of three different groups. "F1? YYYY_v4b_stable_lights.avg_vis.tif" data set was used in our study. The reason for the use of the data in question is that it includes lights, cities and towns with permanent

illumination including gas flare. Short-lived events such as fires are not included in the image data set. In addition, the background noise detection is performed in the image and these values have been changed to zero and the image has been cleared of these values. The images are 8 bits in geoTIF

format and 43201x16801 in size. The cropped images of Istanbul by years shown in Figure 1.

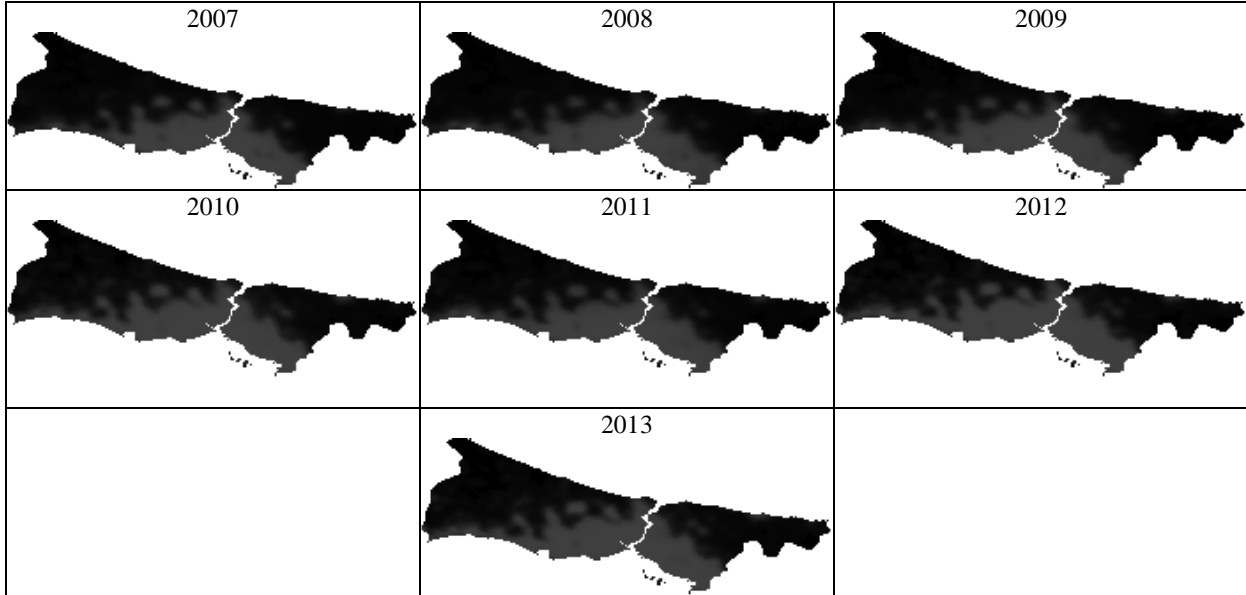


Figure 1. The nighttime imagery of Istanbul city

The population and the electricity consumption data we used in our study were provided from TSE(Turkey Statistical Institute).

III.METHODOLOGY

A. Extraction of the Urban Areas

In the study, cell based image analysis Otsu method was used to determine urban areas. The Otsu method is a threshold detection method that allows the detection of the optimal threshold value that can be used when a gray level image is reduced to two groups. The Otsu method is named from Nobuyuki Otsu, who developed this method. When using this method, it is assumed that the image consists of two color classes, the background and the foreground. Then, for all threshold values, the intra-class variance value of these two color classes is calculated. The threshold value that makes this value the smallest is the optimum threshold value. While the variance value in the class is the minimum value, the variance value among the classes is at the maximum value. Calculating the variance between classes requires less processing, but the calculation of variance between classes for background and foreground pixel classes results in faster results. The method works on gray level images and only looks at how many times the colors are on the image. Therefore, the color histogram of the image is calculated first and all operations are performed on the histogram sequence [7]. After the Otsu method, urban

areas of Istanbul by years (Figure 2.) extracted from nighttime satellite image.

After obtaining urban areas of Istanbul city, the pixels are counted as urban area and rural area. The results of the otsu threshold method 2-class images gathered.

Table 1. Electricity consumption and number of pixels

Year	Total Consumption (MWh)	Number of Pixels
2013	33,925,455	2885
2012	33,084,858	2972
2011	32,672,285	2643
2010	30,525,034	2816
2009	29,147,130	2385
2008	30,008,801	2398
2007	28,501,616	2465

B. Correlation Calculation Between Electricity Consumption and Urban Area

The relationship between free and dependent variables, the level and degree of this relationship is called the method of correlation analysis. In the correlation analysis, if there is only one independent variable, this type of analysis is called simple correlation analysis, and if there are multiple independent

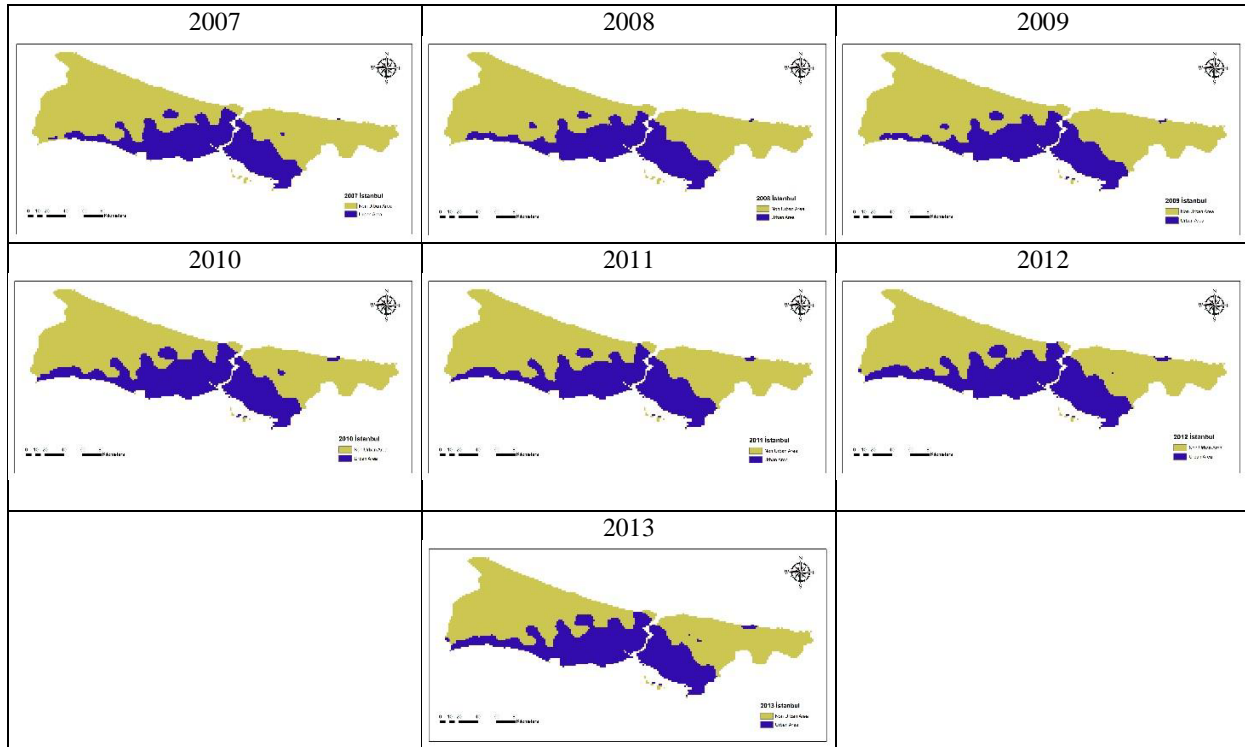


Figure 2. Urban areas of Istanbul by years

variables, this is called multiple correlation. The most commonly used correlation analysis in scientific research [15].

It is very difficult to make the interpretation of the correlation coefficient for the intermediate values except for the exact values. When evaluating coefficients for intermediate values, the number of sample observations (n) is very important. Even in the case of observations based on too many observations, a correlation coefficient of up to 0.25 can be considered meaningful. However, in a small number of 10-15 observations, the correlation coefficient is expected to be above 0.71. According to the value of the correlation coefficient (r), the following comments can be made about the degree of correlation. Pearson correlation coefficients are the most commonly used correlation coefficients used to examine whether there is a relationship between numerical (quantitative) variables in terms of force and direction.

Table 2. Qualification for Pearson r values [8]

0,00 - 0,19	No relationship or unimportantly low
0,20 - 0,39	Weak (low) relationship
0,40 - 0,69	Intermediate relationship
0,70 - 0,89	Strong (high) relationship
0,90 - 1,00	Very strong relationship

Pearson correlation coefficient r; x: Independent variable, y: Equation (1) is used to indicate the dependent variable.

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)} \sqrt{\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}} \quad (1)$$

In this study electricity consumption and urban area correlation are calculated. The electricity consumption values by years were gathered from TUIK. There is a high correlation between night lights data and electric power consumption between 2007 and 2013 years. The correlation was calculated by 0.806.

Table 1. Correlation of electricity consumption and urban area for Istanbul

Image	Year	Energy Consumption (MWh)	Number of Pixels	Correlation
Ist_2013.jpg	2013	33,925,455	2885	0.806
Ist_2012.jpg	2012	33,084,858	2972	
Ist_2011.jpg	2011	32,672,285	2643	
Ist_2010.jpg	2010	30,525,034	2816	
Ist_2009.jpg	2009	29,147,130	2385	
Ist_2008.jpg	2008	30,008,801	2398	
Ist_2007.jpg	2007	28,501,616	2465	

The explanatory coefficient (R^2) gives the explanatory rates of the basic variables. If there is a linear relationship between the two variables, the square of the correlation coefficient is equal to the coefficient of explanation. R^2 ranges from 0 to 1. The approach of R^2 to 1 indicates that most of the change in the dependent variable is explained by the independent variable (Alpar, 2014). According to our correlation result 0.806, the explanatory coefficient is 0.6496. According to the coefficient result, 64% of the variation or variation on the variable X (energy consumption) may be explained by the variation or variation on the Y variable (urban area).

IV.CONCLUSION

Urban area is a parameter that changes every year. The fact that the electricity consumption can be estimated using light activity value and usage of 7-year data for this estimation have created effective results in this study. The R value is 0.806 and R^2 is 0.6496. The result shows that there is strong relationship between urban area and electricity consumption.

V.REFERENCES

- [1] Letu, Husi & Hara, Masanao & Yagi, Hiroshi & Tana, Gegen & Nishio, Fumihiko. (2009). *Estimating the energy consumption with nighttime city light from the DMSP/OLS imagery*, 2009 Joint Urban Remote Sensing Event. 1 - 7. 10.1109/URS.2009.5137699.
- [2] Yücer, Emre & Erener, Arzu. (2018). *GIS Based Urban Area Spatiotemporal Change Evaluation Using Landsat and Night Time Temporal Satellite Data*, Journal of the Indian Society of Remote Sensing. 46. 1-11. 10.1007/s12524-017-0687-5.
- [3] Tilottama Ghosh, Sharolyn J. Anderson, Christopher D. Elvidge, and Paul C. Sutton, *Using nighttime satellite imagery as a proxy measure of human well-being*, Sustainability, 5(12):4988–5019, 2013. ISSN 2071-1050. doi: 10.3390/su5124988.
- [4] Shi, Kaifang & Yu, Bailang & Huang, Yixiu & Hu, Yingjie & Yin, Bing & Chen, Zuoqi & Chen, Liujia & Wu, Jianping. (2014). *Evaluating the Ability of NPP-VIIRS Nighttime Light Data to Estimate the Gross Domestic Product and the Electric Power Consumption of China at Multiple Scales: A Comparison with DMSP-OLS Data*, Remote Sensing. 6. 10.3390/rs6021705.
- [5] Elvidge, C.D., Erwin, E.H., Baugh, K.E., Ziskin, D., Tuttle, B.T., Ghosh, T., et al., 2009. *Overview of DMSP night-time lights and future possibilities*, Urban remote sensing event, pp. 1–5.
- [6] Imhoff, M.L., Lawrence, W.T., Stutzer, D.C., Elvidge, C.D., 1997. *A technique for using composite DMSP/OLS 'city lights' satellite data to map urban area*, Remote Sensing of Environment, 61(3), pp. 361–370.
- [7] H. Atasoy, (2016), *Otsu Eşik Belirleme Metodu*. [Online]. Available: <http://www.atasoyweb.net/Otsu-Esik-Belirleme-Metodu>
- [8] Alpar, R. (2014), *Spor Sağlık ve Eğitim Bilimlerinden Örneklerle Uygulamalı İstatistik ve Geçerlik-Güvenirlilik*, Ankara: Detay Yayıncılık.
- [9] Turan, Muhammed & Yücer, E & Şehirli, Eftal & Karaş, İsmail. (2017). *Estimation Of Population Number Via Light Activities On Night-Time Satellite Images*. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XLII-4/W6. 103-105. 10.5194/isprs-archives-XLII-4-W6-103-2017.
- [10] Levin, N., and Duke, Y., 2012. *High spatial resolution night-time light images for demographic and socio-economic studies*, Remote Sensing of Environment, 119, pp. 1–10.
- [11] Liu, Z., He, C., Zhang, Q., Huang, Q., Yang, Y., 2012. *Extracting the dynamics of urban expansion in China using DMSP-OLS nighttime light data from 1992 to 2008*, Landscape and Urban Planning, 106(1), pp. 62–72.
- [12] Wei, Y., Liu, H., Song, W., Yu, B., Xiu, C., 2014. *Normalization of time series DMSP-OLS nighttime light images for urban growth analysis with Pseudo Invariant Features*, Landscape and Urban Planning, 128, pp. 1-13.
- [13] Huang, X., Schneider, A., Friedl, M.A., 2016. *Mapping sub-pixel urban expansion in China using MODIS and DMSP/OLS nighttime lights*. Remote Sensing of Environment, 175, pp. 92-108.
- [14] Small, C., Elvidge, C.D., 2013. *Night on Earth: Mapping decadal changes of anthropogenic night light in Asia*, International Journal of Applied Earth Observation and Geoinformation, 22, pp. 40-52.
- [15] Türkbâl, A., *Bilimsel Araştırma Metodları Ve Uygulamalı İstatistik*, Atatürk Üniversitesi Basım Evi, 1981

Classification of Flower Species Using Convolutional Neural Networks

E.S. GHRAIRI¹, S. KAHVECI² and İ. BABAOĞLU³

¹ Konya Technical University, Konya/Turkey, elaf_2@yahoo.com

² Konya Technical University, Konya /Turkey, semihkahveci@outlook.com

³ Konya Technical University, Konya /Turkey, ibabaoglu@ktun.edu.tr

Abstract - Having a growing phenomenon on its usage, convolutional neural networks (CNNs) has achieved significant success in many image processing problems such as object detection, segmentation and classification as well as various natural language processing problems. This study presents an investigation of the performance of CNNs on the classification of the flower species. The dataset namely flower-color-images is obtained from the Kaggle website. The dataset consist of 210 images of 10 different flower species. In order to increase the success of the classification process, pre-processing techniques have been applied on the dataset. Data augmentation techniques which are known as mirroring and turning were applied at 4 different angles and the size of the data set was increased. The investigation is evaluated for both color and gray images. The suggested CNN model is based on AlexNet including some modifications, and 5-fold-cross-validation method is also utilized. According to the experiments, the suggested model achieves 84% and 46% classification rates for color and grayscale images respectively.

Keywords - Deep learning, convolutional neural networks, image classification

I. INTRODUCTION

Many machine learning algorithms have been applied for solving image classification problems. Convolutional neural networks (CNNs) have a better generalization capability than any other network model with a deep feed architecture and a fully connected layer in its structure. CNN can learn extremely abstract properties and efficiently identify objects. There are some reasons why CNN is more successful than other models. First of all, the CNN model is greatly reduced the number of parameters that require training. Fewer parameters of CNN provides a smoother model training and less overfitting problems. Secondly, the classification stage is combined with the feature extraction phase both of which use the learning process. Moreover, it is more difficult to apply complex networks in classic artificial neural network (ANN) models than in CNN. Due to its advantages, CNN achieve better successful results than classic ANN models.

Today, the size of data collection and datasets makes CNN more successful. Training of CNN models with advanced hardware and GPU usage realizes high achievements. Today Convolutional neural networks have succeeded in many areas such as object extraction, image

classification, image segmentation, motion detection, natural language processing [2-4].

Flower classification is a difficult problem due to the large flower classes that have similar characteristics. Although they are different species, some flowers have similar color, shape and appearance. The success of some applications, such as plant monitoring systems, plant identification, the flower industry, depends on the result of a successful floral classification. It is possible to make a manual classification to classify flowers, but the high number of flowers and images in a complex image will make the classification process long and tedious [5]. This problem has been addressed in the field of machine learning in order to improve the performance of the flower classification and to accelerate the process.

In this study an investigation of the performance of CNNs on the classification of the flower species are carried out. Color and gray scale images are applied on the CNN model and the results are discussed.

II. RELATED WORKS

İnik, O. And Ülker have presented detailed information about Deep Learning models and CNNs. Authors have made clear explanation about layers of convolution, pooling, ReLu, dropout, fully linked and classification, which are the layers of the architecture of the intersectional neural network. The authors have also described the AlexNet, ZFNet, GoogLeNet, Microsoft ResNet and RCNN architectures that can be considered as basic architectures in Deep Learning [6].

Mercan has used chest radiography for detecting chest tube cancer after treatment with CNNs. The author has developed a software written in C ++, based on CNN. The software was designed as a library and allowed to operate CNNs with different structures. The biggest advantage of this software is that it performs the extraction process very quickly from the images. The performance of the developed software were tested with pixel-based ROC analysis and used to calculate the true positive, real negative, false positive, false negative, and mean accuracy, precision, specificity, percentages in the testing phase. As a result, 99.99% accuracy, 59% sensitivity, 99.99% specificity values was obtained [7].

Cengil and Çınar have tried to solve image classification problem by using the deep learning algorithm. The authors have used CIFAR-100 data set including 8 different class of images such as bus, tractor, train, dinosaur, elephant, butterfly,

chair and television. Totally, 840 images were selected for experiments in the study; 800 images of these images were used in the training process and the remaining 40 images were used as test data. The authors have classified these images by using CNNs [8].

Büyükyılmaz designed a CNN for automatically perceiving of eimeria parasite species and determination of whether they are diseased or not. Eimeria microscopic images of chickens and rabbits were given as inputs to the proposed CNN system. The authors have created and tested different network models by using some existing libraries such as OpenCV and Keras. According to the test results, 87.75% accuracy was determined classification of the chicken and rabbit dataset, and the diseased cells were classified with an accuracy of 78.42% [9].

Anwer used the CNN algorithm in the medical field. Anwer's thesis work has suggested a CNN-based system for breast cancer diagnostics. The proposed CNN-based system has tested on breast cancer dataset in the Wisconsin UCI machine learning repository. Based on the experimental results, the proposed classification-based breast cancer diagnosis system has successfully diagnosed the breast cancer [10].

D. Nkemelu et al. With the Support Vector Machine and K-Nearest Neighbors, which are the two traditional algorithms on the classification of plant seedlings, they worked with the convolutional neural network (CNN) from deep learning models. They used a data set that contained 4.275 different images of 12 species and about 960 seedlings. They observed that Convolutional neural network is more successful than other algorithms. It was found that this study has an important potential to optimize CNN crop yield and increase data efficiency [11].

Q. Li et al. It presented Convolutional neural networks (CNNs) performance in the classification of interstitial lung disease (ILD) and lung image patches. They used the ILD dataset, which was publicly accessible with the CNN structure they had customized. The CNN model has a convolution layer with 7x7 filter, Maxpooling with 2x2 filter and 3 fully connected layers. The only convolution layered architectures they proposed have had good results in their problems [12].

S. Hershey et al. Convolutional neural networks (CNNs) have demonstrated the success of CNNs in audio performance due to their success in image classification. To classify the music of the 70M educational video data set, it used Fully-connected Deep Nerve Networks (DNNs), AlexNet, Inception, VGG, and ResNet. They showed that increasing the size of the training data set increased the success to a certain extent. In order to increase the data set, the best performance of the models is presented by the ResNet-50 model [13].

H. Yalçın and S. Razavi presented the success of the convolutional neural network (CNN) in the field of classification with a CNN model in which they proposed the classification of plant species in images collected from smart agrostations. The proposed architecture consists of 5 layer convolution and 3 full connected layer layers. The models use a pre-trained network. They have compared their results with the Support vector machine classifier and feature identifiers

such as LBP and GIST. According to other algorithms, the proposed model has achieved a success of 97.04[14].

III. MATERIAL AND METHODS

A. Dataset

There are flower pictures of 10 different species of plants in the dataset. These plant species include phlox, rose, calendula, iris, leucanthemum maximum, bellflower, viola, goldquelle, peony and aquilegia. The dataset has 210 images with 128x128 of sizes. The images are in PNG data format. The class label for each image is kept in a separate file in CVS format. The data set is called as "flower-color-images", and it is available on the Kaggle website [15].

B. Data Augmentation

While using CNN's, better success rates are achieved with a large amount of data in many deep learning approaches. In this study, data augmentation pre-processing technique is applied to our data in order to increase success rate on flower classification problem. The data extraction method is applied in 4 different directions including vertical mirror, horizontal mirror, diagonal top left and diagonal top right. By data augmentation, the number of images in the dataset is increased from 210 to 1050 images. The image dimensions also have been reduced from 128x128 to 100x100. The dataset is evaluated by using the original color images together with their gray-scale counterparts. Demonstration of some images obtained by data augmentation are given in Figure 1.



Figure 1: Some of the images after data augmentation

C. The Convolutional Neural Network

A CNN is made out of at least one convolutional layer together with one another subsampling/pooling layer and followed via fully connected layer at the end. In principal, there are three extensions that distinguishes a CNN from a simple feed-forward network; weight sharing, spatial pooling and local receptive fields [16].

Convolutional Layer: This layer is the layer from which the input properties are subtracted. It can be seen as the building block of CNN. In this layer, some filtering techniques such as blurring, edge or corner detection are applied to extract properties from the data. Discrete convolutions are calculated by shifting a filter mask over the input image and calculating the sum of products. The result is written to the current center of the mask. These filter masks produce the so-called local receptive fields [16].

Pooling Layer: This layer is a layer that is frequently added between the Convolutional Layers. The main task of this layer

is to reduce the amount of data and simplify the calculations. There can be more than one pooling process, but the max-pooling method is commonly preferred in CNN.

Fully Connected Layer: Fully connected layer converts the incoming input volume into a n-dimensional vector with defined classes.

Classification Layer: Classification layer comes after the fully connected layer, and the classification process is made in this layer. The output value of this layer is equal to the number of objects to classify. Different classifiers can be used in this layer. Because of its great success, softmax classification is commonly preferred in the classification layer.

In CNNs, raw data are represented as tensor. Tensor concept can be popularized as higher order matrices. For example, the vector is an order (1) tensor, a grayscale image is an order (2) tensor, and an image that has three channels (R, G, & B) is order (3) tensor. The input, intermediate representations and parameters are all represented as tensors in the CNN model [16].

1) AlexNet

At the ImageNet Large Scale Visual Recognition Competition (ILSVRC) which is held in 2010, authors trained a big, deep convolutional neural network for classifying the 1.2 million high resolution images into the 1000 different classes. According to the test results, the authors achieved top-1 and top-5 error rates of 37.5% and 17.0%, which are much better than the former state of the art approaches. Besides, in the aim of achieving a faster training time, they utilized non saturating neurons and a so effective GPU applied of the convolution operation. Then to reduce overfitting in the fully connected layers, a recently-developed regularization manner named (dropout) that proved to be so efficient was utilized. Another proposed CNN model, AlexNet, which was evaluated in the ILSVRC-2012 competition, achieved the top-5 error-rate of 15.3% at the second best. This newly proposed model has 60 million parameters and 650 thousand neurons and consists of eight layers with weights; the first five layers are convolutional and the three remaining are fully connected. The output of the last fully connected layer is fed to a 1000-way softmax, that generates a distribution over the 1000 class labels. The response of normalization layers follows the first and second convolutional layers. Max pooling layers follow both of the response normalization-layers as well as the last convolutional layer (fifth). ReLu non-linearity is applied to the result of all convolution and full connected layers; the standard non-linearity utilized are Tanh or Sigmoid non-linearity. However, the problem with them is that they have “saturating regions”. In these regions the gradient becomes very small. So, gradient descent becomes very slow and training takes a long time. There is only a so small sine, where these functions confer large gradient that enough to make training progress quicker.

The Alexnet model uses a drop-out layer to overcome the overfitting problem. Dropped neurons are not transferred forward in this layer. Therefore, every time an input is presented the neural network at random samples a various architecture. However, all these architectures share-weights. As a result of that, it compels each neuron to depend less, thus

compose complex dependencies, on other neurons forcing it to learn more strong properties that are advantageous in synchronism with many different random subsets of the other neurons. In the test stage, all the neurons are utilized, but their outputs are multiplied by the dropout probability. Dropout is utilized in the first two fully-connected layers (layers 6 & 7) in the AlexNet model [1].

D. Proposed CNN Model

The proposed CNN model was constructed similar to the AlexNet [1] including some modifications on the structure. In the structure of the proposed CNN model, four two-dimensional layers were used; 32 in the second layer, 128 in the third layer, and 256 in the fourth layer. (3x3) kernel dimensions were used in the first Conv2D layer of the proposed CNN model and (2x2) kernel dimensions were used in the rest of the Conv2D layers. MaxPooling2 pooling layers have been used between each of these Conv2D layers. In the last layer of the proposed CNN model, softmax was utilized as the ReLu activation function in the other layers. The structure of the proposed CNN model is given in Table 1.

Table 1: The proposed CNN model

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 84, 84, 32)	896
max_pooling2d_1 (MaxPooling2)	(None, 42, 42, 32)	0
conv2d_2 (Conv2D)	(None, 42, 42, 64)	18496
max_pooling2d_2 (MaxPooling2)	(None, 21, 21, 64)	0
conv2d_3 (Conv2D)	(None, 21, 21, 128)	73856
max_pooling2d_3 (MaxPooling2)	(None, 11, 11, 128)	0
conv2d_4 (Conv2D)	(None, 11, 11, 256)	295168
max_pooling2d_4 (MaxPooling2)	(None, 6, 6, 256)	0
dropout_4 (Dropout)	(None, 6, 6, 256)	0
flatten_1 (Flatten)	(None, 9216)	0
dense_1 (Dense)	(None, 256)	2359552
dropout_5 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 256)	65792
dropout_6 (Dropout)	(None, 256)	0
dense_3 (Dense)	(None, 10)	2570

IV. EXPERIMENTAL RESULT

This study was developed and experimented in the python programming language by using the TensorFlow library published by Google. The model was applied to two different versions of the dataset as color image classification and grayscale image classification. Data is augmented to increase the amount of the images on the dataset. The proposed model was evaluated by employing 5-fold cross validation method for 50 epochs and 100 epochs. The training accuracy and error rate of the color and grayscale images evaluated for 50 epochs are given in Table 2 and Table 3 respectively. The test accuracy of the color and grayscale images evaluated for 50 epochs are given in Table 4 and Table 5 respectively. Besides, the accuracy and error rate graphs respect to the epoch size for the color images and grayscale images evaluated for 50 epochs are given in Fig. 2 and Fig. 3 respectively.

Table 2: The classification accuracy and error rate of the training phase for color images

Classification accuracy (%)								
Epoch	1	2	16	17	25	26	49	50
1. Fold	0.14	0.18	0.39	0.40	0.48	0.49	0.77	0.76
2. Fold	0.13	0.15	0.39	0.42	0.46	0.48	0.72	0.74
3. Fold	0.13	0.16	0.37	0.41	0.48	0.48	0.73	0.71
4. Fold	0.13	0.16	0.38	0.40	0.48	0.51	0.77	0.78
5. Fold	0.10	0.13	0.38	0.39	0.47	0.50	0.69	0.69
Average	0.13	0.15	0.38	0.41	0.47	0.49	0.73	0.74
Error rate (%)								
Epoch	1	2	16	17	25	26	49	50
1. Fold	2.29	2.24	1.69	1.68	1.47	1.42	0.72	0.70
2. Fold	2.30	2.27	1.68	1.66	1.48	1.46	0.77	0.79
3. Fold	2.29	2.26	1.70	1.68	1.48	1.46	0.84	0.83
4. Fold	2.29	2.25	1.69	1.66	1.43	1.41	0.70	0.64
5. Fold	2.30	2.28	1.75	1.71	1.54	1.48	0.90	0.88
Average	2.29	2.26	1.70	1.68	1.48	1.45	0.79	0.77

Table 3: The classification accuracy and error rate of the training phase for grayscale images

Classification accuracy (%)								
Epoch	1	2	16	17	25	26	49	50
1. Fold	0.13	0.20	0.94	0.96	0.98	0.98	1.00	1.00
2. Fold	0.16	0.31	0.93	0.95	0.97	0.98	1.00	0.99
3. Fold	0.12	0.26	0.93	0.94	0.98	0.99	1.00	0.99
4. Fold	0.09	0.07	0.85	0.87	0.98	0.95	0.99	0.99
5. Fold	0.12	0.14	0.93	0.95	0.98	0.98	0.99	0.99
Average	0.12	0.20	0.92	0.93	0.98	0.97	0.99	0.99
Error rate (%)								
Epoch	1	2	16	17	25	26	49	50
1. Fold	13.33	7.86	0.18	0.12	0.08	0.07	0.02	0.02
2. Fold	6.76	2.05	0.20	0.16	0.08	0.06	0.02	0.03
3. Fold	13.31	6.33	0.23	0.16	0.05	0.05	0.01	0.04
4. Fold	14.48	14.82	0.45	0.35	0.08	0.15	0.07	0.06
5. Fold	13.50	10.27	0.22	0.18	0.09	0.08	0.02	0.04
Average	12.28	8.27	0.26	0.20	0.08	0.08	0.03	0.04

When looking at the classification accuracy and error results obtained from the training phase, it can be seen that the proposed CNN model has achieved classification accuracy and error rate as 0.99 and 0.04 for the color images respectively. On the other hand, the proposed CNN model has achieved classification accuracy and error rate as 0.74 and 0.77 for the grayscale images respectively. The proposed CNN model performed prospering classification accuracy and error results on the color images. Besides, the accuracy and error rate graphs respect to the epoch size for the color images and grayscale images evaluated for 50 epochs are given in Figure 2 and Figure 3 respectively. Considering the figures, it can be said that the training phase of the proposed model is carried out very smoothly.

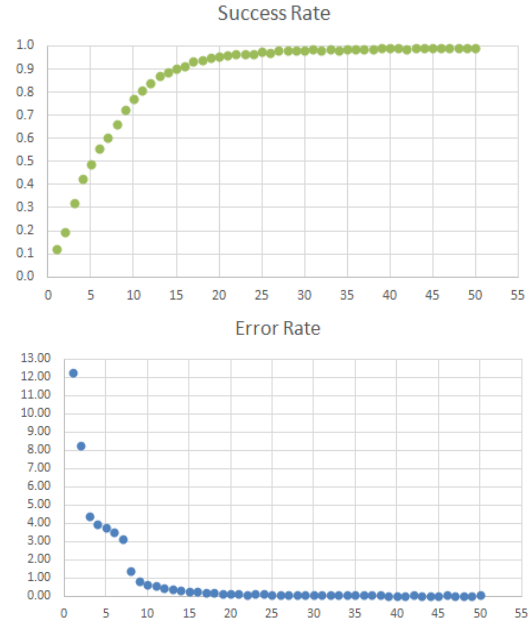


Figure 2: The accuracy and error rate graphs respect to the epoch size for the color images

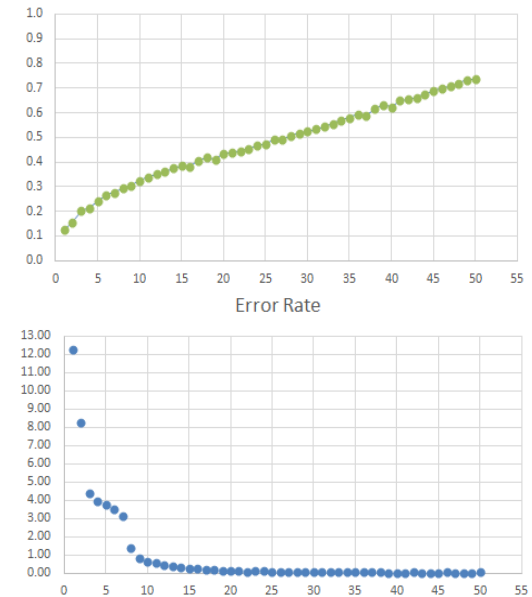


Figure 3: The accuracy and error rate graphs respect to the epoch size for the grayscale images

Table 4: The test accuracy results of the color images

Classification accuracy (%)								
Epoch	1	2	16	17	25	26	49	50
1. Fold	0.12	0.40	0.78	0.79	0.77	0.79	0.80	0.81
2. Fold	0.31	0.41	0.83	0.79	0.78	0.82	0.82	0.82
3. Fold	0.20	0.47	0.83	0.83	0.89	0.87	0.83	0.88
4. Fold	0.10	0.07	0.80	0.78	0.79	0.84	0.86	0.86
5. Fold	0.12	0.37	0.81	0.82	0.83	0.85	0.84	0.86
Average	0.17	0.34	0.81	0.80	0.81	0.83	0.83	0.84

Table 5: The test accuracy results of the grayscale images

Classification accuracy (%)								
Epoch	1	2	16	17	25	26	49	50
1. Fold	0.17	0.24	0.33	0.46	0.38	0.47	0.54	0.50
2. Fold	0.17	0.18	0.35	0.37	0.40	0.42	0.53	0.51
3. Fold	0.15	0.18	0.35	0.36	0.40	0.41	0.44	0.45
4. Fold	0.14	0.18	0.36	0.41	0.40	0.42	0.43	0.42
5. Fold	0.12	0.23	0.34	0.41	0.40	0.43	0.45	0.47
Average	0.15	0.20	0.35	0.40	0.40	0.43	0.48	0.47

The training phase of color images took place an average of 22 seconds per-epoch, however, the training phase of grayscale images has an average of 8 seconds per-epoch. According to the results it can be said that the training phase of grayscale images are faster than the training phase of color images.

When the experimental results in Table 4 and Table 5 are examined, the proposed CNN model recorded an accuracy with 0.84 on the test data set during the testing phase. For the gray data set, the proposed CNN model achieved 0.47 classification accuracy on the test data set during the testing phase. In general, the proposed CNN model performed very well classification accuracy and error results on the color data set.

In addition, the classification results of the proposed CNN model are obtained using 100 epochs and 5-fold cross-validation. The training accuracy and error rate of the color and grayscale images evaluated for 100 epochs are given in Table 6 and Table 7 respectively. The test accuracy of the color and grayscale images evaluated for 100 epochs are given in Table 8 and Table 9 respectively.

Table 6: The classification accuracy and error rate of the training phase for color images

Classification accuracy (%)										
Epoch	1	2	16	17	49	50	75	76	99	100
1. Fold	0.11	0.12	0.88	0.90	0.99	0.99	1.00	1.00	1.00	1.00
2. Fold	0.15	0.30	0.95	0.96	1.00	0.99	1.00	0.99	1.00	1.00
3. Fold	0.10	0.12	0.10	0.10	0.10	0.10	0.95	0.96	0.99	0.99
4. Fold	0.15	0.29	0.95	0.93	0.99	0.99	1.00	1.00	0.99	0.99
5. Fold	0.11	0.23	0.91	0.93	0.99	0.99	1.00	1.00	0.99	1.00
Average	0.13	0.21	0.76	0.77	0.81	0.81	0.99	0.99	0.99	0.99
Error rate (%)										
Epoch	1	2	16	17	49	50	75	76	99	100
1. Fold	14.22	14.10	0.33	0.28	0.03	0.05	0.02	0.01	0.00	0.01
2. Fold	10.85	2.30	0.18	0.12	0.01	0.03	0.00	0.02	0.03	0.01
3. Fold	14.47	14.02	14.43	14.43	14.44	14.43	0.13	0.14	0.03	0.05
4. Fold	8.91	2.27	0.20	0.19	0.02	0.04	0.02	0.01	0.05	0.01
5. Fold	13.51	9.37	0.24	0.23	0.03	0.03	0.00	0.03	0.03	0.02
Average	12.39	8.41	3.08	3.05	2.91	2.92	0.04	0.04	0.03	0.02

Considering the classification accuracy and error results obtained from the training phase, the proposed CNN model has achieved classification accuracy and error values on the color images with 100 epochs during the training phase 0.99 and 0.02, respectively. On the other hand, the proposed CNN model has achieved classification accuracy and error values on the grayscale images with 100 epochs during the training phase 0.98 and 0.07, respectively. The proposed CNN model performed very well classification accuracy and error results on the color images with 100 epochs.

Table 7: The classification accuracy and error rate of the training phase for grayscale images

Classification accuracy (%)										
Epoch	1	2	16	17	49	50	75	76	99	100
1. Fold	0,12	0,14	0,38	0,40	0,71	0,71	0,94	0,92	0,98	0,99
2. Fold	0,13	0,17	0,41	0,39	0,72	0,74	0,95	0,94	0,99	0,97
3. Fold	0,10	0,18	0,39	0,42	0,76	0,75	0,96	0,95	0,99	0,98
4. Fold	0,12	0,15	0,41	0,41	0,74	0,74	0,93	0,93	0,98	0,98
5. Fold	0,11	0,16	0,39	0,42	0,75	0,76	0,95	0,94	0,98	0,98
Average	0,12	0,16	0,40	0,41	0,74	0,74	0,94	0,93	0,98	0,98
Error rate (%)										
Epoch	1	2	16	17	49	50	75	76	99	100
1. Fold	2,29	2,27	1,72	1,70	0,84	0,83	0,21	0,26	0,07	0,05
2. Fold	2,29	2,26	1,70	1,71	0,80	0,77	0,19	0,17	0,06	0,08
3. Fold	2,30	2,27	1,65	1,65	0,70	0,71	0,18	0,16	0,04	0,07
4. Fold	2,30	2,27	1,67	1,68	0,80	0,75	0,22	0,24	0,06	0,08
5. Fold	2,30	2,27	1,68	1,67	0,78	0,74	0,18	0,19	0,05	0,07
Average	2,30	2,27	1,69	1,68	0,78	0,76	0,20	0,21	0,06	0,07

The training phase of color images with 100 epochs took place on average in 22 seconds per-epoch, however, the training phase of grayscale images with 100 epochs has an average of 8 seconds per-epoch. According to these results, it can be said that the training phases of grayscale images with 100 epochs faster than the training phase of the color images with 100 epochs.

Table 8: The test accuracy results of the color images

Classification accuracy (%)										
Epoch	1	2	16	17	49	50	75	76	99	100
1. Fold	0.10	0.11	0.74	0.73	0.80	0.81	0.81	0.82	0.80	0.80
2. Fold	0.35	0.53	0.75	0.70	0.81	0.87	0.80	0.86	0.89	0.85
3. Fold	0.11	0.10	0.10	0.10	0.10	0.10	0.84	0.82	0.88	0.87
4. Fold	0.38	0.47	0.77	0.83	0.83	0.84	0.85	0.79	0.79	0.79
5. Fold	0.23	0.34	0.81	0.72	0.84	0.82	0.83	0.81	0.85	0.82
Average	0.23	0.31	0.64	0.62	0.68	0.69	0.83	0.82	0.84	0.83

Table 9: The test accuracy results of the grayscale images

Classification accuracy (%)										
Epoch	1	2	16	17	49	50	75	76	99	100
1. Fold	0.11	0.24	0.40	0.41	0.48	0.47	0.51	0.53	0.50	0.52
2. Fold	0.10	0.20	0.42	0.42	0.48	0.50	0.54	0.54	0.56	0.55
3. Fold	0.21	0.19	0.41	0.39	0.46	0.50	0.50	0.52	0.51	0.55
4. Fold	0.12	0.20	0.40	0.34	0.45	0.43	0.47	0.48	0.48	0.50
5. Fold	0.17	0.18	0.36	0.40	0.50	0.49	0.48	0.49	0.44	0.50
Average	0.14	0.20	0.40	0.40	0.48	0.48	0.50	0.51	0.50	0.52

When the experimental results in Table 8 and Table 9 are examined, the proposed CNN model recorded an accuracy with 0.83 on the test data set during the testing phase. For the grayscale images, the proposed CNN model achieved 0.52 classification accuracy on the test data set during the testing phase. In general, the proposed CNN model performed promising classification accuracy and error results on the color images.

V. CONCLUSION

The aim of this study is to increase the success rate of the flower classification problem in order to achieve successful and faster results in applications in the flower industry. CNN, one of the deep learning algorithms used for classification aim, was applied to the flower data set. The model was designed in two different schemes: colored and gray-scale. Considering

color images, proposed model achieved a success rate of 0.84, while the proposed model achieved a success rate of 0.47 considering grayscale images. The model performs a shorter time evaluation process for grayscale images but achieves higher success rate for color images. The CNN algorithm was more successful in comparison to grayscale images on CNN color images because of its more distinctive and more features on the convolution layer color images. As a result of the study, a faster and more successful classification model was provided in the flower industry.

VI. REFERENCES

- [1] A. Krizhevsky, I. Sutskever and G.E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks", *Advances in neural information processing systems*, 1097-1105, 2012,
- [2] M. Sornam, K. Muthusubash and V. Vanitha, "A Survey on Image Classification and Activity Recognition using Deep Convolutional Neural Network Architecture", *2017 Ninth International Conference on Advanced Computing (ICoAC)*, IEEE, 121-126, 2017.
- [3] S. Indolia, A.K. Goswami, S.P. Mishra, and P. Asopa, "Conceptual Understanding of Convolutional Neural Network-A Deep Learning Approach", *Procedia Computer Science*, 132, 679-688, 2018.
- [4] T. M. Mitchell, "Machine Learning", McGraw-Hill, Inc., New York, NY, USA, 1 edition, 1997.
- [5] M. Das, R. Manmatha, E. Riseman, 'Indexing flower patent images using domain knowledge', *IEEE Intelligent Systems and their Applications*, 14, (5), pp. 24-33, 1999.
- [6] Ö. İnik and E. Ülker, "Derin Öğrenme ve Görüntü Analizinde Kullanılan Derin Öğrenme Modelleri", *Gaziosmanpaşa Bilimsel Araştırma Dergisi*, 6(3), 85-104, 2017.
- [7] C. Wang, and Y. Xi, "Convolutional Neural Network for Image Classification", *Johns Hopkins University Baltimore*, MD, 21218.
- [8] E.Ç. Kaptan, "Image Segmentation and Texture Mapping on Pillows Using Fully Convolutional Neural Networks", Master's Thesis, *Bahçeşehir University*, Istanbul, 1-48, 2018.
- [9] E. Cengil and A. Çınar, "A New Approach for Image Classification: Convolutional Neural Network", *European Journal of Technic*, 6(2), 96-103, 2016.
- [10] Büyükyılmaz, M., "Mikroskobik Görüntüler Üzerinde Derin Öğrenme Algoritmaları Kullanarak Hastalıklı Hücrelerin Otomatik Tanımlanması", Yüksek Lisans Tezi, *Necmettin Erbakan Üniversitesi*, Konya, 1-53, 2017.
- [11] D. Nkemelu, D. Omeiza and L. Nancy, "Deep Convolutional Neural Network for Plant Seedlings Classification", *ArXiv* 2018, November 2018
- [12] Q. Li, W. Cai, X. Wang, Y. Zhou, D.D. Feng and M. Chen, "Medical Image Classification with Convolutional Neural Network", *ICARCV* 2014, 10-12th December 2014.
- [13] S. Hershey, S. Chaudhuri, D.P.W. Ellis, J.F. Gemmeke, A.Jansen, R.C. Moore, M. Plakal, D. Platt, R.A. Saurous, B. Seybold, M. Slaney, R.J. Weiss, K. Wilson, "CNN Architectures For Large-Scale Audio Classification" 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5-7 March 2017.
- [14] H. Yalçın, S. Razavi, "Plant Classification using Convolutional Neural Network", 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics), 18-20 July 2016.
- [15] <https://www.kaggle.com/olgabelitskaya/flower-color-images> (15.03.2019)
- [16] B. Taşhan, "Road Lane Detection System with Convolutional Neural Network", Master's Thesis, *Bahçeşehir University*, Istanbul, 1-54, 2017.

Real-Time Diseases Detection of Grape and Grape Leaves using Faster R-CNN and SSD MobileNet Architectures

Shekofa GHOURY¹, Cemil SUNGUR² and Akif DURDU³

¹Konya Technical University, Konya/Turkey, shekofaghouri@hotmail.com

²Konya Technical University, Konya/Turkey, csungur@ktun.edu.tr

³Konya Technical University, Konya/Turkey, akifdurdu@gmail.com

Abstract - Plant Diseases are a major problem in the agricultural sector. Accurate and rapid detection of diseases and pests in plants can help to develop an early treatment technique while greatly reducing economic losses. It is very difficult and time-consuming to follow and test the grape disease manually and requires extensive processing time, a tremendous amount of procedures and expertise in some grape diseases. A fast automated processing system is needed to detect grape disease. In this paper, the task of grape and grape leaf disease detection is performed. The goal is to differentiate healthy and diseased grapes and grape leaves. This task is done using transfer learning by using two pre-trained deep learning models, Single SSD_MobileNet v1 and Faster R-CNN Inception v2. Faster-R-CNN Inception v2 model classified 95.57% of all the testing images correctly with a classification accuracy of 78% to 99%. The accuracy level of this model was higher but with longer processing time. SSD_MobileNet v1 only classified 59.29% of all the testing images correctly but with better processing time. This model was giving pretty high classification accuracy for the images which are organized, contains lesser noise, have uniform background and sizes. Which was between 90% to 99% of accuracy. But for the images which contain noise, have different backgrounds and resolutions it was giving very poor classification accuracy which was between 52% to 80% alongside a high percentage of misclassifications. According to these results, it's been concluded that for the current time SSD_MobileNet v1 model may not be useful for real-time classifications for this task but the Faster-R-CNN Inception v2 model could perform quite better.

Keywords - Grape Disease Detection, Convolutional Neural Networks, Deep Learning, Faster Regional - CNN, Inception, SSD, MobileNet.

I. INTRODUCTION

AGRICULTURE is the most important and old profession in Afghanistan. Approximately 80% of the population of Afghanistan is engaged in agriculture and farming, 31% of Afghanistan's national income comes from agricultural productions, and pays the utmost attention to food production. The grape plant is a product that is largely produced in the field of agriculture. Grape cultivation has social and economic importance in Afghanistan. The quality of grapes has deteriorated for several reasons over the last few years. One of

the important reasons is the diseases of the grapes. In order to prevent diseases, farmers are spraying large amounts of pesticides, but this leads to an increase in production costs. In addition, farmers cannot identify diseases manually. Diseases are diagnosed only after infection, but it takes a lot of time and has adverse effects on the ligament. It is very difficult and time-consuming to follow and test the grape disease manually and requires extensive processing time, a tremendous amount of procedures and expertise in some grape diseases.

Nowadays the goal of modern agriculture is to produce the maximum amount of yield with less requirement of energy and time, thus modern image processing, machine learning, and deep learning techniques are used. Agriculture products need proper quality control to gain more products. Taking these reasons into consideration, it is important to detect the diseases of grape and grape leaves using an automated processing system. Plant disease identification using images is an important challenge. The recent movements in the use of various machine learning algorithms for plant disease classification have yielded promising results in several selected plant diseases and crops.

This paper presents a detection of grape and grape leaf disease detection and also a differentiation between grape and grape leaf. This work is done using Tensorflow object detection API on two pre-trained models, Single Shot Detector (SSD) MobileNet v1 and Faster R-CNN Inception v2. These models had been retrained on grapes dataset which had been collected from the internet and PlantVillage dataset [1] and prepared specially for this work. This paper is organized in the following structure: a brief idea of previous work and techniques used for plant disease detection and recognition is presented in Section II, materials and methods are in Section III, experimental results and discussions are in Section IV and finally, the conclusion of the work is described in Section V.

II. RELATED WORK

Many researchers had worked on different plant disease detection using different techniques and gained pretty good results on their suggested methods. Some of them are concluded as below:

Kakade and Ahire, proposed a new approach based on image processing and neural networks based classification to detect and classify grape leaf diseases. Their aim was to detect, identify and accurately measure the initial symptoms of the disease. As a result they achieved 92.94% accuracy of disease detection and classification [2]. Sjadojevic et al., developed a system to identify 13 types of diseases out of healthy ones in 5 crops using CNN based and fine-tuning concepts on the images which they had downloaded from the internet. They achieved 96.3% as Top-1 success and 99.99% as a top-5 success for 100,000 iterations [3]. Mohanty et al., had analysed plant leaf diseases using transfer learning approach in which they retrained two pre-trained deep learning based architectures (AlexNet and GoogLeNet) on a public dataset having 54,306 images of diseased and healthy plant leaves, the models were able to classify 26 diseases of 14 crop species, and they had achieved over 99.35% of classification accuracy [4]. Fuentes et al., proposed a robust deep-learning-based approach for the detection of diseases and pests in tomato plants. They used Faster R-CNN, R-FCN and SSD as deep learning meta-architectures and combined each of them with deep feature extractors (VGG net and Residual Network (ResNet)). Their system was able to recognize 9 different types of diseases and pests [5]. Mude et al., suggested a method to detect cotton leaf diseases using image processing techniques and artificial neural networks. They tried to detect 4 types of diseases (Rendening, Rust, Bacterial, Fussarium). Their proposed method provided approximately 84% of efficiency [6]. Kothawale et al., used image processing techniques and SVM classifier technique to detect and classify grape leaf disease. They trained their system on 90 images of healthy and diseased grape leaves. As an output result, they obtained 89.90% of accuracy [7]. Rangarajan et al., performed tomato crop disease classification task using two pre-trained deep learning algorithms (AlexNet and VGG16 net) on images of tomato leaves (6 diseases and a healthy class) from plantVillage dataset. They achieved a classification accuracy of 97.29% for the VGG16 net and 97.49% for AlexNet [8]. Horea and Mihai presented a new, high-quality dataset containing 55244 images of fruits with 81 different classes and also experimented detection of fruits using the deep neural network. First, the filmed fruits by rotating them using a motor, then they extracted fruit frames from it. They obtained different results for different network configurations on the dataset, they achieved 99.42% as maximum training accuracy and 96.52% as maximum test accuracy [9].

III. MATERIALS AND METHODS

Two deep neural networks are used for the task of grape and grape leaf disease detection using transfer learning. In particular, deep learning is a subset of machine learning which includes a specific class of functions inspired by neural networks and a particular way of training them [10]. In recent years, depending on the large data and powerful hardware, deep learning has become increasingly the actual choice for the processing of complex high-dimensional data such as text, images and audio signals [10].

Generally there are two choices of training a deep neural network: one is to create a deep neural network from scratch

and train it on the desired dataset, which is called training from scratch and the other is to adopt a pre-trained network which is similar to our task and retrain it on the desired dataset for our task with bringing some changes into it, which is called transfer learning. The training of deep neural networks with millions of parameters requires large amounts of data in order of millions of samples to properly limit optimization [11]. To train these networks from scratch there is a need for large amounts of training data and high computational resources, this had led researchers to adapt pre-trained networks to the desired task domain with domain-specific data by meaning of transfer learning which applies the knowledge learned from a problem to another, different but related problem [11]. Because the dataset which is used for this task in this paper is not a very big dataset in order to train a deep neural network from scratch, the transfer learning method had been chosen for this task to fine tune the model parameters to our dataset. Thus two deep learning models called SSD_MobileNet v1 and Faster-R-CNN Inception v2 which are pre-trained COCO-Tensorflow object detection models [12].

A. System Configuration

Models of deep learning need very strong Graphics Processing Unit (GPU) and Compute Unified Device Architecture (CUDA) [8]. This task had been implemented in Python programming language with Tensorflow library under Windows 10, using GPU NVIDIA GeForce 740M (version 417.35) with CUDA 9.0, 2G video memory and 4G Random Access Memory (RAM).

B. SSD MobileNet

Single Shot Multibox (SSD) model with the MobileNet detector and classifier had been combined together and pre-trained on the COCO (Common Objects in Context) dataset. This dataset contains 2.5 million labeled instances in 328k images (91 object categories) [13].

1) *MobileNet*: Had been introduced for mobile and embedded vision applications in [14], which is based on depthwise separable convolutions. Basically, MobileNet is a class of lightweight deep convolutional neural networks that are vastly smaller in size and faster in performance than many other popular models. Depth-wise separate convolutions first apply a single filter on each input to filter the input data, followed by 1×1 convolutions which combine these filters into a set of output features. These depth-wise separable layers almost mimic the function of typical convolution layers but with much faster speed and with a slight difference (typical convolution filters and combines into output features both, but in depthwise separable convolution this is divided into two layers, one separate layer for filtering and one separate layer for combining). This minimizes the model size and reduces computational power demands. All layers are followed by a batchnorm and ReLU nonlinearity except the final fully connected layer which feeds into a softmax layer for classification having no nonlinearity. MobileNet has 28 layers without Counting depthwise and pointwise convolutions. Figure 1 demonstrates the general architecture of MobileNet.

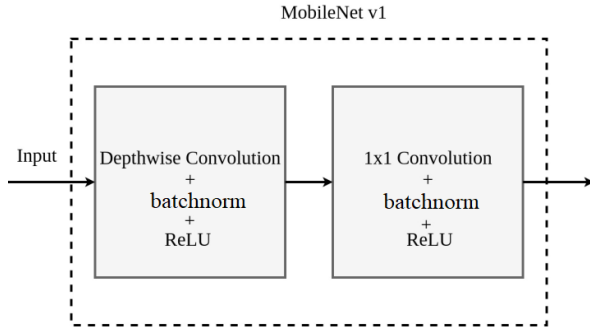


Figure 1: MobileNet Architecture

2) **SSD**: Had been proposed for object detection in [15]. The ideas behind many networks like RPN in faster R-CNN, YOLO and multi-scale CONV features have been combined in SSD to achieve fast detection speed while maintaining high detection quality [16; 15]. In SSD, for the precedence of each object category inside each bounding box, a score is generated, followed by an adjustment step to the box to better match the object shape/ the final detection. The CNN network in the SSD is fully convolutional, based on VGG-16 as a base network. Followed by, many auxiliary conv layers which are progressively decreasing in size. In order to detect small objects, SSD uses more shallow layers with higher resolution. For objects having different sizes, SSD performs multiple scales detection by operating on multiple conv feature maps, each of them for bounding boxes of proper sizes predicts category scores and box offsets Figure 2 demonstrates a high-level diagram of SSD for generic object detection.

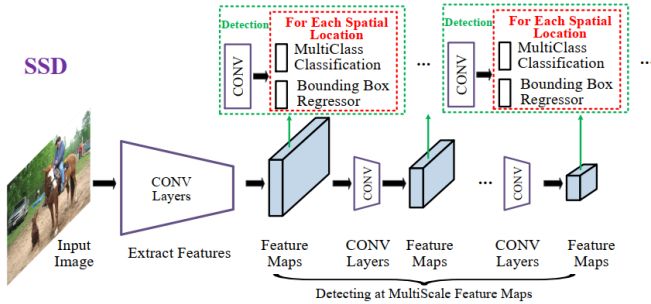


Figure 2: High-level diagram of SSD [16] for generic object detection

C. Faster-RCNN Inception

Faster R-CNN model with the Inception module had been combined together and pre-trained on the COCO (Common Objects in Context) dataset.

1) **Faster R-CNN**: This is a Region proposal Network (RPN) introduced in [17] for object detection purposes which depend on region proposal algorithms to predict object locations. They used a single network for the task of RPN which generates region proposal operation and Fast R-CNN classifies regions. In Faster R-CNN, the RPN shares full-image convolutional features with fast R-CNN, this enables almost cost-free region proposals. An RPN is a fully-convolutional network that predicts object bounds and scores at the same time simultaneously and this makes Faster R-CNN a purely CNN

based model no handcrafted features been used. The high-quality region proposal which is generated by RPNs after being trained end-to-end is used by Fast R-CNN for region detection. An RPN takes an input image having any size and gives a set of rectangular object proposals with an object score for each as an output. First $n \times n$ reference boxes (Each sliding window) are initialized by RPN that at each conv feature map location has different scales and aspect ratios. Each sliding window is mapped to a lower dimensional vector (such as 256 for Zeiler & Fergus model and 512 for VGG), which is then fed into 2 siblings fully-connected layers (box classification layer and a box regression layer). To the output of the $n \times n$ conv layer ReLUs are applied. For the very deep VGG16 model, Faster R-CNN is able to test at 5fps (including all steps) on a GPU, while achieving state-of-the-art object detection accuracy on PASCAL VOC 2007 using 300 proposals per image. Figure 3 demonstrates the high-level diagram of Faster R-CNN for generic object detection.

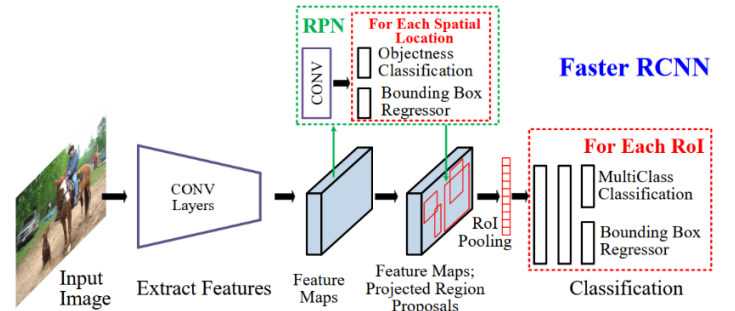


Figure 3: High-level diagram of Faster R-CNN [16] for generic object detection

2) **Inception-v2**: The Inception model is a pre-trained deep convolutional neural network having millions of parameters that had been trained on ImageNet dataset using very powerful and expensive computers, training this model takes weeks to finish. The first version of Inception (inception-v1) was introduced in [18] as GoogleLeNet. Later the architecture of inception was improved in [19] using the batch normalization technique and different other ways and the new version (Inception-v2) came out. In which based on the factorization technique 7×7 convolution have been factorized into three 3×3 convolutions to improve computational speed. There are 3 traditional inception modules in the inception part of the network and using the grid reduction technique the 35×35 grid with 288 filters each is reduced to a 17×17 grid with 768 filters. Followed by 5 instances of the factorized inception modules. Using grid reduction it is reduced to an $8 \times 8 \times 1280$ grid. There are two Inception modules at the coarsest 8×8 level, and for each tile an output filter bank having a size of 2048 is concatenated. remove the bottleneck filter banks in the module made wider instead of deeper.

D. Dataset

The images of healthy and diseased grapes were collected from the internet. A total number of 136 healthy grapes images and 124 diseased grapes images were obtained. Some of the images of grape leaves were collected from the internet and some were collected from the PlantVillage dataset [1]. A total

number of 201 images of healthy grape leaves and 195 images of diseased grape leaves were collected. So there is a total number of 656 number of prepared images for this task. Then these were divided into two folders, a total number of 543 images for training and 113 number of images for testing. Then using LabelImg software [20] all these images had been labeled into 4 classes (Healthy Grape (HG), Diseased Grape (DG), Healthy Grape Leaf (HGL), Diseased Grape leaf (DGL)) one by one. This software creates a .xml file which then had been converted into a .csv file in order to be read by the Tensorflow. Finally, two .csv files had been created named training.csv and testing.csv which contains information about all the images and their corresponding classes.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

As mentioned before the transfer learning approach is used for this task which using a pre-trained model for another task. At first label map is created which would tell classifier what each image contains by defining a mapping class ID numbers (Healthy Grape:1, Diseased Grape:2, Healthy Grape Leaf:3, Diseased Grape Leaf:4). Then some parameters of both models (SSD_MobileNet v1 and Faster-R-CNN Inception v2) had been altered by parameters according to our task. In the configuration files of both models the `eval_input_reader` (points to testing data), `num_examples` (number of testing images), `num_classes` (number of output class), `train_input_reader` (points to training data) and `fine_tune_checkpoints` sections had been altered according to this task. The batch size in the SSD_MobileNet v1 model is changed from 24 to 1 because it was exceeding the memory size of our computer.

A. Training:

Faster-R-CNN Inception v2 had been trained for 22 hours which was a total number of 6981 steps of training. This model was training slow, it was taking approximately 9 to 17 sec per image to process them. But it was training pretty good, the loss values started high and it was decreasing drastically. Figure 4 demonstrates the Total Losses graph generated for Faster-R-CNN Inception v2 by TensorBoard while training the model in this task.

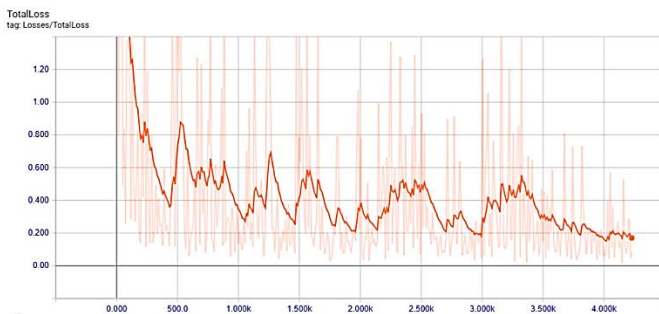


Figure 4: Total Losses of Faster-R-CNN Inception v2

SSD_MobileNet v1 had been trained for 12 hours which was a total number of 16229 steps of training. This model was training pretty faster than Faster-R-CNN Inception v2, it was taking approximately 3 to 2 sec per image to process them. But

it wasn't training well, the loss values were higher too and it was taking a lot of steps to decrease the loss values, in other words, the loss values were decreasing very late (it was taking thousands of steps to decrease a bit) that is why this model had been trained with more steps than the previous model to get better results. Figure 5 demonstrates the Total Losses graph generated for SSD_MobileNet v1 by TensorBoard while training the model in this task.

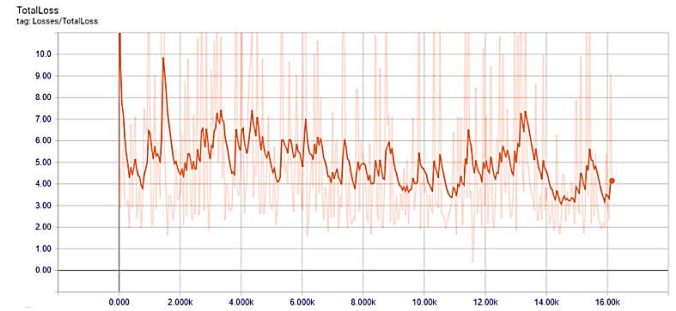


Figure 5: Total Losses of SSD_MobileNet v1

B. Testing:

The models had been tested on 113 images in which 26 of them contains HG images, 16 of them contains DG images, 41 of them were HGL images and 30 of them were DGL images. To examine that how these models will perform in real time, it's been tried to keep these images as natural as possible, which means that the images were in different sizes having different resolutions taken by different cameras. Some of the images had uniform backgrounds while some had very noisy backgrounds.

First, all the images had been tested on the Faster-R-CNN Inception v2 model one by one. And the results were pretty good. This model classified almost **95.57%** of all the testing images correctly. As shown in Figure 6 this model was able to classify a high percentage of images correctly from total testing images. In which it was able to classify 92.31% of HG images, 100% of HGL images, 93.75% of DG images and 93.30% of DGL images correctly from a total of them each.

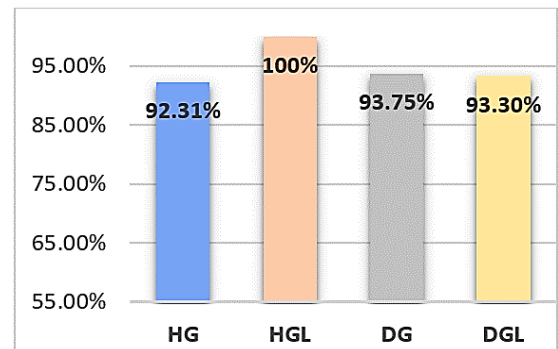


Figure 6: Testing Results of Faster-R-CNN Inception v2

The classification accuracy obtained using the Faster-R-CNN Inception v2 model was between **78% to 99%**. This model had pretty high classification accuracy but with a bit slow processing time, the average processing time of this model was between 25 to 30 seconds per image. A demonstration of

the classification accuracy of this model during testing is presented in Figure 7.

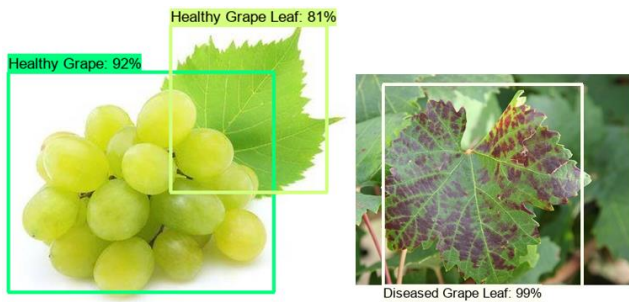


Figure 7: Testing accuracy of Faster-R-CNN Inception v2

After that, the same 113 testing images had been used to test SSD_MobileNet v1 model one by one. But the results which this model showed were not good enough. This model classified almost **59.29%** of all the testing images correctly with better processing time, which means it was taking almost 14 to 15 seconds to process each image. As shown in Figure 8 this model was able to classify a lower percentage of images correctly from total testing images. From which it was able to classify 88.46% of HG images, 85% of HGL images, only 1% of DG images and 30% of DGL images correctly from a total of them each.

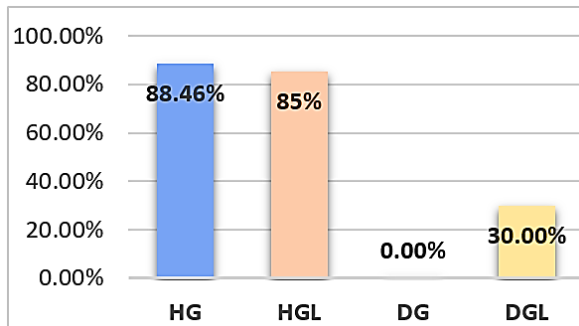


Figure 8: Testing Results of SSD_MobileNet v1

As it's been observed that, this model was able to classify images collected from the PlantVillage dataset [1], which are in the pretty good format, in other words, the images which had uniform backgrounds and also had the same size and resolution. The classification accuracy of this model for most of those images very pretty good which means it was between **90% to 99%** of accuracy. But for the images which contain noise, have different backgrounds and resolutions it was giving very poor classification accuracy which was between **52% to 80%** alongside a high percentage of the wrong classification. This can lead to poor results during real-time testing. A demonstration of the classification accuracy of this model is presented in Figure 9.

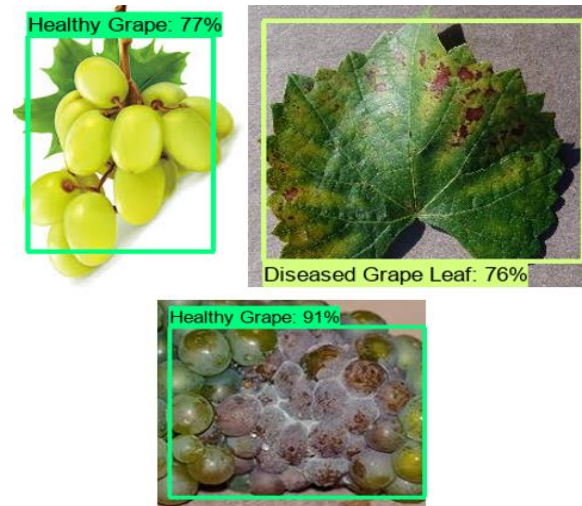


Figure 9 Testing accuracy of SSD_MobileNet v1

C. Discussions:

As shown in Table 1, the Faster-R-CNN Inception v2 model was able to classify testing images better than the SSD_MobileNet v1 model. This means the Faster-R-CNN Inception v2 model correctly classified 108 number of images out of 113 images with a very high classification accuracy, even though the images were in different shapes, sizes, resolutions, and backgrounds. But this model was taking a bit longer time to process each image. This means that this model can be quite useful for real-time classification of images for this task.

On the other side, the SSD_MobileNet v1 model wasn't able to classify all the images correctly. It only classified 67 number of images out of 113 images. This model was giving pretty high classification accuracy for the images which are organized, contains lesser noise, have uniform background and sizes. But for the other images classification accuracy was lower and also it had a very high percentage of miss-classification for those images. But this model was processing images faster than other models. According to results obtained for this model, it's been concluded that for now, this model may not be useful for real-time classifications for this task.

Table 1: Both Models Results Comparisons

	Faster-R-CNN Inception v2				SSD_MobileNet v1			
	HG	HG L	DG	DG L	HG	HG L	DG	DG L
Total Images	26	41	16	30	26	41	16	30
Correctly Classified	24	41	15	28	23	35	0	9
Total Images	113				113			
Correctly Classified	108				67			
Total Percentage	95.57%				59.29%			

V. CONCLUSION

Grape and Grape Leaf real-time disease detection has been performed with the images collected from the internet and the PlantVillage dataset. This task is done using Tensorflow object detection API on two pre-trained deep learning models, Single Shot Detector (SSD) MobileNet v1 and Faster R-CNN Inception v2.

Faster-R-CNN Inception v2 model classified almost **95.57%** of all the testing images correctly with a classification accuracy of **78% to 99%**. The accuracy level of this model was higher but with longer processing time. The average processing time of this model was 25 to 30 seconds per image. This means that this model can be quite useful for real-time classification of images for this task.

On the other hand, the SSD_MobileNet v1 model wasn't able to classify all the images correctly. This model only classified almost **59.29%** of all the testing images correctly but with better processing time, means it was taking almost 14 to 15 seconds to process each image. This model was giving pretty high classification accuracy for the images which are organized, contains lesser noise, have uniform background and sizes. It was between **90% to 99%** accuracy. But for the images which contain noise, have different backgrounds and resolutions it was giving very poor classification accuracy which was between **52% to 80%** alongside a high percentage of misclassifications. According to these results, it's been concluded that for the current time this model may not be useful for real-time classifications for this task.

REFERENCES

- [1] D. P. Hughes, and M. Salathe. (2015). An Open Access Repository of Images on Plant Health to Enable the Development of Mobile Disease Diagnostics. *CoRR*. arXiv:1511.08060. [Online]. Available: <https://arxiv.org/abs/1511.08060>, Accessed: 24 March 2019.
- [2] N. R. Kakade, and D. D. Ahire, "Real Time Grape Leaf Disease Detection", *International Journal of Advance Research and Innovative Ideas (IJARIIE)*, 1(4) pp. 598-610, 2015.
- [3] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification", *Computational Intelligence and Neuroscience*, 2016 pp. 11, 2016.
- [4] S. P. Mohanty, D. P. Hughes, and M. Salathe, "Using Deep Learning for Image-Based Plant Disease Detection", *Frontiers in plant science*, 7 pp., 2016.
- [5] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition", *Sensors*, 17(9) pp. 2022. MDPI AG, 2017.
- [6] S. Mude, D. Naik, and A. Patil, "Leaf Disease Detection Using Image Processing for Pesticide Spraying", *International Journal of Advance Engineering and Research Development*, 4(4) pp. 1129-32, 2017.
- [7] S. S. Kothawale, S. R. Barbade, and P. P. Mirajkar, "Grape Leaf Disease Detection Using Svm Classifier", *International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)*, 6(4) pp. 3288-94, 2018.
- [8] A. K. Rangarajan, R. Purushothaman, and A. Ramesh, "Tomato Crop Disease Classification Using Pre-Trained Deep Learning Algorithm", in *Conf. International Conference on Robotics and Smart Manufacturing (RoSMa2018)*. pp. 1040-47, 2018.
- [9] M. Horea, and O. Mihai, "Fruit Recognition from Images Using Deep Learning", *Acta Universitatis Sapientiae, Informatica*, 10(1) pp. 26-42, 2018.
- [10] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola. (2018). Dive into Deep Learning. [Online]. Available: <http://en.diveintodeeplearning.org>, Accessed: 19 March 2019.
- [11] M. Mehdipour-Ghazia, B. A. Yanikoglu, and E. Aptoula, "Plant Identification Using Deep Neural Networks Via Optimization of Transfer Learning Parameters", *Neurocomputing*, 235 pp. 228-35, 2017.
- [12] "Tensorflow Detection Model Zoo", (2019). [Online]. Available: https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md, Accessed: 20 March 2019.
- [13] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft Coco: Common Objects in Context", in *Conf. Computer Vision – ECCV 2014. Springer International Publishing*, pp. 740-55, 2014.
- [14] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. (2017). Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861. [Online]. Available: <https://arxiv.org/abs/1704.04861>, Accessed: 18 March 2019.
- [15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single Shot Multibox Detector", in *Conf., Springer International Publishing*, pp. 21-37, 2016.
- [16] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen. (2018). Deep Learning for Generic Object Detection: A Survey. arXiv:1809.02165. [Online]. Available: <https://arxiv.org/abs/1809.02165>, Accessed: 21 March 2019.
- [17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-Cnn: Towards Real-Time Object Detection with Region Proposal Networks", *Advances in Neural Information Processing Systems 28 (NIPS)*, pp. 91-99, 2015.
- [18] C. Szegedy, L. Wei, J. Yangqing, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions", in *Conf. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 1-9, 7-12 June 2015.
- [19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision", in *Conf. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 2818 - 26, 27-30 June 2016.
- [20] "Labelimg", (2018). [Online]. Available: <https://github.com/tzutalin/labelimg>, Accessed: 13 February 2019.

Improved Network Intrusion Detection System Using Deep Learning

M. RIFAAT¹ and M. GÜNDÜZ¹

¹ Konya Technical University, Konya/Turkey, mustafarifaat539@yahoo.com

¹ Konya Technical University, Konya/Turkey, mgunduz@ktun.edu.tr

Abstract - Intrusion Detection Systems is considered a main factor in information security. With the increasing of Electronic attacks, the safety of information security becomes an important affair of all over the world. The classical machine learning models (e.g. neural networks (NN), random forests (RF) and support vector machines (SVM), etc.) have produced good intrusion detection rates and low false alarm rates for computer networks for unknown attacks in the network. Recently, the development of machine learning led to discovering deep learning. Deep learning networks are presenting better options than the classical machine learning models to train various classifiers and get the best rates in intrusion detection. In this study, a Convolutional Neural Network - Network Intrusion Detection System (CNN-NIDS) model is proposed for improving intrusion detection in computer networks. The proposed model involves network traffic data, extracted from the NSL-KDD dataset, which has been converted to image-types to train a Convolutional Neural Network (CNN). The assessment of the results is made using a binary classification system which includes two options; Attack and Normal. The results present accuracies of up to 80.20% compared with 75.75% of a Deep Neural Network (DNN).

Keywords - Information security, convolutional neural network, deep learning, network intrusion detection system.

I. INTRODUCTION

DESPITE more people and societies increasingly joining and using the Internet for various life-changing tasks, the Internet platform is on the other hand posing great security threats. In recent years a variety of different type of attacks has emerged within cyberspace. Numerous research efforts have been carried out to combat these attacks. Network Intrusion Detection System (NIDS), that is implemented using machine learning algorithms [1], is one of the measures that have been proposed to detect computer network intrusions both in industry and academia. Security experts continue to seek solutions that would offer better NIDS performance that presents the highest detection rates and low false alarm rates [2, 3].

Improving the performance of machine learning models can be carried out by developing better or fine-tuning parameters of these models [4, 5]. Deep learning models that were proposed by Hinton [6] have led to rapid development in the field of machine learning. Utilization of these models can greatly improve the success of machine learning algorithms consequently improving the efficiency of NIDS. The deep learning theories and technologies have promoted a very rapid

development that has opened up newer artificially intelligent intrusion detection system's solutions [7].

Intrusion detection considered as a classification issue, i.e. a binary classification problem. This usually involves identifying if the network traffic behavior is a normal or their attack has been detected. The base target of the network intrusion detection systems is improving the precision of the classification problem and effectively identify or predict a possible network anomaly behavior. CNN has recently become a widely used deep learning model of choice, especially for classification problems. CNNs have proven very effective for utilization in the wide variety of applications like image classification, object detection, visual recognition, speech recognition and Natural Language Processing (NLP) [8]. Moreover, these models have enabled successful industrial applications such as video indexing, face recognition, intelligent surveillance, people tracking and multimedia understanding, image annotation and powering self-driving cars and robot vision [9]. This can be attributed to the ability of CNN to extract best feature representations of the input data.

The deep learning method usages are increasing rapidly, by the continuous development of huge data and computing power. In this study, the CNN-NIDS model that's inspired by the success of CNN models is proposed for the classification of computer network traffic for a network intrusion detection system. CNN model is trained using artificially generated image-type data from the NSL-KDD dataset. The performance of the CNN-NIDS model was compared with Deep neural network (DNN) proposed in [10].

II. RELATED STUDIES

Deep learning approaches have shown to present better results than classical machine learning approaches. Javaid et al. [11] utilize a Self-Taught Learning model that employed a deep learning algorithm to classify the unauthorized network signals using the NSL-KDD data set. The classification task achieved accuracies of up to 79.10% in a 5-class classification task.

Tavallae et al. [12] presented a study of several machine learning algorithms for the network Intrusion Detection System (IDS). In this analysis, the authors also use NSL-KDD data set. And provide a clear description and guidelines that are needed to understand intrusion detection systems. In this study, results showed that the best accuracy rate achieved was 82% with NB-Tree. Other machine learning algorithms such as Random Tree

have 81% accuracy and Multilayer Perceptron has 77% accuracy.

Panda and Patra [13] propose experimental results in 10% from the KDD'99 datasets that included 65,525 connections. Using a naïve Bayes classifier model. The IDS tested the activities in a system for suspicious behavior or patterns that inform whether a system attack has occurred on the computer network. The machine learning algorithm's performance is measured by checking on how it detects the 5 classes presented in the dataset. These classes include DOS, Probe, U2R, R2L and normal. The authors' model presents a detection rate of up to 95%, with an error rate of 5%.

Shone et al. [14] proposed a non-symmetric deep auto-encoder for unsupervised feature learning, for network intrusion detection. To handle the intrusion detection system through the experiment on two databases, the first one is KDD cup 99 Dataset where they used 10% the accuracy results were 97%. The second database was NSL KDD used in 13 classes and the accuracy results were 89%.

Tang et al. [10], used a deep neural network for flow-based anomaly detection, as a deep learning based approach. The accuracy results were 75.75% when applied on both the training & testing dataset of 6 features from NSL KDD full dataset. The current work has been compared with this work.

III. THE PROPOSED METHOD

The main idea behind Convolutional Neural Network is to learn complex features from pixel-level contents, capitalizing on a sequence of operations such as normalization, altering, pooling, etc.

There are four main operations that are typical of CNN. These operations include; convolutional operations, rectifier linear units, pooling and operations on the classification layer of the CNN model. These operations are carried out using three essential layers in the network architecture. The first layer is convolutional, the second layer is pooling and the third layer is the classification. For example, LeNet5 [15] architecture is one of the classic CNN which can be used to understand how these layers interconnect. The architecture of LeNet5 is shown in Figure. 1.

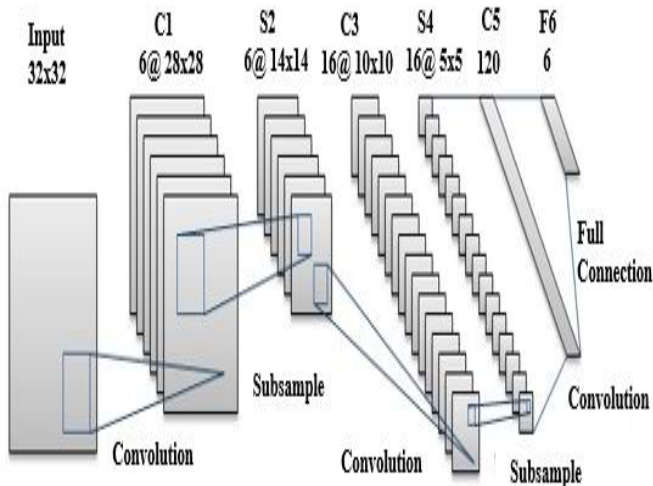


Figure 1: LetNet5 CNN architecture.

The proposed method implements a CNN for the NIDS classification problem. As it is presented in Figure. 2, (3 convolution layers, 1 max pooling Layer, 2 average pooling layers, and 3 fully connected layers used) the result is a two-class classification (normal & attack).

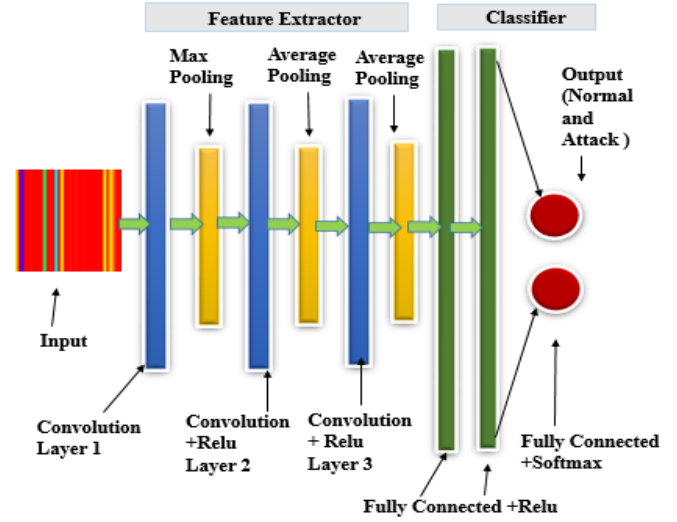


Figure 2: Proposed structure.

The convolutional layer used to extract features from the input image and outputs them as feature maps. This operation preserves the spatial data in the image pixels through learning features by using kernels. A kernel or filter in CNN is like a matrix or a small window that is slide over the input image (as shown in Figure 3). The convolutional layer computes the dot product of the corresponding image pixels and produces an activation map or a feature map. The filter act as a feature detector from the image and various values of feature matrices will output various feature maps.

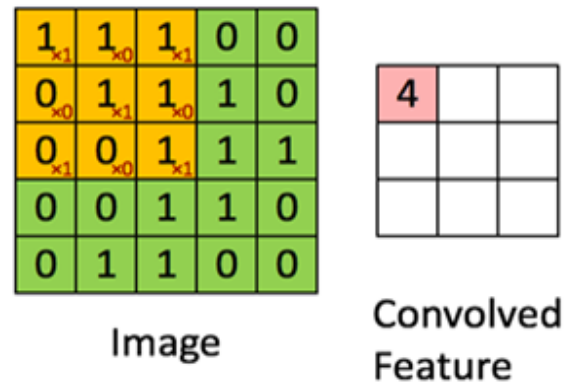


Figure 3: 3x3 Filter on a 5x5 input image.

The pooling layer operation can be performed using various techniques. This can include the average pooling or maximum pooling methods. Average & maximum pooling help to decrease the dimensionality of the generated feature maps and are commonly employed due to their computational capacity. Maximum pooling has shown to improve shift invariance within the feature maps developed from convolving through the

input image and other feature maps as inputs at each convolutional layer [16].

The average pooling method calculates the mean of all the feature map values in the window. Whereas, in each window, the max pooling only selects the maximum value as shown in Figure 4. This method can ensure that most of the important information is captured from the feature maps.

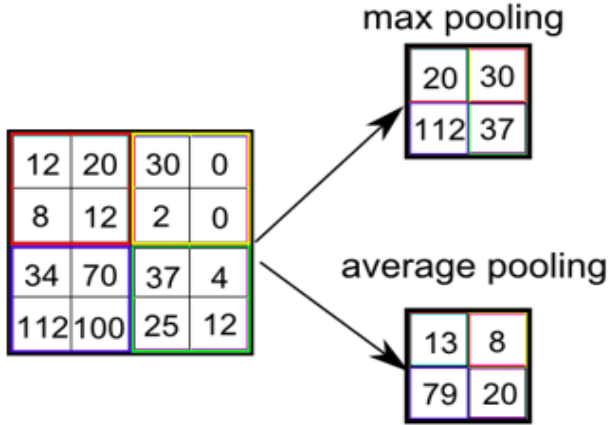


Figure 4: Max and average pooling.

The classification layer is made up of a layer of fully connected. The fully connected layer is the Multi-Layer Perceptron with an activation function like the Rectifier Linear Unit [17]. In this layer, every neuron in each layer is linked to neurons in the successive layers. The features extracted from the input layer by the convolutional layer is fed to the classification layer. A softmax classifier can be attached at the end of the classification layer before connecting it to the output layer that will be used to identify the class that an input image is found [18].

IV. DATASET OVERVIEW

The (NSL-KDD) dataset is reduced and improved version of the DARPA (KDD Cup 99) dataset [12]. DARPA KDD Cup99 is one of the world's leading network intrusion detection competitions. However, recently it has been found that the DARPA KDD CUP99 dataset had some problems that create bias in the results obtained by the learning algorithm. These network traffic data were inaccurate due to the repetition of its records both within the training & testing datasets. On the other hand, this dataset was used in the previous to evaluate the performance of a network intrusion detection system by researchers and thereby presented significant results.

The NSL-KDD dataset contains: (1) KDD Train+ is the training dataset that has 125973 network traffic samples and 23 attack types that are grouped as shown in Table 1. (2) KDD Test+ is the test dataset contains 22554 network traffic samples and 39 attack types shown in Table 2. There are 41 features; 3 of these features are non-numeric features and 38 are numeric features. There are 4 attacks from the NSL-KDD data set that include User to Root (U2R), Denial of Service (DoS), Probe &

Remote to Local (R2L). U2R are types of attack where a user who is allowed to enter the system, but is not an administrator tries to do things that require administrator permission. The DOS attack occurs when an attacker tries to prevent legitimate users from using a service. Probe attacks have been developed to learn the valid IP addresses, active ports, or operating system of any machine on a server side. Finally, R2L in this type of attack, the attacker exploits some security holes in the network to access as a local user of this device and send packets to the machine over a network. The types of attacks were shown in table 1 and table 2.

Table 1: Attack types in KDD train+ the NSL-KDD dataset.

Category	Attack name in KDD train + dataset
DoS	Pod, Back, Smurf, Teardrop, Land, Neptune
U2R	Loadmodule, Perl, Rootkit, Buffer-overflow
Probe	Satan, Nmap, Portsweep, Ipsweep
R2L	Spy, Ftp-write, Phf, Gess_pass, Multihope, Warezclicent, warezmater, Imap
Normal	Normal

Table 2: Attack types in KDD test+ the NSL-KDD dataset.

Category	Attack name in KDD test + dataset
DoS	Smurf, Land, Teardrop, Pod, Back, Neptune, Processtable, Mailbomb, Worm, Apache2, Udpstorm,
U2R	Perl, loadmodule, Buffer-overflow, Rootkit, Sqlattack, Ps, Xterm,
Probe	Saint, Ipsweep, Portsweep, Satan, Mscan, Nmap,
R2L	Fpt-write, Multihop, Guess-passwd, Snpmpguess, Spy, Warezmater, Xsnoop, Xlock, snmpgetattack, httptunnel, Sendmail, Named, Imap, Phf
Normal	Normal

V. DISCUSSION AND RESULTS

In this study, the experiments designed to analyze the performance of the CNN-NIDS model. A 2-class classification (Normal, Attack) is implemented to evaluate the effectiveness of the model to detect network attacks. Several authors have shown the ability to convert numeric network traffic data into image type data [19, 20]. Using these image conversion techniques, images are artificially generated from the NSL-KDD dataset to obtain image-type data as shown in Figure 5. These images are used to train the CNN-NIDS model. The experiments are performed using Matlab R2017b, Windows 10, Intel Core i7-8700k CPU @ 3.70 to 4.50 GHz, RAM 16 GB @ 3000GHZ & GTX 1060 NVidia - 6GB GPU. The values from numeric features were used to generate images of size 256*256 pixels. The numeric features in Table 4 are directly converted to image-type. However, the non-numeric features are converted into numerical data before image-type conversion as shown in Table 3.

Table 3: Corresponding values of non-numeric data.

Protocol name	Value
ICMP	0
TCP	1
UDP	2

NSL-KDD data consist of the 5-class that include 4 attack types (DoS attack, U2R attack, Probe attack, & R2L attack) and a normal class from the NSL KDD dataset as Table 5.

Table 4: Features of NSL-KDD dataset [21].

NO	Features	Types	NO	Features	Types
1	Duration	Continuous	22	Is guest	Symbolic
2	Protocol_Type	Symbolic	23	Count	Continuous
3	Service	Symbolic	24	Srv_Count	Continuous
4	Flag	Symbolic	25	Serror_Rate	Continuous
5	Src_Bytes	Continuous	26	Srv_Serror_Rate	Continuous
6	Dst_Bytes	Continuous	27	Rerror_Rate	Continuous
7	Land	Symbolic	28	Srv_Rerror_Rate	Continuous
8	Wrong_Fragment	Continuous	29	Same_Srv_Rate	Continuous
9	Urgent	Continuous	30	Diff_Srv_Rate	Continuous
10	Hot	Continuous	31	Srv_Diff_Host_Rate	Continuous
11	Num_Failed_Logins	Continuous	32	Dst_Host_Count	Continuous
12	Logged_in	Symbolic	33	Dst_Host_Srv_Count	Continuous
13	Num_Compromised	Continuous	34	Dst_Host_Same_Srv_Rate	Continuous
14	Root_Shell	Continuous	35	Dst_Host_Diff_Srv_Rate	Continuous
15	Su_Attempted	Continuous	36	Dst_Host_Same_src_Port_Ra	Continuous
16	Num_Root	Continuous	37	Dst_Host_Srv_Diff_Host_Rat	Continuous
17	Num_File_Creation	Continuous	38	Dst_Host_Serror_Rate	Continuous
18	Num_Shells	Continuous	39	Dst_Host_Srv_Serror_Rate	Continuous
19	Num_Access_Files	Continuous	40	Dst_Host_Rerror_Rate	Continuous
20	Num_Outbound_Cmde	Continuous	41	Dst_Host_Srv_Rerror_Rate	Continuous
21	Is_Host_Login	Symbolic			

Table 5: NSL-KDD data.

Traffic	Training	Testing
Normal	67343	9711
DOS	45927	7458
U2R	52	67
Probe	11656	2421
R2L	995	2887

The NSL-KDD dataset was downloaded and imported in Excel datasheet for preprocessing. The preprocessed data converted to the image types. Then generated images supplied to the CNN model for training & testing of the model. A sample image-type attack is shown in Figure. 5.

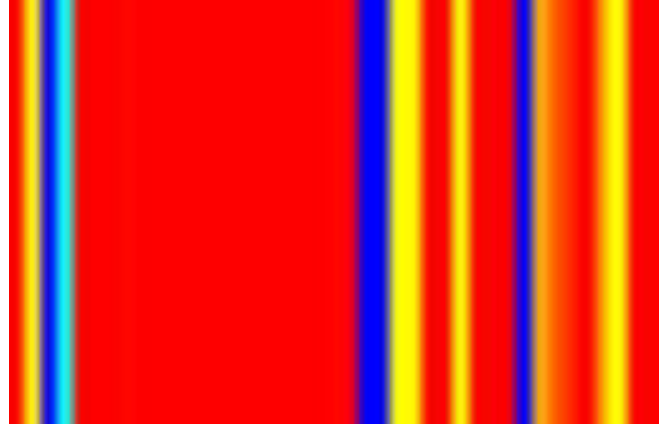


Figure 5: One of the image attacks of NSL-KDD dataset the size images 256x256x3.

In this study a two-class binary classification model implemented (Normal, Attack), the performance of the model compared with deep neural network proposed in [10], it is clear the accuracy value of the CNN approach is better than the accuracy value of the deep neural network. Figure. 6 shows the test accuracy values of CNN and DNN models. The experiment shows that the performance accuracy of the model CNN is 80.20% and the deep neural network is 75.75%, so shown that is the CNN it was better than the deep neural network.

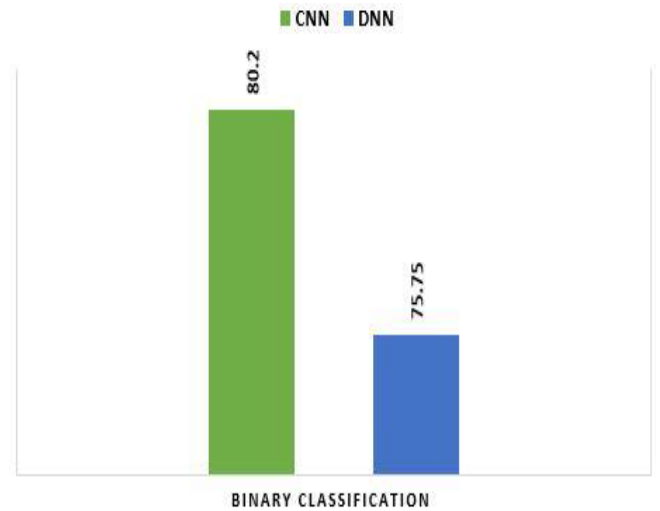


Figure 6: Type of binary classification.

Training accuracy results of the convolutional neural network and the deep neural network [10] are presented in Figure 7. As pointed in the figure the accuracy of this study is %98.20 and the accuracy of the deep neural network is 91.62%. This shows the convolutional neural network (CNN) proposed is better than the deep neural network (DNN).

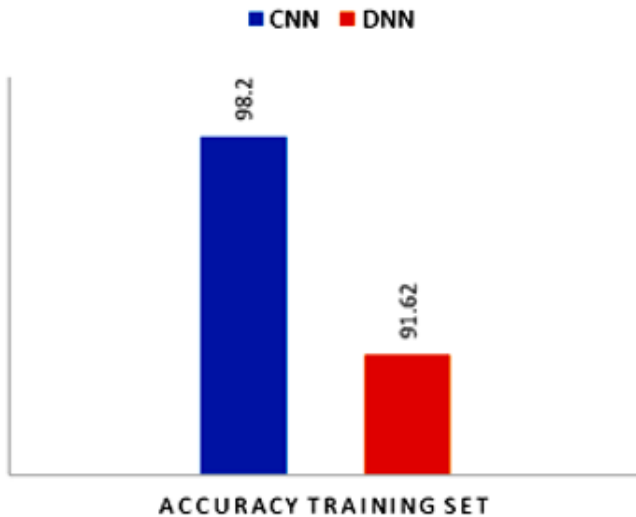


Figure 7: CNN and DNN training accuracies.

VI. CONCLUSION

Computer network threats are increasingly becoming an information security hazard. For this reason, it is necessary today to consider protecting these networks against such vulnerabilities. In this paper, a CNN-NIDS model has been constructed for network intrusion detection systems. This model has successfully applied a convolutional neural network to classify network traffic as normal or a possible attack. The proposed model has been evaluated and compared the results with novel machine learning models such as the DNN on the NSL-KDD dataset. The proposed model was able to achieve accuracies of up to 80.20%. The CNN-NIDS model has therefore shown to be a better model for the network intrusion detection problem as compared with the DNN model.

REFERENCES

- [1] K. A. da Costa, J. P. Papa, C. O. Lisboa, R. Munoz, and V. H. C. de Albuquerque, "Internet of Things: A Survey on Machine Learning-based Intrusion Detection Approaches," *Computer Networks*, vol. 151, pp. 147-157, 2019.
- [2] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85-117, 2015.
- [3] N. Moustafa, J. Hu, and J. Slay, "A holistic review of Network Anomaly Detection Systems: A comprehensive survey," *Journal of Network and Computer Applications*, vol. 128, pp. 33-55, 2019.
- [4] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *Security and Privacy (SP), 2010 IEEE Symposium on*, 2010, pp. 305-316.
- [5] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, 2018.
- [6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, p. 436, 2015.
- [7] A. Mathew, J. Mathew, M. Govind, and A. Mooppan, "An Improved Transfer learning Approach for Intrusion Detection," *Procedia computer science*, vol. 115, pp. 251-257, 2017.
- [8] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, et al., "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354-377, 2018.
- [9] G. Yao, T. Lei, and J. Zhong, "A review of Convolutional-Neural-Network-based action recognition," *Pattern Recognition Letters*, vol. 118, pp. 14-22, 2019.
- [10] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho, "Deep learning approach for network intrusion detection in software defined networking," in *2016 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 2016, pp. 258-263.
- [11] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, 2016, pp. 21-26.
- [12] M. Tavallaei, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, 2009, pp. 1-6.
- [13] M. Panda and M. R. Patra, "Network intrusion detection using naive bayes," *International journal of computer science and network security*, vol. 7, pp. 258-263, 2007.
- [14] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, pp. 41-50, 2018.
- [15] Y. LeCun, L. Jackel, L. Bottou, A. Brunot, C. Cortes, J. Denker, et al., "Comparison of learning algorithms for handwritten digit recognition," in *International conference on artificial neural networks*, 1995, pp. 53-60.
- [16] Y. Gong, L. Wang, R. Guo, and S. Lazebnik, "Multi-scale orderless pooling of deep convolutional activation features," in *European conference on computer vision*, 2014, pp. 392-407.
- [17] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, 2013, p. 3.
- [18] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440-1448.
- [19] L. Mohammadpour, T. C. Ling, C. S. Liew, and C. Y. Chong, "A Convolutional Neural Network for Network Intrusion Detection System," *Proceedings of the Asia-Pacific Advanced Network*, vol. 46, pp. 50-55.
- [20] S. Seok and H. Kim, "Visualized malware classification based-on convolutional neural network," *Journal of the Korea Institute of Information Security and Cryptology*, vol. 26, pp. 197-208, 2016.
- [21] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," *Ieee Access*, vol. 5, pp. 21954-21961, 2017.

Sentiment Analysis for Hotel Reviews with Recurrent Neural Network Architecture

K. M. KARAOĞLAN¹, V. TEMİZKAN², O. FINDIK³

¹ Computer Technologies, Karabuk University Karabuk/Turkey, kkaraoglan@karabuk.edu.tr

² Tourism and Hotel Management, Karabuk University, Karabuk/Turkey, vtemizkan@karabuk.edu.tr

³ Computer Engineering, Karabuk University, Karabuk/Turkey, oguzfindik@karabuk.edu.tr

Abstract - Online marketing platforms have turned into large volumes of information and opinion for customers with the transition to Web 2.0. Customers refer to these resources in order to obtain information before they purchase a product and to reach the potential views of others about possible experiences. Businesses also need customer feedback to improve the services they provide and to explore which reviews are more valuable product specifications. In this study, Sentiment Analysis (SA) was performed with 2-pole (positive-negative) classification about hotel businesses on an opinion dataset created by users. Deep Learning based Recurrent Neural Network (RNN) architecture was used in these analyzes. With the results of the RNN architecture, the results of the classification based on score conditional and editorial interpretation were compared and it was observed that the performance of classification with RNN was successful.

Keywords - Sentiment Analysis, Recurrent Neural Network, Classification, Natural Language Processing, Deep Learning.

I. INTRODUCTION

SOCIAL networks that become information and opinion sources has radically changed ways of expressing ideas and feelings of people using web communities, blogs, wikis and other online media platforms with the transition to Web 2.0. [1]. This situation has led to the emergence [2] of ideas mining and SA, which are interested in acquiring subjective and useful information from the texts used in data mining and Natural Language Processing (NLP) techniques.

SA [3-4] is a method used to determine the ways in which sentiment are expressed in texts and whether these expressions contain positive or negative opinions on a particular product or service. For instance [5], consumers need reviews from others to learn about possible shopping experiences before buying a product or service.

Businesses need consumer reviews to improve the quality of services they provide and to explore which aspects are more valuable in their product characteristics [6,7]. In this context SA, market intelligence [8] has helped public opinion and [10] decision-making processes on many issues [9] such as customer satisfaction measurement.

A properly constructed SA system [5] can eliminate the need for surveys and change the way traditional market research is conducted. SA mainly focuses on the polarity of ideas [11], in other words the automatic classification of positive or negative. At sentence level, sentiment [12] expressed for each sentence are classified. This classification [13] first determines whether the sentence is subjective or neutral. If there is a subjective sentence extraction, SA calculates [11] whether the sentence is positive or negative.

According to Balazs [14], approaches to basic NLP techniques, based on the rules obtained from grammar [98], uncontrolled dictionary-based (semantic-based) approaches and controlled machine learning approaches. In SA [15], Naive Bayes for polarity classifications, Support Vector Machiner (SVM) classifiers are stated as effective machine learning models.

Gamon et al., [16] presented a controlled learning approach to the classification of sentiments on a data set of film and product evaluations. In Hu and Liu studies [17], have been identified as a product of a product frequently reviewed by customers. In the next stage, sentences containing ideas were determined and their polarities were found. Finally, the results are summarized according to their weight.

Marrese-Taylor et al., [18] they have achieved high accuracy results by working on TripAdvisor reviews by Bing Liu [19]. You and colleagues [20], with the Deep Learning-based RNN architecture, have presented an approach for security auditing to classify short messages in prisons as safe and unsafe.

In this study, of which Turkey is a high tourism potential, Karabuk, one of the destinations for business hotel located in Safranbolu district, the Turkish reviews made by the customers on the website (otelpuan.com) and the satisfaction scores were used as data set in SA. 2 polarity (positive-negative) classification was performed on the data set with RNN architecture. The results obtained with RNN classifier were compared with static classifications based on financial and editor. In the second part of the study, detailed information about the model data set and RNN architecture is given. In the third chapter, analyzes of RNN and other static classification results are presented. In the fourth chapter, evaluation and results are given.

II. DATASET AND METHOD

A. Dataset

In this study, the customer reviews and points of the hotels in the Safranbolu district of Karabük province, which is on the UNESCO world heritage city list, were taken from the website and data set was created. To ensure the reliability of the SA results to be taken on this dataset, corrupt and missing data has been extracted. The descriptive statistical data for the data set are presented in Table 1.

Table 1: Descriptive statistics of dataset.

Total Count of Reviews	Length of Word Index	Count of Tokens Without Stopwords
665	2747	9353
Mean Token Count of Reviews	Max. Word Count of Tokens	Standard Deviation of Tokens
15,13	130	13,97

B. Method

The simple RNN [21] was originally designed by Jeff Elman and published in the article 'Finding structure in time'. Therefore, a simple RNN is known as "Elman Network".

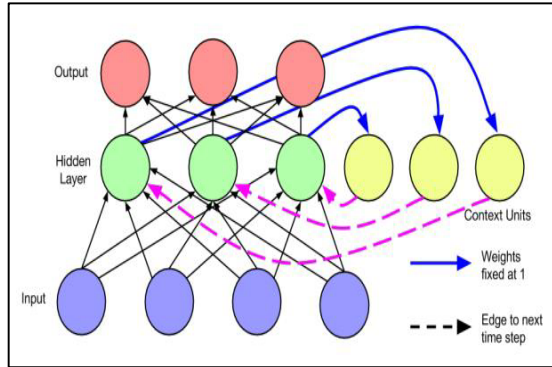


Figure 1: RNN as described by Elman [26].

Figure 1 presents a graph diagram of Elman Network. In this network, hidden layer nodes are connected to the content unit. In the next step, it is linked to the hidden layer as feedback. RNN [22,23], a special type of neural network, has been widely used as a powerful tool because of its superior performance in parallel processing and recursive calculation. RNN [26] is a model that introduces the concept of time to a forward-fed neural network. RNN architecture is the class of artificial neural networks where the links between the units form a directed loop. In this architecture, the same task is processed for each element (words in the sentence) of an array based on the previous output. While a Multi Layer Perceptron can only be mapped from input to output vectors, the RNN can in principle be mapped from the history of previous entries to each output.

Feed-Forward Neural Network (FFNN) structures are processed only forward. In this structure, input data is processed from the network and gets an output value.

In RNN, the value of the inputs is evaluated according to the previous inputs. Since the RNN structures use their output as input in the next process, they are separated from the feedforward structures. In this context, RNN structures need memory. Figure 1 shows a typical MLP-specific FFNN and RNN architecture.

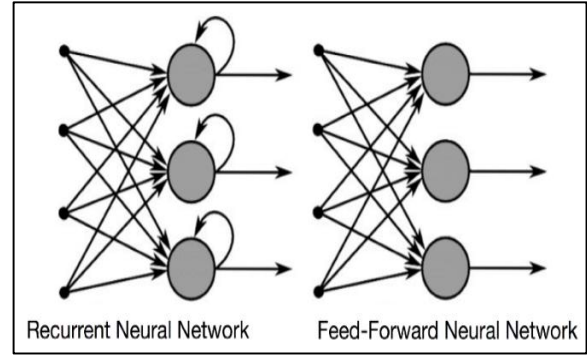


Figure 2: RNN and FFNN in graph form.

Long-Short Term Memory (LSTM) [25] and Gated Recurrent Unit (GRU) [24] networks based on RNN architecture are widely used in NLP applications. The most important reason for this situation is that this architecture is in accordance with the principle of temporal series.

III. SENTIMENT ANALYSIS STUDY

In this study, the static data of the hotel reviews and the dynamic SA performed with RNN architecture are compared. The RNN classifier was divided into 80% training (532 Data) and 20% test data (133 data). RNN architecture was built with GRU structure and 3 hidden layers were used as 32, 16 and 8 nodes. 'Sigmoid' function is used because of its superiority in accuracy performance as an activation function in applied architecture. Binary cross entropy is used as a function of loss in terms of performance. In addition, the model was tested with randomly generated sentences at the end of the studies.

Statically, the data set in the sentiment classification is based on the mean of the arithmetic score given to the interpretations. Editor-based classified data has also been added to test the accuracy of the system in the static data set. All static classification results are tested with RNN. Table 2 presents the statically generated classification cases.

Table 2: Static classified cases.

Case 1	if CASE>5: Sentiment=1 Else: Sentiment=0
Case 2	if CASE>6: Sentiment=1 Else: Sentiment=0
Case 3	if CASE>7: Sentiment=1 Else: Sentiment=0
Case 4	if CASE>8: Sentiment=1 Else: Sentiment=0
By Editor	Given by Editor

Table 3 presents a descriptive table with answers to the system when all data sets are applied to the model. When this table is examined, it is seen that the best performance for test data is obtained for Case1, Editor and Case 2, respectively.

Table 3: RNN performance outputs for all cases.

Initial Case of Sentiment	Loss Value of Training	Accury Value of Training	Loss Value of Testing	Accury Value of Testing	Length of Incorrect Prediction-Ratio
Case 1	0,0268	0,9972	0,0480	0,9925	1- %0,75
Case 2	0,0295	0,9981	0,7036	0,7143	35- %26,3
Case 3	0,0737	0,9850	0,8980	0,6541	40- %30
Case 4	0,1100	0,9756	0,6247	0,7594	42- %31,5
By Editor	0,0410	0,9944	0,3702	0,8722	17- %12,7

Some misclassified examples for all static classifications are given in Table 4. In this table, static classifications and responses received by RNN classifier are compared for misclassified samples. When this table is analyzed, it can be seen that the classification based on points cannot give the correct results for some reviews. However, the performance of the RNN classifier appears to be better for SA.

Table 4: Some examples of RNN outputs and static classification.

Initial Case of Sentiment	Reviews	Predictions	Sentiment Value Due to Prediction	Static Sentiment Value
Case 1	I don't recommend the rooms because it is bad. There was no positive aspect.	0,07645	0	1
Case 2	The rooms were not good at all and according to the requested price I will not stay at the hotel once again. It was an old room. There was a closed car park.	0,13782	0	1
Case 3	The location was good. A clean facility. The rooms were fine.	0,9717	1	0
Case 4	The location was nice. Close to the city. The staff was concerned. Breakfast was nice.	0,90579	1	0
By Editor	The hotel room was old but was nice in terms of hygiene. The hotel staff were friendly people but still not so mediocre as the hotel olmaması gerekiyor.	0,53291	1	0

Finally, in order to measure the accuracy of the system from another perspective, randomize generated examinations are given as inputs to trained models for each case. For this reason, the 10 randomized reviews that will be given to the system are presented in Table 5.

Table 5: Random reviews given to trained model

Index Number	Reviews
1	"The hotel was nice, certainly I would recommend."
2	"The rooms were not clean and our room smelled of moisture.."
3	"The room was not clean. The staff was indifferent."
4	"The hotel is excellent, you can choose this hotel with peace of mind."
5	"The hotel is great but the waiters were delayed in food service and the pool was bad, I don't recommend."
6	"The air conditioner was out of order. it smelled of cigarettes."
7	"The staff was rude. The meals are not tasty."
8	"I've never seen such a bad place."
9	"Everything was smooth and beautiful, but the bed lining were dirty."
10	"The rooms are normal but the staff disinterested. Too bad."

Table 5 shows the RNN-based classifier results of all models presented. These classifier results are shown graphically in Figure 3. When Table 5 is examined, it gives a high rate of accurate classification results in all cases according to the actual sentiment classification result.

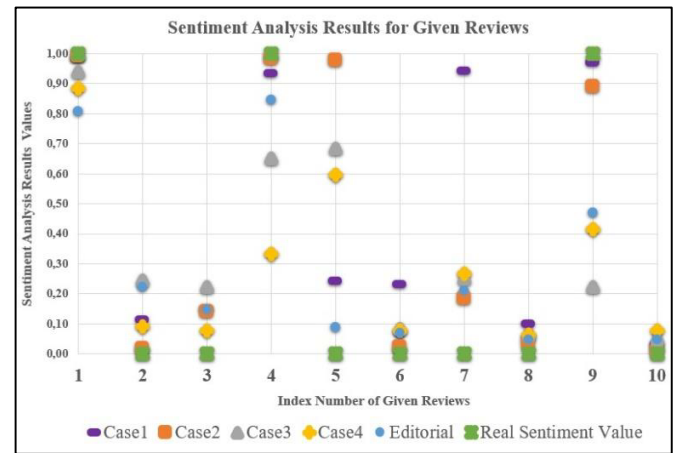


Figure 3: RNN-based classification results including all cases.

IV. RESULTS AND EVALUATION

Customer reviews about products and services constitute a valuable data set for businesses. SA provides the possibility to convert the data set into meaningful information using customer reviews. By evaluating this opportunity, businesses can respond to customer requests quickly by analyzing the target audience. Such predictive analysis helps to understand the potential customer audience at a deeper level.

In this study, RNN architecture-based SA was performed on Turkish hotel reviews of Safranbolu district that has a tourism potential in Turkey. RNN model was applied on points and editor-based static classified data. The classification results of given model were compared with the static classification results and the performance was evaluated. Examples of incorrectly calculated sentiments for the data set for the test are presented with results of RNN architecture. In addition, randomized hotel reviews were applied to the trained model and the results were observed. As a result of the studies performed, it was observed that RNN architecture produces correct results for SA. This study showed that RNN architecture, which was determined in accordance with the 'time series' principle, has been successfully applied in NLP and SA studies based rich data.

REFERENCES

- [1] Cambria, Erik, et al. "An ELM-based model for affective analogical reasoning." *Neurocomputing* 149 (2015): 443-455.
- [2] Giatsoglou, Maria, et al. "Sentiment analysis leveraging emotions and word embeddings." *Expert Systems with Applications* 69 (2017): 214-224.
- [3] Nasukawa, Tetsuya, and Jeonghee Yi. "Sentiment analysis: Capturing favorability using natural language processing." *Proceedings of the 2nd international conference on Knowledge capture*. ACM, 2003.
- [4] Ravi, Kumar, and Vadlamani Ravi. "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." *Knowledge-Based Systems* 89 (2015): 14-46.
- [5] Cambria, Erik, and Bebo White. "Jumping NLP curves: A review of natural language processing research." *IEEE Computational intelligence magazine* 9.2 (2014): 48-57.
- [6] G.L. Urban, J.R. Hauser, "Listening in" to find and explore new combinations of customer needs, *J. Market.* 68 (2) (2004) 72-87, <http://dx.doi.org/10.1509/jmkg.68.2.72.27793>.
- [7] O. Netzer, R. Feldman, J. Goldenberg, M. Fresko, Mine your own business: market-structure surveillance through text mining, *Market. Sci.* 31 (3) (2012) 521-543, <http://dx.doi.org/10.1287/mksc.1120.0713>.
- [8] Li, Yung-Ming, and Tsung-Ying Li. "Deriving market intelligence from microblogs." *Decision Support Systems* 55.1 (2013): 206-217.
- [9] Kang, Daekook, and Yongtae Park. "Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach." *Expert Systems with Applications* 41.4 (2014): 1041-1050.
- [10] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." *Foundations and Trends® in Information Retrieval* 2.1-2 (2008): 1-135.
- [11] Soleymani, Mohammad, et al. "A survey of multimodalRecurrenysis." *Image and Vision Computing* 65 (2017): 3-14.
- [12] Choi, Heeyoul, Kyunghyun Cho, and Yoshua Bengio. "Context-dependent word representation for neural machine translation." *Computer Speech & Language* 45 (2017): 149-160.
- [13] Pang, Bo, and Lillian Lee. "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts." *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2004.
- [14] Balazs, Jorge A., and Juan D. Velásquez. "Opinion mining and information fusion: a survey." *Information Fusion* 27 (2016): 95-110.
- [15] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: sentiment classification using machine learning techniques, *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, vol. 10, 2002. pp. 79-86.
- [16] Aue A, Gamon M. "Customizing sentiment classifiers to new domains: a case study". *International Conference on Recent Advances in Natural Language Processing (RANLP)*, Borovets, Bulgaria, 21-23 September 2005.
- [17] M. Hu, B. Liu, Mining and summarizing customer reviews, in: *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2004, pp. 168-177, <http://dx.doi.org/10.1145/1014052.1014073>.
- [18] E. Marrese-Taylor, J.D. Velásquez, F. Bravo-Marquez, Y. Matsuo, Identifying customer preferences about tourism products using an aspect-based opinion mining approach, *Proc. Comput. Sci.* 22 (2013) 182-191, <http://dx.doi.org/10.1016/j.procs.2013.09.094>.
- [19] Liu, Bing, and Lei Zhang. "A survey of opinion mining and sentiment analysis." *Mining text data*. Springer, Boston, MA, 2012. 415-463.
- [20] You, Lina, et al. "A deep learning-based RNNs model for automatic security audit of short messages." *2016 16th International Symposium on Communications and Information Technologies (ISCIT)*. IEEE, 2016.
- [21] Elman, Jeffrey L. "Finding structure in time." *Cognitive science* 14.2 (1990): 179-211.
- [22] H. Deng, L. Zhang, X. Shu, Feature memory-based deep recurrent neural network for language modeling, *Appl. Soft Comput.* 68 (2018) 432-446.
- [23] G. Manjunath Patel, A.K. Shettigar, P. Krishna, M.B. Parappagoudar, Back propagation genetic and recurrent neural network applications in modelling and analysis of squeeze casting process, *Appl. Soft Comput.* 59 (2017) 418-437.
- [24] Cho, Kyunghyun, et al. "On the properties of neural machine translation: Encoder-decoder approaches." *arXiv preprint arXiv:1409.1259* (2014).
- [25] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.
- [26] Lipton, Zachary C., John Berkowitz, and Charles Elkan. "A critical review of recurrent neural networks for sequence learning." *arXiv preprint arXiv:1506.00019* (2015).

Prediction of Absenteeism at Work with Machine Learning Algorithms

K.KARADAĞ¹ and M.E. TENEKECİ²

¹ Harran University, Sanliurfa/Turkey, k.karadag@harran.edu.tr

² Harran University, Sanliurfa/Turkey, etenekeci@harran.edu.tr

Abstract The increase in competitiveness in business life and the development of economic production depending on expert personnel have increased the importance of trained specialists. However, it is a long-term and costly process to train and develop talented workforce. For this reason, it is very important to use expert personnel effectively in business life. Nevertheless, due to various reasons, employees have being absenteeism to work, causing loss of time and money for companies. Generally, the reasons for absenteeism to work may be unforeseen circumstances such as physical and psychological disease, birth, etc., as well as premeditate times such as legal leave and holidays. Therefore it is very important to be able to predict and manage all these absenteeism cases because of serious costs. In our work, machine-learning algorithms are used to estimate the absenteeism durations. Dataset, which is collected from a specific company in 3 years by A. Martiniano, is used for training and testing of the predictive model. Medium Gaussian Support Vector Machine, Cosine K-Nearest Neighbor and Medium Decission Trees, which are methods of the machine learning algorithms, were used for estimating absenteeism at work. It has been see into that the used methods successfully predicted absenteeism duration. Thus, by determining the absenteeism periods of employees, it will be possible to take precautions in the workplace.

Keywords - Absenteeism at work; Prediction; Cosine K-Nearest Neighbor; Medium Gaussian Support Vector Machine; Medium Decission Trees

I. INTRODUCTION

The human factor is important in all areas as well as in the success or failure of enterprises. Like many factors that adversely affect the purpose of companies, absenteeism is also an important problem that needs to be taken caution [1]. It is important to reduce absences in the effective and efficient use of employees. The most general definition of absenteeism can be defined as the fact that employees do not come to a pre-planned shift. There are two types of absenteeism: these, willingly absenteeism (controllable) and unintentional absenteeism (uncontrollable). Disease caused by factors other than the will of the employee, attendance to the funeral cannot be controlled, Absences with personal and arbitrary reasons without a valid excuse are expressed as checked absences [2].

The main reasons for absenteeism are the disease, the death of one of the family members, the emergence of emergency situations related to the family members, and the stay asleep. Other reasons put forward by the employees are; low job

satisfaction, non-standard working conditions, unfair behaviors of superiors, continuously of boring and monotonous work, and physically unsafe working conditions [3]. Absenteeism creates cost in many ways for organizations. These can be addressed directly and indirectly. Production losses, overtime, reserve personnel and procurement costs expresses to direct costs. Productivity, quality and order losses mean indirect costs. absentee staff, also affect the performance of other employees in the organization negatively [4]. Absence is a loss of wages for companies as well as for employees. There are also some studies that show the positive aspects of absenteeism such as increasing motivation of employees and reducing occupational accidents [5]. When the studies on the causes of absenteeism are examined, Stewartc examined the cost of absenteeism for the economy [6]. Keller has reached the conclusion that his work is an important indicator of the absence of age [7]. Markussen et al. have dealt absenteeism caused by gender [8]. Drakopoulos and Grimani researchs absenteeism and dissatisfaction have dealt the relationship [9]. Martiniano et al. using the same set of data used in our study, they showed the most common causes of absenteeism [10]. Akgeyik conducted a study on the factors affecting the absences of employees [11].

The aim of the study is to take precautionary measures in the selection of personnel and to determine the parameters which cause the absenteeism by taking into account the criteria that cause absenteeism in the enterprises or organizations. Machine learning algorithms were used to make sensitive and accurate predictions.

II. MATERIALS AND METHODS

2.1. Data Set Used in Work

For the training and testing of the prediction model developed, 3-year data from a specific enterprise collected by A. Martiniano were used. In the data, variables were defined as 21 properties. Table 1 is provided features that enable us to obtain information about absenteeism. Another feature is the group of diseases. This group of diseases has been evaluated among 28 categories [10].

Table 1: Feature information causing absenteeism.

Number	Attribute Information
1	Individual identification (ID)
2	Reason for absence (ICD) (Diseases)
3	Month of absence
4	Day of the week
5	Seasons
6	Transportation expense
7	Distance from Residence to Work
8	Service time
9	Age
10	Work load Average/day
11	Hit target
12	Disciplinary failure
13	Education
14	Son
15	Social drinker
16	Social smoker
17	Pet
18	Weight
19	Height
20	Body mass index
21	Absenteeism time in hours

In Table 2, the international codes of diseases which are shown as the reasons of absenteeism and other diseases than these are included [10]. Absenteeism and come to work states are expressed as zero (0) and one (1).

Table 2: international codes of diseases.

Number	International Code of Diseases (ICD) and Other Diseases
1	Certain infectious and parasitic diseases
2	Neoplasms
3	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
4	Endocrine, nutritional and metabolic diseases
5	Mental and behavioural disorders
6	Diseases of the nervous system
7	Diseases of the eye and adnexa
8	Diseases of the ear and mastoid process
9	Diseases of the circulatory system
10	Diseases of the respiratory system
11	Diseases of the digestive system
12	Diseases of the skin and subcutaneous tissue
13	Diseases of the musculoskeletal system and connective tissue
14	Diseases of the genitourinary system
15	Pregnancy, childbirth and the puerperium
16	Certain conditions originating in the perinatal period
17	Congenital malformations, deformations and chromosomal abnormalities
18	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
19	Injury, poisoning and certain other consequences of external causes
20	External causes of morbidity and mortality
21	Factors influencing health status and contact with health services
22	Patient follow-up
23	Medical consultation
24	Blood donation

25	Laboratory examination
26	Unjustified absence
27	physiotherapy
28	dental consultation

2.2. Recommended System for Absenteeism

The process flow chart for the determination of the absenteeism proposed within the scope of our study is given in Figure 1. Firstly, service time and absenteeism time in hours was proportioned among themselves. then qualitative absenteeism as well as quantitative absenteeism was determined. The new absenteeism status is determined by machine learning algorithms; K-Nearest Neighbor (KNN), Support Vector Machine (SVM) and Decision Trees (DT).

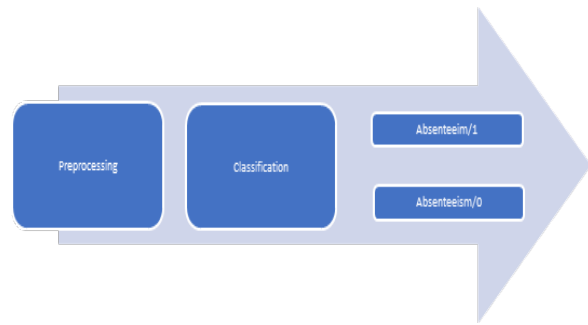


Figure 1: Flow chart followed in the study.

2.3. Methods Used for Absenteeism

Three different classification methods have been tried in the study; KNN, SVM, DT. The basic principle of the KNN Algorithm looks at the k sample classes nearest to the class to be classified and the new data is sent to the highest class. One of the factors affecting the success of the algorithm in the classification process is the correct selection of k. The choice of k value too large or too small can cause the patterns that need to be found in the same class to be assigned to different classes. Determining the most successful k value by trying different k values is the most widely used method [12]. In SVM, It is a controlled machine learning algorithm. In this algorithm, data components are designed as a space in the number of properties. The relationship between each property and the coordinate value is found [13]. The DT forms the classification model of the tree structure. It divides the data set into subgroups and creates decision trees at the same time. As a result, a tree is created from nodes and leaves. Leaf node assigns a class and decision node tests an attribute [14].

III. APPLICATION AND ACHIVEMENTS

Five of the most common diseases among the leading causes of absenteeism were given in figure 2. Of the 740 specimens, 149 are medical consultation, 112 are dental consultation, 69 are physiotherapy, 55 are diseases, including musculoskeletal system and connective tissue, and 40 are cases of injury, poisoning and certain other causes.

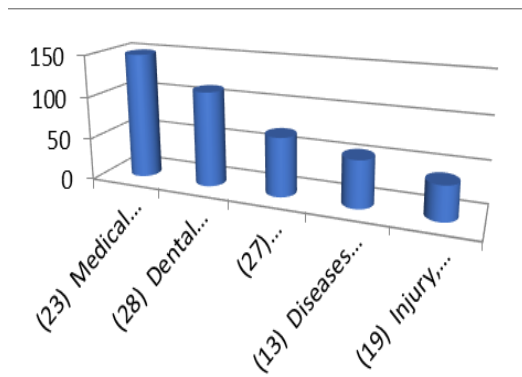


Figure 2: Five of the most common diseases among the leading causes of absenteeism.

Since there are many types of each classification method, the method with the highest results is preferred. In this study, Cosine (CKNN) from KNN varieties, Medium Gaussian (MGSVM) from SVM varieties and Medium (MDT) from DT varieties was used. The classification methods were compared among themselves and the three methods in which the highest success result is obtained are get used. The average values of the three highest classification methods are given in the table 2-3. These classification methods are CKNN, MGSVM and MDT. At first, 75 % of the data were used in education, 25 % were used for testing and accuracy rates were obtained. The results are shown in the table 2.

Table 2: Accuracy of absenteeism for 75 % education and 25 % testing.

True	160	1	
Class	17	7	%
MGSVM	Predict Class		
			90.2
True	153	8	
Class	14	10	%
CKNN	Predict Class		
			88.1
True	153	8	
Class	15	9	%
MDT	Predict Class		
			87.6

Then, with fivefold cross validation, all data were found using both training and testing and accuracy rates and the results are shown in the table 3.

Table 3: Accuracy of absenteeism for cross validation.

True	635	9	
Class	58	38	%
MGSVM	Predict Class		
			90.9
True	636	8	
Class	58	38	%
CKNN	Predict Class		
			91.1
True	623	21	
Class	48	48	%
MDT	Predict Class		
			90.7

When the average success rates of the classification methods were examined, the highest achievements for cross-validation were obtained using CKNN with 91.1%, MDT with 90.9% and MGSVM with 90.7%. CKNN gave higher results than other classification methods. The highest success rates for training and testing are; MGSVM with 90.2% was obtained using CKNN with 88.1% and MDT with 87.6%. MGSVM gave higher results than other classification methods. When the use of training and test data were compared, the success rates of cross validation were higher.

IV. CONCLUSION

In this study, it is thought that the absenteeism which has a direct contribution to work efficiency will be made and it will be thought that the diseases that will cause absenteeism will be known and contribute to the efforts to take measures in advance. The methods used and the methodology applied were found to be suitable for this study.

REFERENCES

- [1] Faruk Şahin. (2011) “ İşe Devamsızlığın Nedenleri, Sonuçları ve Örgütler için Önemi.” Niğde Üniversitesi İİBF Dergisi, 4 (1) : 24-39.
- [2] Banks, Jessie, Patel, Cynthia J., Moola, Mohammed A., (2012) “ Perceptions of Inequity in the Workplace: Exploring the Link with Unauthorised Absenteeism.” SA Journal of Human Resource Management, 10 (1): 1-8.
- [3] Bünyamin BACAK ve Yusuf YİĞİT. (2010) “ Devamsızlığın Nedenleri, Ekonomik Sonuçları ve Azaltılması İçin Alınması Gereken Önlemler.” Girişimcilik ve Kalkınma Dergisi 5(1): 29-44.
- [4] Mashonganyika, Oswald. (2004) “ The Relationship Between Job Satisfaction and Absenteeism: A Study of The Shop Floor Workers in a Motor Manufacturing Plant.” Master Thesis, Rhodes University, South Africa.
- [5] Australasian Faculty of Occupational Medicine, Workplace Attendance and Absenteeism Report, 1999, Australia.
- [6] Stewart, Nicole, “Missing in Action: Absenteeism Trends in Canadian Organizations”, www.conferenceboard.ca, September 23, 2013.
- [7] Keller, Joseph A., (2008) “ Examination of Gender, Pay, Age, Tenure, and Flexible Hours on Absenteeism.” Ph.D Thesis, Northcentral University, Arizona.

- [8] Markussen, Simen, Roed, Knut, Røgeberg, Ole J. and Gaure, Simen, (2011) "The Anatomy of Absenteeism", *Journal of Health Economics*, 30 (2): 277–292.
- [9] Drakopoulos, Stavros, A. and Aikaterini, Grimani, (2013) " The Relationship between Absence from Work and Job Satisfaction: Greece and UK Comparisons." *International Journal of Human Resource Management*, 24 (18): 3496-3511.
- [10] Martiniano, A., Ferreira, R. P., Sassi, R. J., & Affonso, C. (2012) " Application of a neuro fuzzy network in prediction of absenteeism at work." In *Information Systems and Technologies (CISTI)*, 7th Iberian Conference on (pp. 1-4). IEEE.
- [11] Akgeyik, T., (2017) " Çalışan Devamsızlığını etkileyen Faktörler." *Sosyal Siyaset konferansları*, 72, 35-54.
- [12] Özerdem, Mehmet Sıraç and Karadağ, Kerim. (2014). " Parmak hareketlerine ilişkin ECoG örüntülerin AR tabanlı öznelilikler ile sınıflandırılması." *Dicle Üniversitesi , Mühendislik dergisi*, 5(2): 89-97.
- [13] <https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/> Support Vector Machines
- [14] Naoufal Er-raji, Faouzia Benabbou, Mirela Danubianu, Amal Zaouch. (2018) "Supervised Machine Learning Algorithms for Priority Task Classification in the Cloud Computing Environment." *IJCSNS International Journal of Computer Science and Network Security*, 18(11):176-181.

Automated Diagnosis of Tuberculosis using Deep Learning Techniques

Pike Msonda ¹ and Sait Ali Uymaz ²

¹ Konya Technical University, Konya/Turkey, pike.msonda@gmail.com

² Konya Technical University, Konya/Turkey, sauymaz@ktun.edu.tr

Abstract – Tuberculosis (TB), one of the world’s most infectious disease is the cause of millions of deaths every year. Developing countries have the highest number of TB victims due to poor public health systems. Chest Radiography is one of the most used techniques in diagnosing TB. This scenario sets up an opportunity to use Chest X-Ray (CXR) output to provide features for classifying the patients having TB. Using common deep learning architectures, we present the potential of accurately classifying TB from a given CXR image. AlexNet, GoogLeNet and ResNet50 are the (Deep Convolutional Neural Network) DCNN architectures used to demonstrate this potential. The results presented in this report use a public dataset obtained from Montgomery County, Maryland, USA and Shenzhen, China.

Keywords – Tuberculosis, Chest-X-Ray, Deep Learning, CNN, Chest Radiography.

I. INTRODUCTION

Tuberculosis scientifically known as *Mycobacterium tuberculosis* (TB) is an infectious disease that is responsible for up to 1.3 million deaths each year [1]. Developing countries are the biggest victims of this deadly disease due to lack of proper medical facilities to accurately diagnose and treat TB [2, 3]. As a consequence of high HIV (Human Immunodeficiency Virus) rates, people in developing countries have a higher probability of contracting TB [2]. Among the several types of TB, pulmonary TB is the one we are most concerned [4]. This type of TB is unique as it mainly affects the pulmonary region of the body. One of the most common and effective ways to diagnose patients with pulmonary TB is from chest X-Ray (CXR) or chest radiographs (CR) [5, 6]. The output of this diagnosis methodology outputs a radiographic snapshot of the mapping of organs in the chest region. The use of CRs presents a perfect opportunity to use actualise the potential of deep learning models for image classification. Since the problem presented involves images, deep convolutional neural network (DCNN) models are the most effective [6, 7]. DCNN models are simply an up step from the standard convolutional neural network (CNN) models. The aim is to achieve near the same or more accuracy in diagnosis as a Radiologist would achieve. If integrated into a wider Clinical Decision Support System (CDSS), these models can be able to reduce the number of errors, thereby increasing the accuracy of results obtained from patient diagnostics [6, 8, 9]. One of the earliest proposed solutions was presented a solution that combines lung segmentation using graph cut algorithm and classification is performed by Support Vector Machines (SVM) [5]. Although

the focus of this study is specifically deep learning methods, that research also highlights the potential of machine learning algorithms to solve the same problem.

In a recent study, a 121-Layer CNN model was built and used to classify pneumonia CRs [10]. Although pneumonia and TB are fundamentally different diseases, the approach to CR classification is the same. A powerful feature of deep learning models is the ability to pre-train models which allows their various weights and probabilities to be transferred to another architecture [11]. A research using a pre-trained CNN model to extract features for classification demonstrates a particular solution where the resulting feature vector was then finally fed into an SVM classifier [12]. The ability of deep learning to assist in solving the problem of TB in a real-world setting is made evident with a solution to improve TB diagnosis as well as the mobile health technologies of poor communities in Peru in a recently published research [13]. To achieve this, they built a CR database and used DCNN architectures to train models for classification [13]. Further, in 2017 a report published by Radiology.org, written by [14], provides the biggest recommendation of using automation to diagnose Tuberculosis. They present a set of results obtained from 2 models; AlexNet and GoogLeNet. Each consists of pre-trained model results as well as untrained model results. On top of this, they propose an ensemble model which utilises half of the weighted averages of probability scores from AlexNet and GoogLeNet [14].

The subsequent sections of this paper will discuss in brief the deep learning architectures of AlexNet, GoogLeNet, and ResNet50, as well as detail experimentations and results, obtain from training the mentioned models on a public data set provided by Montgomery County, Maryland, USA and Shenzhen, China. Finally, we will present different algorithms or methods to improve the presented results.

II. DATASET

Lack of publicly accessible data is one of the main reasons why few papers have been published in regards to this research. However, this paper presents results obtained from utilizing a public dataset obtained from Montgomery County in a joint collaboration with the Department of Health and Human Services, Montgomery County, Maryland, USA. This set contains precisely 138 CRs where 58 are images for patients diagnosed with TB and the remaining 80 are those patients that having a normal condition. Each CR is exported into PNG (Portable Network Graphics) s 12-bit grayscale images as given in Figure 1. The resolution of each image borders between 4,020×4,892 or 4,892×4,020 pixels.

The second dataset has similar properties to the first however it contains more images which were obtained from People's Hospital, Guangdong Medical College, Shenzhen, China. It contains a total of 662 images divided as 326 positive TB cases and 336 normal cases. This dataset also in the same PNG format, 12-bit grayscale as well as having a 3000 x 3000 pixels resolution. A combination of the two datasets brings the total number of images to be used to 800. A summary relating to the datasets used is presented in Table 1.

Table 1: Summary of public datasets to be used.

Name	Dataset 1	Dataset 2
Source	Montgomery County, USA	Shenzhen, China
Extension	.PNG	.PNG
Normal	80	336
Abnormal	58	326
Total	138	662
Grand total	800	

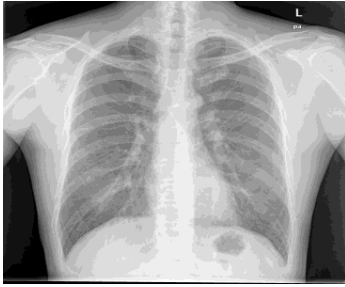


Figure 1: Chest Radiograph image sample

III. METHODOLOGY

The advent of Deep Networks in 2006 sparked a massive interest in deep learning [15, 16]. What deep learning tries to achieve is to arrange (mostly hierarchical) layers of information units and by transferring the features of data found in the first layer to subsequent layers where the data fine-tuned to make good predictions [15]. A contributing factor to the rise of deep learning lies in its ability to solve the different problem of various research domains with remarkable results [17]. In particular, we have used CNN's because they are objectively better at handling grid-like data like images and videos. It is therefore important to note the term DCNN is a result of adding more layers to a standard CNN. Convolutional, subsampling and fully-connected layer are the three main layers that make up a CNN model in that order as presented in figure (2) [9, 17]. The first layer gets the input and through convolution, the returns a set of weights and features that subsequently shared throughout the network (1). Given $x = \{x_1, x_2 \dots x_n\}$ the resulting convolutional matrix is denoted by:

$$y_j = f(\sum_i K_{ij} \otimes x_i + b) \quad (1)$$

Where: y_j is the output of the convolutional kernel denoted by K . This kernel is now mapped unto the x_i with bias b . The function f represents an activation non-linear function.

This is sometimes referred to as weight-sharing which is a very important component of a CNN model. Secondly, the features dimensions are reduced by applying pooling, depending on the problem; average pooling or max pooling can

be used. Finally, the fully-connected layer applies a SoftMax function (4) for classification [18].

$$G(a) = \frac{e^{a_i}}{\sum_i e^{a_i}} \quad (4)$$

Where a_n is the probability distribution of a given set of features.

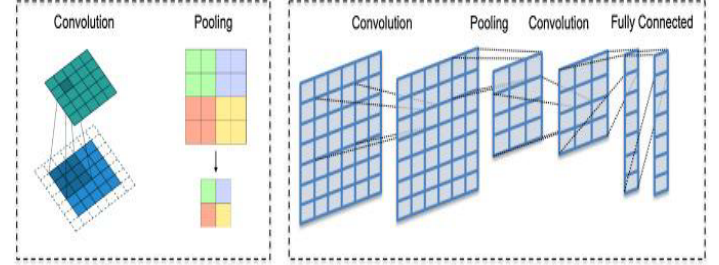


Figure 2: Various CNN layers.

The success of CNN has led to successful implementation of handwriting recognition, face detection, behavior recognition, speech recognition, image classification and many more [15, 17]. The following are the DCNN architectures we have used for this paper:

A. AlexNet

AlexNet is a one of the winners of 2012 ImageNet challenge, developed and presented by [19]. AlexNet prefers to use *rectified linear unit* (ReLU) activation functions in its convolutional layers as opposed to Tanh [20]. This decision increases its speed by fold as well as adding non-linearity to the data which is essential to training deep learning models [18, 19]. The dropout rate is kept at 0.5 with pooling size of 3. This architecture consists of 5 convolutional layers with ReLU activation as aforementioned; the input to the network is set to a size of 224 x 224 x 3 as described in [19] paper. Following the convolutional layers, there are 2 fully-connected layers that finally lead to the output layer [19]. Figure 3 shows a parallel AlexNet design showing various layers from the input to the output layer [19]. The fully connected layer is a set of neurons or hidden units of 4096 each. The output layer is 1000 hidden units that finally leads to a softmax function that performs the classification [19].

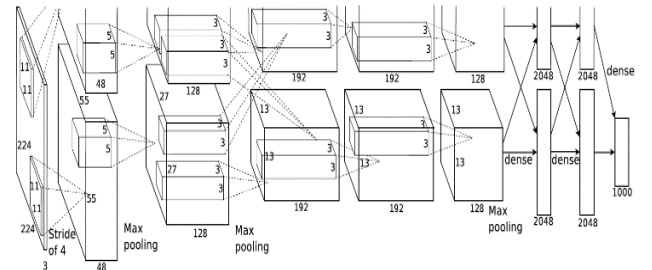


Figure 3: AlexNet Parallel Design.

B. GoogLeNet

GoogLeNet also known as the InceptionV1, is also a winner of the ImageNet Large Scale Visual Recognition Competition (ILSVRC) in 2014 [21]. The InceptionV1 architecture is mainly comprised of 22 layers which try to replicate human

neurological visual image processing. Similar to AlexNet, it uses ReLU activation function (5) in all of its inception layers [21]. One of the most defining features of this architecture is that it does not follow standard CNN design where convolutional layers followed by normalization and max-pooling rather a string of inception modules contribute to filter learning and dimensionality reduction. Regardless of max pooling contribution to drop in accuracy, in this architecture, the repeated use of inception layers compensates for that by reducing the dimension wherever the computational requirements are most needed [22].

This architecture is very deep with accounting 22 layers without pooling and 27 layers all together when pooling layers are also included. A total of 9 inception layers and 2 convolutional layers. The input size of 224 x 224 x 3 with dropout set to 0.4. Further, with this architecture, the number of trainable parameters is reduced from 62.3 Million to only 4 Million [21].

C. ResNet50

Residual Network (ResNet) which usually comprises of more than 152 layers however, other variants containing 34, 50 and 101 layers have also been presented [23]. Again, this architecture came out on top in the 2015 ILSVRC competition, achieving near human accuracy [24]. The concept of Residual Networks tries to solve the problem that arises when you go deeper with convolutional networks; some information is lost. A set of residual functions are used to move the input information on each layer onto the next layer [23]. As already evident the deeper you go the more accurate models become, however that scenario is not always the case as the deeper networks have saturated accuracy which over the course of several layers starts to degrade continuously resulting in a higher training error. This behavior also known as the degradation problem is solved by implementing residual learning [23].

Residual Learning aims at re-routing a mapping produced by normal stacked convolutional layers by reformulating the mapping on the layers by explicitly approximating the residual function. This is under the assumption that multiple nonlinear layers can asymptotically approximate complicated functions, then it is equivalent to hypothesize that they can asymptotically approximate the residual functions [23]. In essence, residual learning can be expressed as follows:

$$F(x) = H(x) - x \quad (2)$$

$$H(x) = F(x) + x \quad (3)$$

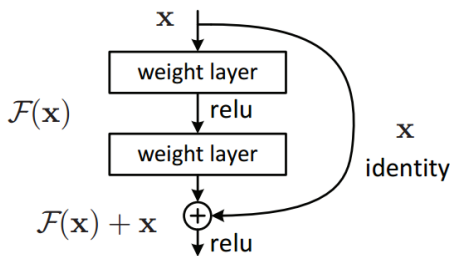


Figure 4: Residual Learning Module.

$H(x)$ here represents the mapping of stacked previous layers of the architecture. Whilst $F(x)$ is the activation function (usually ReLU(5)).

IV. EXPERIMENTS

Naturally, given such lean datasets, data augmentation techniques can be used to reduce overfitting and make the model more robust against subtle changes in test data. However, the results presented below were obtained without any data augmentation or data pre-processing save for resizing the images to fit the standard input of each architecture. Consequently, automatic feature extraction of DCNN leaves very little data pre-processing to be done. When it comes to the models, we used ReLU activation functions (5) in all the layers except the output layer in which we used a SoftMax function (4) [18, 25].

$$R(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{else,} \end{cases} \quad (5)$$

Here $R(x)$ is the ReLU function where x is the probability distribution of given features, similar to (4).

Model loss was calculated using *categorical cross-entropy* (8) which is a product of a Softmax function and a standard cross-entropy (7) [18, 26]. Training was done by resizing the images to 224x224 or 227x227 depending on the model default sizes. From dataset 1 CR images, 20% was separated as a test set. The remaining 110 images we used as the training set with 20% being used for validation of the model. Training on dataset 2 also follows the same parameters as applied on dataset 1. Further, we use stochastic gradient descent (SGD) for model optimization. The optimizer is initialized with a learning rate of 0.001, a momentum of 0.9 and a decay rate of 0.0005 for all the models.

Cross-entropy is calculated as log-likelihood of distribution of y given input x as shown in (6). From equation (4) and (6) we can derive categorical cross-entropy as follows:

$$L\left(f\left(\frac{x}{\theta}\right), y\right) = -\text{Log}(P(f = y|x, \theta)) \quad (6)$$

$$CE = G(x) \cdot L\left(f\left(\frac{x}{\theta}\right), y\right) \quad (7)$$

$$CE = -\text{Log}\left(\frac{e^{a_i}}{\sum_i e^{a_i}}\right) \quad (8)$$

Where CE is the categorical cross-entropy of a given distribution.

We presented results obtained from the training of the models was conducted over two datasets; dataset 1 and dataset 2, see Table 1 for more details. Training of the models yielded results presented in Table 2, where the training accuracy, loss, validation loss and validation accuracy is shown for each model. Model prediction results are presented starting from Table 3 to Table 8. We use contingency tables to calculate for sensitivity, specificity and accuracy of the prediction as shown in Table 9.

Table 2: Results from training models.

	Model	AlexNet	GoogleNet	ResNet50
Dataset 1	Accuracy	0.96439	0.96903	0.91988
	Loss	0.09807	0.07857	0.30472
	Validation accuracy	0.71325	0.64128	0.65416
	Validation loss	1.88020	3.33816	3.26669
	Epochs	120		
Dataset 2	Accuracy	0.95159	0.96686	0.94850
	Loss	0.12271	0.08285	0.15443
	Validation accuracy	0.72059	0.811478	0.80361
	Validation loss	1.54202	1.01642	1.00404
	Epochs	120		

Table 3: Contingency table showing results for AlexNet on dataset 1

AlexNet Dataset 1		Predicated		Total
		Positive	Negative	
Actual	Positive	12	2	14
	Negative	8	6	14
Total		20	8	28

Table 4: Contingency table showing results for AlexNet on dataset 2

AlexNet Dataset 2		Predicated		Total
		Positive	Negative	
Actual	Positive	56	31	87
	Negative	5	41	46
Total		61	72	133

Table 5: Contingency table showing results for GoogLeNet on dataset 1

GoogLeNet Dataset 1		Predicated		Total
		Positive	Negative	
Actual	Positive	6	8	14
	Negative	1	13	14
Total		7	21	28

Table 6: Contingency table showing results for GoogLeNet on dataset 2

GoogLeNet Dataset 2		Predicated		Total
		Positive	Negative	
Actual	Positive	45	7	52
	Negative	16	65	71
Total		51	72	133

Table 7: Contingency table showing results for ResNet50 on dataset 1

ResNet50 Dataset 1		Predicated		Total
		Positive	Negative	
Actual	Positive	10	4	14
	Negative	1	13	14
Total		11	17	28

Table 8: Contingency table showing results for ResNet50 on dataset 2

ResNet50 Dataset 2		Predicated		Total
		Positive	Negative	
Actual	Positive	57	15	72
	Negative	15	46	61
Total		72	61	133

Table 9: Area under the ROC curve (AUC) results for each model

AUC	Model	AlexNet	GoogLeNet	ResNet50
Dataset 1	Sensitivity	85.7	42.8	71.4
	Specificity	42.8	92.8	92.8
	Accuracy	64.2	67.8	82.1
Dataset 2	Sensitivity	64.3	86.5	79.1
	Specificity	89.1	91.5	75.4
	Accuracy	72.9	82.7	77.4

We used AUC (area under the ROC curve) scores to measure the performance of our models [27]. On dataset 1, AlexNet achieved the highest sensitivity of 85.7 conversely it slumped to the lowest in terms of specificity and accuracy. This is to say that they ability of AlexNet to classify positive cases of TB is higher than the other models on the first dataset. GoogLeNet in contrast to AlexNet achieved the highest specificity of 92.8 however it performs poorly when it comes to sensitivity. Overall GoogLeNet got a higher accuracy rate than AlexNet but failed to achieve the consistency of ResNet50 which got scores of (71.4,92.8, 82.1). Further, results obtained from dataset 2 are more consistent, this can be because dataset 2 has more images than dataset 1. Nonetheless, AlexNet has the lowest sensitivity and accuracy as compared to GoogLeNet and ResNet50. On this dataset, GoogLeNet achieves the highest a sensitivity, specificity and accuracy of all the models.

V. CONCLUSION

In this paper, we present a solution to automatically diagnose TB patients. We use already designed deep learning architecture which are trained on public datasets from Montgomery, USA and Shenzhen, China. The results are very encouraging depending on the small dataset. Nonetheless, this present a huge opportunity to integrate the trained models to a CDSS that could greatly assist doctors in diagnosis of patients with TB, more especially in developing countries where the availability of trained professionals is not readily available. Specifically, AlexNet's overall performance was the worst as compared to the other models, this could be because the model's architecture is not as deep

as its counterparts. In contrast, ResNet50 performed considerably consistent on both datasets where as GoogLeNet was the best performer on dataset 2.

VI. FUTURE WORK

In the future we will try improve the used models' overall accuracy by implementing basic data augmentation techniques. On top of that, as observed the performance of the model on two different datasets was quite different because of that, we wish to create our own dataset of CRs images.

VII. REFERENCES

- [1] J. Furin, H. Cox, and M. Pai, "Tuberculosis," *The Lancet*, 2019/03/20/ 2019.
- [2] W. H. Organization, *Global tuberculosis report 2013*. World Health Organization, 2013.
- [3] W. H. Organization, *Global tuberculosis report 2018* World Health Organization 2018.
- [4] G. Abebe, Z. Bonsa, and W. Kebede, "Treatment outcomes and associated factors in tuberculosis patients at Jimma University Medical Center: A 5-year retrospective study," *International journal of mycobacteriology*, vol. 8, no. 1, p. 35, 2019.
- [5] S. Jaeger *et al.*, "Automatic tuberculosis screening using chest radiographs," *IEEE Trans. Med. Imaging*, vol. 33, no. 2, pp. 233-245, 2014.
- [6] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60-88, 2017.
- [7] L. Saba *et al.*, "The present and future of deep learning in radiology," *European Journal of Radiology*, vol. 114, pp. 14-24, 2019/05/01/ 2019.
- [8] M. P. McBee *et al.*, "Deep Learning in Radiology," *Academic Radiology*, vol. 25, no. 11, pp. 1472-1480, 2018/11/01/ 2018.
- [9] A. Maier, C. Syben, T. Lasser, and C. Riess, "A gentle introduction to deep learning in medical image processing," *Zeitschrift für Medizinische Physik*, 2019/01/25/ 2019.
- [10] P. Rajpurkar *et al.*, "CheXnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.
- [11] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345-1359, 2010.
- [12] U. Lopes and J. F. Valiati, "Pre-trained convolutional neural networks as feature extractors for tuberculosis detection," *Computers in biology and medicine*, vol. 89, pp. 135-143, 2017.
- [13] Y. Cao *et al.*, "Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor and marginalized communities," in *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 2016, pp. 274-281: IEEE.
- [14] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574-582, 2017.
- [15] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11-26, 2017/04/19/ 2017.
- [16] R. Wason, "Deep learning: Evolution and expansion," *Cognitive Systems Research*, vol. 52, pp. 701-708, 2018/12/01/ 2018.
- [17] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Information Fusion*, vol. 42, pp. 146-157, 2018/07/01/ 2018.
- [18] C. Cao *et al.*, "Deep Learning and Its Applications in Biomedicine," *Genomics, Proteomics & Bioinformatics*, vol. 16, no. 1, pp. 17-32, 2018/02/01/ 2018.
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.
- [20] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, 2013, vol. 30, no. 1, p. 3.
- [21] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
- [22] R. McCoppin and M. Rizki, "Deep learning for image classification," in *Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR V*, 2014, vol. 9079, p. 90790T: International Society for Optics and Photonics.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [24] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: Generalizing residual architectures," *arXiv preprint arXiv:1603.08029*, 2016.
- [25] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 2011, pp. 315-323.
- [26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980-2988.
- [27] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145-1159, 1997/07/01/ 1997.

A system based on image processing and deep CNN features for classification of defective fruits

Rhudie Grace Zang Edzang¹ and Mehmet Göktürk¹

Corresponding author Address: Department of computer engineering, Gebze Technical University, 41400 Gebze-Kocaeli. E-mail: rgrace@gtu.edu.tr

Corresponding author Address: Department of computer engineering, Gebze Technical University, 41400 Gebze-Kocaeli.

ABSTRACT :

Faced with the growing demand for quality products in markets and industries in agricultural sectors . from consumers, sellers, and producers. The research aims to respond to this, by using effective and efficient technologies, namely image processing and computer vision, to automate the process of inspection and evaluation of the quality of agricultural products. Conducted for many years by human experts. The use of these technologies concerns both the detection of fruit diseases and their classification. To follow this concept, we chose to focus our work on fruits classification. whose purpose is to separate infected fruits from those that are not affected. However, based on the concepts of computer vision, the proposed technique is centered on three steps. The first step concerns preprocessing and segmentation, in which, we resize and improve the quality of the images. The second step involves, deep pre-trained models (Alex-Net), these are used for the extraction of the features in the different fruits (banana, apples, oranges). And finally, we performed the classification, by using multi-class SVM. The studies were conducted on the public dataset, the Kaggle dataset, to obtain a 99.50% classification accuracy and on a new dataset for a accuracy of 85.7 %. The results clearly show that the proposed method works well in terms of improvement in classification accuracy and precision.

Keywords: Classification, Deep features Extraction, Fruits diseases, Image processing, Segmentation.

I. Introduction

The visual quality control of the products is currently unavoidable in many sectors. As the only means of recognition, and classification, the image serves as a tool that enables technology to replace and even surpass the traditional means (human eyes) of product control, such as fruits in the agricultural industry. Computer vision is the technology that allows us to replace human vision in this process of fruit inspection and grading in today's modern world, it also allows us to do its automation. This problem has generated a lot of interest in recent years. An automatic inspection system can be described by the combination of sophisticated image processing and advanced computer vision techniques. This coupling has a great reputation of being the means per excellence to offer good result and rapid identification, thus reducing man's efforts and the cost of labor. The main factors that can affect the quality of fruits are biotic and abiotic. To find an efficient and effective solution to this problem, many approaches have been proposed for identifying and classifying fruits by their diseases, and by separating fruit infects from those that are not. (e.g. apple, dates, oranges, and other citrus fruits). Then, many surveys and comparative studies on the classification and identification of fruit diseases based on image analysis techniques can be found in the literature[1], [2]. in image processing, the operation of dividing an image and gathering the pixels together according to pre-defined criteria is called segmentation.

Effective segmentation techniques have been introduced in several studies such as thresholding [3],[4]; K-Means clustering [5],[6],[7], and edge detection [8]. The next step plays an important role in the representation of an image, it is called the feature extraction. which is another way of understanding and seeing an image. There are several methods for extracting and describing features, However many of them have been proposed for citrus fruits like apples and oranges. Which include Local binary pattern and complete local binary pattern [7], color histogram, deep features [9]. Other technique can be found in the literature. These characteristics are then classified by several classification methods such as support vector machine (SVM)[1], [10], multi-class SVM (M-SVM) [11],[9], artificial neural network (ANN)[12],and Bayesian network of classification[13].

Statement of the problem

All these classifiers are excellent and offer great accuracy, they can provide very good results, and their performance strongly depends on the quality of features previously found. but when classes are more relevant to each other, the accuracy of the whole system decreases [13]. The results can then be improved by creating new feature extraction methods that are more adapted to the studied images, or by using a "better" classifier. So, at the annual ILSVRC Computer Vision Competition, a new Deep Learning algorithm explodes the records! This is a convolutional neural network called AlexNet. Its methodology is similar to those traditional methods of supervised learning: it receives input images, detects the features of each of them, then trains a classifier on it. However, the features are learned automatically. It is designed to do all the hard work of extraction and description of features: during the training phase, the classification error is minimized to optimize the parameters of the classifier and the features! In addition, the specific architecture of the network can extract features of different complexities, from the simplest to the most sophisticated.

Contribution

Then, in this work we propose an automatic system for the classification of apples, oranges, and bananas; with and without diseases by applying the traditional method of image processing combined with Deep Feature extraction. Composed of three major steps, including pre-processing and segmentation, feature extraction and finally classification.

- for the pre-processing stage, the images were first resized for a size of 227 * 227, then to improve and erase the salt noise on the images the median filter was applied.
- Segmentation was applied by the thresholding method to separate the background and the foreground.
- the classification of the different fruits was carried out by the implementation of a deep architecture in which a pre-trained model was used. and the choice is Caffe-Alex Net to extract the deep features. And finally, we use a multi-SVM for classification.

the structure of the rest of the article is as follows. Sections I and II give a brief description of the different methods and results proposed in the literature. Experimental datasets, methods, and results are described in Section IV. and finally, in the last section, a conclusion is drawn.

II. Literature Review

In this section, recent research in the use of digital image processing and convolution neural network architectures in agricultural applications is discussed. Prior to image processing and machine learning techniques have been used to classify different plants diseases.[1];[14];[15];[11].Generally, these systems implement the following steps: First, images are acquired using a digital camera, followed by the image processing's techniques, such as image enhancement, color space conversion, image filtering to improve the quality of the images. Secondly, the Segmentation was required to detect the region of interest in the image. Then, From the segmented images, the important features are extracted and used as an input for the machine learning classifier [6].

With this technique, the overall accuracy is dependent on the type of image processing and feature extraction techniques used [16], but also on the classifier. during these last few years, studies have achieved that state of the art with the use of convolution neural networks, a special deep learning network specializes in image processing and pattern recognition being multilayer supervised networks, can automatically learn entities from datasets. CNNs have achieved peak performance in almost all-important image classification tasks in their last years. They have the noble distinction of performing both extraction and classifications functions in the same architecture. One of the biggest advantages of CNNs is that they can automatically learn the features of images without a need of

segmentation method [16]. It has been proposed to use various CNN architectures for object recognition, for example. LeNet, AlexNet, GoogLeNet etc.

This deep learning architecture has been incredibly successful in plant disease detection studies. In [17] propose for the classification of banana leaves including healthy (1643 images) and unhealthy (2057 images) fruits, the LeNet architectures as a convolutional neural network. They applied a technique called transfer learning. In [18] A dataset of 54306 images was used. They applied AlexNet and GoogleNet models based on CNNs to detect and identify 26 diseases, the final accuracy was 99.35%. For the identification of soybean plant Disease, [16] use a CNN architecture consist of convolutional layers, each followed by a max-pooling layer. And the fully connected MIP was used at the final layer. They achieved 99.32% of accuracy.

But to use this type of model, a huge amount of image is needed, which is computationally intensive, then [19] proposes a novel classification architecture by using a pre-trained deep learning network combine with traditional classifiers. They utilized AlexNet, GoogleNet architecture, and VGG for feature extraction and then the features are classified by traditional classifiers which are SVM, ELM and KNN. For a dataset of 1965 images consisting of the real plants with pest disease. They obtained the highest-level accuracy of 97.86%, with the ResNet50 model and SVM classifier. To obtain an accuracy of 98.60%, [9] propose an architecture base on deep CNN features where two deep pre-trained models VGG16 and Caffe Alex-Net are used for feature extraction of selected diseases (banana diamond leaf spot and deightoniella leaf and fruit spot). Then for classifying the diseases, they are used multi-class SVM. In conclusion to this literature, the results demonstrated the feasibility of CNN in-depth image processing techniques and image processing for classification of plant diseases.

III. Materials and Methods

To classify apples, oranges, and bananas fruits diseases a large collection of the fruit's images is demanded. The images are downloaded from the Kaggle-dataset. In this section the methodology followed is discussed in detail.

Dataset

A complete and adequate dataset is required for all classification research during the training and testing phase. The dataset for the experiment is downloaded from the Kaggle database containing different fruits images. It contains a collection of images taken in different environments. A dataset containing 13221 fruits images from six classes including healthy and infected fruits is downloaded. The class samples of the dataset are summarized in Table 1.

Label	Type of fruits	Train	Test	Number
1	fresh apples	1632	408	2040
2	fresh banana	1500	375	1875
3	fresh oranges	1392	348	1740
4	rotten apples	2405	601	3006
5	rotten banana	2031	507	2538
6	rotten oranges	1618	404	2022

Table 1. Details of the utilized dataset used for the classification.

The proposed method

All the details of the proposed method are illustrated in figure 1. As seen in fig 1, the proposed method is composed of four layers. To begin with, let introduce the concept of Deep Feature Extraction in the next subsection.

Deep Feature Extraction

Based on the extracting features learned from a pre-trained convolutional neural network. Deep feature extraction can be obtained by using the fully-connected layer of pre-trained networks to extract features from an input The features represent the most important image, whereby it will be necessary for the classification, whereby it will be necessary for the classification. To conclude, we used one of the most popular pre-trained CNN models called Alex-Net to extract effective deep features. This deep CNN known to the public during a computer vision competition called ImageNet in 2012 is composed of 25 layers where 8 of them are composed of learnable weight. whereby five of the eight are called convolutional layers and the other three are called fully connected. In this architecture, the convolutional layers use a variable kernel size and each of them is followed by max-pooling layers.

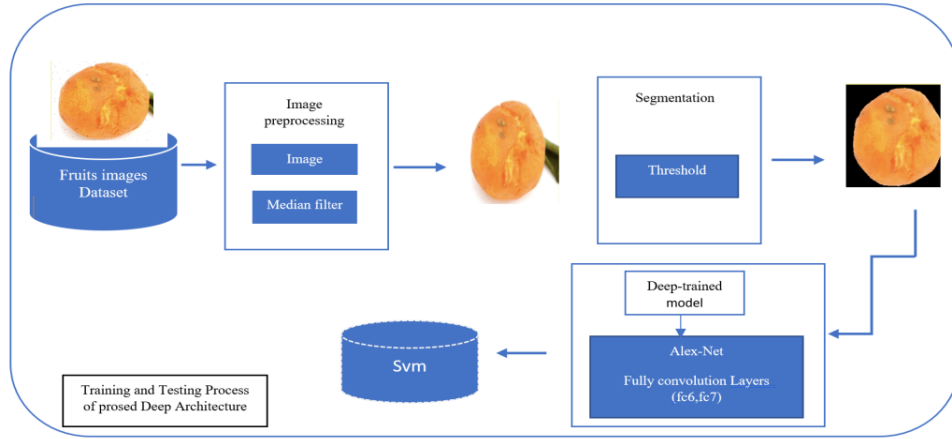


Fig.1.The architecture of proposed disease identification and classification.

IV. Experimental Setup and Results

For the experimental studies, MATLAB software was used for coding on a computer having an Intel Core i7-7500U CPU with 8 GB memory. The data were numerically split for training and testing, for each category of fruits species the details are given in table 1.

Now we detail step by step the proposed architecture:

Step 1: Pre-processing the input image, by resizing the image according to Alex-Net into 227 x 227 using bilinear interpolation, then we applied the media filter to remove the salt noise on input images.

Step 2: From the enhanced image we use a threshold method to separate the background from the foreground by using otsu's method which computes a global image threshold.

Let $I_e(x,y)$ be the resized and enhanced image. The image, which is in the RGB color model is converted to YCbCr color space. Each enhanced image are consists of three channels in which y is for the luminance, and Cb and Cr are the Chrominance channels. Following the equation from [20], the Y, Cb, and Cr are extracted.

$$\begin{pmatrix} I_y \\ I_{Cb} \\ I_{Cr} \end{pmatrix} = \begin{pmatrix} 16 \\ 128 \\ 128 \end{pmatrix} + \begin{pmatrix} 0.299.IeR & +0.587.IeG & +0.114.IeB \\ -0.169.IeR & -0.331.IeG & +0.500.IeB \\ +0.500.IeR & -0.419.IeG & -0.081.IeB \end{pmatrix}$$

once extracted from the RGB image, the background will be separated from the foreground by the following equation.

$$E(w) = \begin{cases} 0 & \text{if } I_e(x,y) > T \\ 1 & \text{if } I_e(x,y) < T \end{cases}$$

Where T represents the threshold value taken using the histogram of the image. The threshold was performed on the I_eY channel. Then, After obtaining the segmented image we perform some morphological operation to remove noise on the segmented image. Opening(2) and closing(1) were used:

$$Ie \bullet S = (Ie \oplus S) \ominus S \quad 1$$

$$Ie \circ S = (Ie \ominus S) \oplus S \quad 2$$

Where S represents structure element which is a matrix 3*3. The result of the segmentation is given in fig 2.

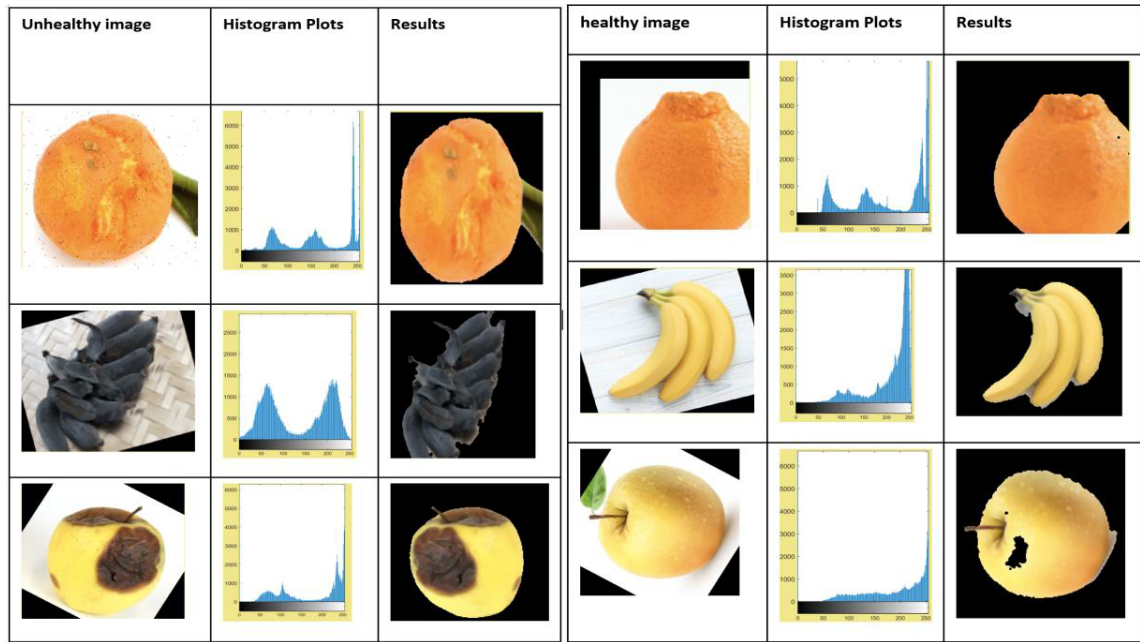


Figure 2: Example of the pre-processing process from the left to the right, original images and resize and improving images and the segmented images.

From the result obtained in step 2, we extracted the deep features with the fully-connected layer based on Alex-net pre-trained model, but also from two other pre-trained model vgg16 and vg19. The deep features are extracted from the fc6, fc7 layers, and by the combination of the two layers. The results of the deep-features are given in Table 2.

	Unhealthy Apple	Unhealthy Banana	Unhealthy Orange	Healthy Apple	Healthy Orange	Healthy Banana
Unhealthy Apple	1208	0	1	9	0	0
Unhealthy Banana	0	1217	0	0	1	0
Unhealthy Orange	1	0	1211	1	0	5
Healthy Apple	8	0	0	1200	0	10
Healthy Orange	0	0	0	0	1218	0
Healthy Banana	2	0	3	3	0	1210

Figure 3: confusion Matrix of fc7 (Legend:1. fresh apples; 2. fresh banana; 3. fresh oranges; 4. rotten apples; 5. rotten banana; 6. rotten oranges; 7.)

As it can be observed in table 2, the deep features extracted from the fc6 layer of the Alex-net model has the best deep features. And has also, the highest accuracy level among the three models. The use of segmentation slightly reduces the accuracy of the

system. The use of image processing techniques not increase nor improve the accuracy of the model. The result of the deep feature from Alex-net can be seen from the confusion matrix above in fig 3. Where the mean of the matrix was calculated to obtain the accuracy.

For more, we had also made some test on a new dataset. The dataset consists of 126 apple images of unhealthy fruits and 500 images of healthy apple. This new dataset is composed only of apple images. the results and the comparison with the Kaggle dataset are given below. We extract the deep features from the Alex-net model for the new dataset.

The result of the preprocessing and segmentation is given in fig 4.

	Kaggle-Dataset			New-Dataset		
	FC6	Fc7	Fc6+Fc7	Fc6	Fc7	Fc6+Fc7
With segmentation	99.56	99.26	99.03	85.7	85.2	80.26
Without segmentation	99.60	99.40	99.30	84.2	82.4	77.3

Table 4 : Comparasion between Kaggle-Dataset and New Dataset

	Alex-net			Vgg16			Vgg19		
	Fc6	Fc7	Fc6+fc7	Fc6	Fc7	Fc6+Fc7	Fc6	Fc7	Fc6+fc7
With segmentation	99.56	99.26	99.03	99.47	99.28	98.7	99.1	98.8	98.1
Without segmentation	99.60	99.40	99.30	99.30	98.58	98.3	99.35	98.5	99.1

Table2: Examination of the dataset.

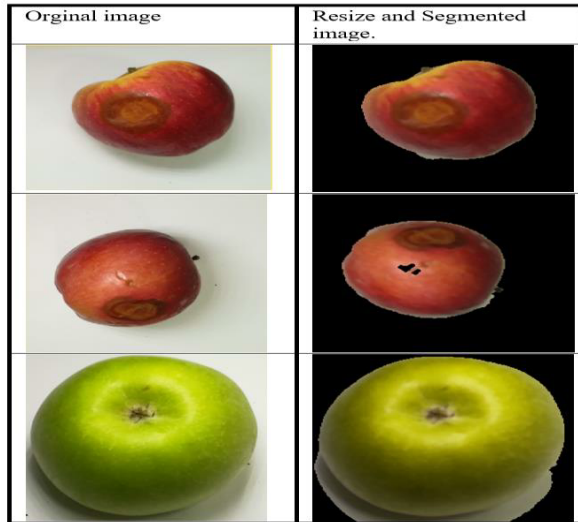


Figure 4 : Result of the pre-processing process, original images and resize and improving images and the segmented images.

model used. The advantage of this approach in the classification of diseased fruits is the automatic of, vectors descriptors. And also the use of preprocessing and segmentation methods may not be required but can improve the classification accuracy. In conclusion, this approach reduces the size of the problem by replacing conventional but sophisticated image processing methods. As we believe that we can get better results while adding a few more diseases. In our future work, we will no longer go deeper on the feature extraction part, in order to authenticate the proposed method.

For the new dataset, the feature was also extracted from the fc6, fc7. From the result, the fc6 layer has the highest accuracy. But for the new dataset, the highest accuracy was obtained by applying the image processing techniques.

V. CONCLUSIONS

This work was an investigation of image processing and machine learning methods to classify apple, banana and orange fruit diseases in the proposed method. Our main focus was on the segmentation and feature extraction step. These Preliminary results show the effectiveness of the proposed approach, even under difficult conditions, such as a complex background and a different image quality. It should also be noted that using a pre-trained model as a feature extractor means that the efficiency of the system depends on the

REFERENCES

- [1] M. Bhange and H. A. Hingoliwala, "Smart Farming: Pomegranate Disease Detection Using Image Processing," *Procedia Comput. Sci.*, vol. 58, pp. 280–288, 2015.
- [2] Z. Bin Husin, A. Y. Bin Md Shakaff, A. H. Bin Abdul Aziz, and R. B. S. Mohamed Farook, "Feasibility study on plant chili disease detection using image processing techniques," *Proc. - 3rd Int. Conf. Intell. Syst. Model. Simulation, ISMS 2012*, pp. 291–296, 2012.
- [3] "measuring the severity of fungi caused disease on leaf using triangular thresholding method by Richard marfo (bed it) a thesis submitted to the department of computer science
- k [4] R. P. Shaikh and S. A. Dhole, "Citrus leaf unhealthy region detection by using image processing technique," *Proc. Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2017*, vol. 2017–Janua, pp. 420–423, 2017.
- [5] M. S. Kim, A. M. Lefcourt, Y. R. Chen, and Y. Tao, "Automated detection of fecal contamination of apples based on multispectral fluorescence image fusion," *J. Food Eng.*, vol. 71, no. 1, pp. 85–91, 2005.
- [6] M. T. Habib, A. Majumder, A. Z. M. Jakaria, M. Akter, M. S. Uddin, and F. Ahmed, "Machine vision based papaya disease recognition," *J. King Saud Univ. - Comput. Inf. Sci.*, pp. 0–9, 2018.
- [7] S. R. Dubey and A. S. Jalal, "Detection and classification of apple fruit diseases using complete local binary patterns," *Proc. 2012 3rd Int. Conf. Comput. Commun. Technol. ICCCT 2012*, pp. 346–351, 2012.
- [8] X. Li and W. Zhu, "Apple grading method based on features fusion of size, shape and color," *Procedia Eng.*, vol. 15, pp. 2885–2891, 2011.
- [9] M. A. Khan *et al.*, "CCDF: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features," *Comput. Electron. Agric.*, vol. 155, no. October, pp. 220–236, 2018.
- [10] D. Abdelhamid, R. Salah, B. Mohamed Chaouki, and T.-A. Abdelmalik, "Classification des images des dattes par SVM: contribution à l'amélioration du processus de tri," *Cosi 2010*, pp. 159–167, 2010.
- [11] M. Sharif, M. A. Khan, Z. Iqbal, M. F. Azam, M. I. U. Lali, and M. Y. Javed, "Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection," *Comput. Electron. Agric.*, vol. 150, pp. 220–234, Jul. 2018.
- [12] G. Capizzi, G. Lo Sciuto, C. Napoli, E. Tramontana, and M. Woźniak, "Automatic Classification of Fruit Defects based on Co-Occurrence Matrix and Neural Networks," vol. 5, pp. 861–867, 2015.
- [13] D. Unay, B. Gosselin, O. Kleynen, V. Leemans, M. F. Destain, and O. Debeir, "Automatic grading of Bi-colored apples by multispectral machine vision," *Comput. Electron. Agric.*, vol. 75, no. 1, pp. 204–212, 2011.
- [14] Z. Iqbal, M. A. Khan, M. Sharif, J. H. Shah, M. H. ur Rehman, and K. Javed, "An automated detection and classification of citrus plant diseases using image processing techniques: A review," *Comput. Electron. Agric.*, vol. 153, no. September 2017, pp. 12–32, 2018.
- [15] M. Kumar, "Engineering characterization of kinnow and tomato for grading using image processing," 2017.
- [16] S. Walleign, "Soybean Plant Disease Identification Using Convolutional Neural Network," pp. 146–151, 2017.
- [17] J. Amara, B. Bouaziz, and A. Algergawy, "A Deep Learning-based Approach for Banana Leaf Diseases Classification," *Btw*, pp. 79–88, 2017.
- [18] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Front. Plant Sci.*, vol. 7, 2016.

- [19] J. Garba and S. Sani, "Coordinated Placement and Tuning of STATCOM and SSSC for Oscillation Damping in a Wide Area Networks of Power System .," pp. 11–13, 2018.
- [20] M. K. Tripathi, "A Framework with OTSU ' S Thresholding Method for Fruits and Vegetables Image Segmentation," vol. 179, no. 52, pp. 25–32, 2018.

A Survey of Joint Routing and Energy Optimization Techniques for Underwater Acoustic Sensor Networks

H. Yetgin¹

¹ Bitlis Eren University, Bitlis/Turkey, hyetgin@beu.edu.tr

Abstract - Underwater sensor networks (UWSNs) are composed of a certain amount of sensors and/or vehicles that interactively glean data from underwater environment and cooperatively perform predetermined tasks. Each of these battery-limited sensors dissipates a considerable amount of energy during sensing, communicating and data processing activities, where replenishing the drained batteries is impractical and time consuming, and interrupts the ongoing communication. Therefore, energy conservation and routing strategies are vitally important for accomplishing a certain task for a particular underwater application. In this paper, we review the recent advances in UWSNs, including their applications and joint routing optimization and energy conservation techniques. Finally, several appealing directions for future research on UWSNs are presented.

Keywords - Routing Optimization, Energy Conservation, Energy Dissipation, Underwater Acoustic Sensor Networks.

I. INTRODUCTION

ABOUT 71% of earth's surface is covered by different forms of water, such as oceans, seas and rivers, which indicates a plethora of precious resources lying underwater is essential to be explored [1]-[3]. Recent advances in underwater wireless sensor networks (UWSNs) have led to potential exploration of underwater with the aid of sensors. Today, UWSNs play a significant role in plenty of underwater applications, ranging from instrument tracking, early warning systems of natural disaster and exploration of oil underneath ocean to oceanographic data collection, pollution monitoring and tactical surveillance [4], [5]. UWSNs also support the control of autonomous underwater vehicles (AUVs) [2].

Underwater communication operates with a set of sensors transmitting gleaned information to buoyant gateways that relay the data to the nearest coastal station, so-called remote station. This task can also be achieved via a satellite, which can be depicted from Fig. 1. An ongoing two-dimensional (2-D) underwater communication architecture can also be illustrated as in Fig. 1. Mainly, acoustic transceivers are utilized at the time of communication [1]-[4]. With the help of acoustic waves, data can be relayed over long distances thanks to its long wavelength, but at the cost of small bandwidth due to low frequency waves. Although UWSNs

offer promising solutions to the above-mentioned applications, a large portion of the electromagnetic frequency spectrum suffers from unpredictable characteristics of underwater environment, such as severe attenuation, frequency dispersion, large propagation delays, multipath fading due to the reflections from the sea ground and the surface [6]. All these drawbacks necessitate solid application deployment of UWSNs for the sake of extended operations and uninterrupted communication, which indeed requires the optimization of routing and energy dissipation.

Domingo *et al.* [7] analyzed total energy consumption in UWSNs considering shallow water and deep water, where they focused their attention solely on the energy consumption analysis of the proposed routing protocols. Noting that this paper is a more recent contribution on routing optimization and energy dissipation strategies of UWSNs than Domingo *et al.* [7] from 2008. In this paper, we have also provided a comprehensive classification of routing decisions and energy conservation techniques. Moreover, a limited number of papers on energy-efficient routing protocols are surveyed in [8]. However, to the extent of our knowledge, a comprehensive classification of routing optimization and energy conservation strategies have not been disseminated in the literature. The aim of this treatise is to survey and present recent advances in increasing the performance of UWSNs in terms of routing optimization and energy conservation techniques.

The remainder of this paper is organized as follows. We survey routing optimization and energy conservation techniques of UWSNs in Section II, where we have also classified these techniques in terms of their objectives and design constraints. Then, the current issues as well as the potential research ideas are discussed in Section III. Finally, we conclude with a summary in Section IV.

II. ROUTING OPTIMIZATION AND ENERGY DISSIPATION

Coverage hole is the phenomenon induced by the death of a sensor or occurs due to the energy hole, which can also be caused by an uneven distribution of sensors in the network. Random deployment can also be another reason for the network field to be more populated at some points and to be scarcely populated at the remaining parts. In turn, coverage holes lead to energy holes due to unbalanced and recurring

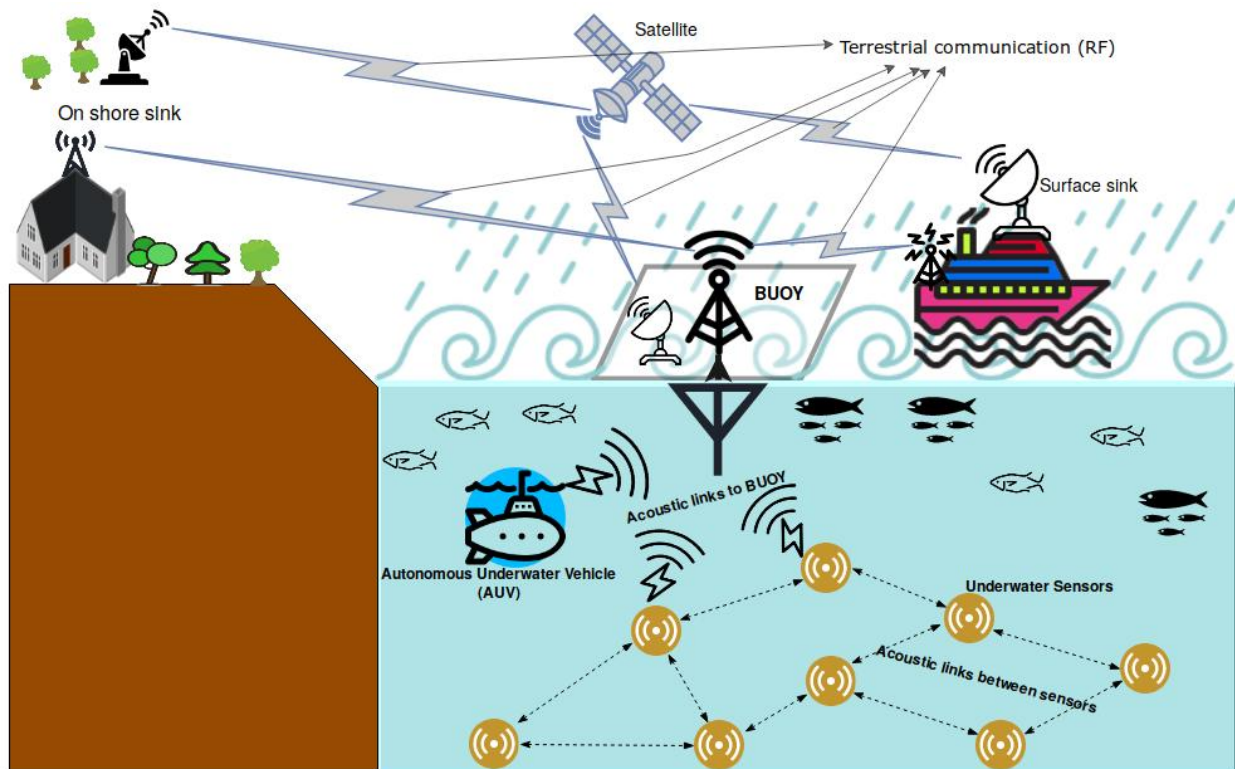


Figure 1: A typical 2-D architecture for UWSNs.

transmission of the sensed data through the same route. To cope with the above-mentioned issues, Latif *et al.* [10] designed a coverage hole repair technique for UWSNs, where depth based routing (DBR) is first exploited for finding the energy-hungry spots. Then, the proposed technique identifies the redundant coverage areas in order to move the sensor from redundant areas to scarce areas to avoid creation of holes. This technique aids in maximizing the network lifetime and throughput at the cost of increased delay. Identical approaches to avoid coverage and energy holes are also followed by [11] and [12]. Similarly, Chen *et al.* [13] investigated a mobile geocast problem in three-dimensional (3D) UWSNs, which tackles with the hole problem in order to minimize energy dissipation of sensors, while maximizing data collection. To solve this specific problem, an apple slice technique is exploited to build multiple layers to cover the hole and to guarantee path continuity, where simulation results demonstrated the performance improvement in successful delivery rate, power dissipation and message overhead. Rou *et al.* [14] employed an energy-balanced unequal layering clustering (EULC) algorithm for enhancing the energy-efficiency of acoustic sensors, which is devised with unequal layering using node depth information. This approach aids in avoiding hole issues by constructing clusters of varying sizes within the same layer. EULC algorithm manages to balance energy dissipation throughout the UWSNs and thus assists in extending the network lifetime. Geethu *et al.* [15] considered an energy-efficient routing protocol that relies on the optimal hop distance, where forwarding nodes are determined based on depth, remaining energy, and distance from the computed optimal hop position

to source node. Simulation results of [15] demonstrated that energy dissipation is considerably reduced, leading to increased lifetime of UWSNs, whereas courier sensors are benefited for handling coverage-hole issue, which results in improved packet delivery ratio and increased throughput.

Internet of underwater things enabled the deployment of smart things in UWSNs in order to facilitate the discovery of vast unexplored seabed. Therefore, inexpensive packet forwarding and energy conservation are inevitable functions that a routing protocol has to contain. To address this challenge, Zhou *et al.* [16] proposed an enhanced channel aware routing protocol (E-CARP) in order to accomplish location-free and greedy hop-by-hop packet forwarding technique, which opportunistically routes data through the most appropriate relay node to initiate data transmission at each time point. Similarly, Javaid *et al.* [17] proposed balanced energy adaptive routing (BEAR) to balance the overall energy dissipation and to prolong the lifetime of UWSNs. BEAR operates with the remaining energy and location information, so that the least energy consuming neighbour sensors can be selected and sensors with relatively higher residual energy can be exploited for relaying data. BEAR outperformed its counterparts in terms of network lifetime in simulation results of [17].

Highly dynamic nature of UWSN links entails for an adaptive, scalable and energy-efficient routing schemes for UWSNs. DBR, for example, suffers from void region and detoured forwarding. To tackle these issues, Guan *et al.* [18] designed a distance-vector based opportunistic routing (DVOR) scheme, which utilizes a mechanism that records the smallest hop counts toward the surface sink. Then, packet is

opportunisticly forwarded based on the distance vectors of sensors constituting the smallest hops. DVOR has been advocated to have low overhead, the ability to avoid void region and long detour, which are then validated by simulations, outperforming the existing protocols in terms of packet delivery ratio, energy-efficiency, and average end-to-end delay. It is widely-known that acoustic channel is energy hungry, which deteriorates the lifetime performance of UWSN and diminishes the performance of monitoring applications. These challenges are separately tackled with duty-cycling and opportunistic routing in the literature. Coutinho *et al.* [19] amalgamated these into a hybrid scheme taking both the symbiotic design of anypath routing and duty-cycling into account for sustainable UWSNs, where strobed preamble low power listening (LPL) and low power probing (LPP) techniques are implemented for the design of asynchronous duty-cycling protocols. Moreover, Hsu *et al.* [20] applied the idea of opportunistic routing for maximizing the goodput while satisfying the end-to-end delay requirements for time-critical UWSN applications, where the packet is dropped if its deadline cannot be met. However, the amount of missed deadlines has to be minimized for the sake of energy efficiency. To efficiently solve this problem, Hsu *et al.* [20] developed a two-step heuristic algorithm, which is constituted by per-node forwarding node selection and packet forwarding prioritization. Similarly, Rahman *et al.* [21] proposed an energy-efficient cooperative opportunistic routing (EECOR) protocol to forward packets towards the surface sink. In order to forward packets, a set of forwarding relay is determined with the help of local information of forwarder and a fuzzy logic-based relay selection scheme is employed to identify the best relay by taking into account the energy consumption rate and packet delivery probability of the forwarder. EECOR protocol is conducted on the Aqua-sim platform and compared with its existing routing protocols with regard to the performances of packet delivery ratio, end-to-end delay, energy dissipation and network lifetime. Another approach on opportunistic routing is proposed by Noh *et al.* [22], who targeted the design of an efficient anycast routing algorithm for reporting underwater sensor events to any surface buoy, where authors proposed hydraulic-pressure-based anycast routing protocol, so-called hydrocast, exploiting the measured pressure levels to relay data to the surface buoy, which is based on an opportunistic routing mechanism to maximize the energy efficiency and to limit cochannel interference. An appealing technique for forwarding packets to buoy is introduced by Jin *et al.* [23], who devised a Dempster-Shafer evidence theory-based opportunistic routing (EBOR) protocol, where source node takes residual energy and packet delivery probability into account for determining the next optimal hop. Simulation results of [23] demonstrated that EBOR assists in prolonging the lifetime of UWSNs by means of uniformly distributing the energy usage, and achieving a higher packet delivery ratio.

Another compelling application of UWSNs can be found in [24], where sensor measurements are sent to a short-range data collectors, namely the sink sensor nodes, through multihop acoustic communication. Following, an autonomous

underwater vehicle (AUV) periodically visits the data collectors to retrieve data through visible light communication (VLC). Finally, mobile AUV can regularly surface to offload the gleaned data to buoys. With this data collection strategy introduced by Yan *et al.* [24], unbalanced energy dissipation problem can be mitigated at the cost of an unduly large delay. Balancing energy usage can be achieved by leveraging data collectors rather than the sensors, which are closer to the sink that rapidly deplete their battery-energies. Simulation results attain that the network lifetime is extended with the aid of topology optimization scheme, and dynamic value-based path planning strategy can enhance the value of information (VoI) on AUV. A machine learning approach is proposed by Hu *et al.* [25], who developed an adaptive and lifetime aware routing protocol based on reinforcement learning, which aims for balancing the energy use with a rewarding function throughout the routing process, where 20% longer lifetime is yielded compared to that of vector-based forwarding routing algorithm. Zorzi *et al.* [26] proposed a class of routing schemes considering main characteristics of UWSNs, such as attenuation, absorption, propagation latency and energy consumption, where the authors of [26] focused their attention on the channel characteristics effecting energy consumption and delay performances. They demonstrated that the characteristics of hop distance in radio networks also applies to underwater acoustic networks. Nonetheless, routing protocol proposed in [26] exhibited large amount of energy savings and achieved quasi-optimal energy dissipation. Zhang *et al.* [27] implemented a routing protocol that can cope with the beam width and 3D direction characteristics of acoustic sensors, which are mostly ignored by the existing underwater routing protocols. These characteristics usually lead to serious network connectivity issues and a large number of asymmetric links, which indeed steeply reduces the performance of UWSNs. Therefore, Zhang *et al.* [27] employed a link detection mechanism in order to identify link state information and adopted an adaptive routing feedback method in order to fully benefit from underwater asymmetric links, which can lead to conservation of a great amount of energy. Ensuring continuous communication and monitoring tasks from all sensor locations necessitates the replacement of energy-depleted sensors, which can be very expensive in remote underwater locations, and hence Mohapatra *et al.* [28] developed an effective policy involving with routing and node replacement decisions in order to minimize the replacement costs per unit time.

Due to the characteristics of acoustic channels, accomplishing reliable data transmission for time-critical applications is quite challenging, particularly for energy-constraint UWSNs. Conventional retransmission techniques are not applicable for time-critical applications, whereas widely-used approaches without retransmission dissipates a great amount of energy. To tackle with this particular problem, Zhou *et al.* [29] proposed a new multipath power control transmission (MPT) scheme, which ensures a satisfactory end-to-end packet error rate and a likely equal balance between the overall energy dissipation and the end-to-end packet latency. MPT consumed much less energy

Table 1: Summary of recent advances in UWSNs in terms of routing optimization and energy conservation techniques.

Author(s), (Year)	Objective(s)	Outcome	Algorithm/Approach
Latif <i>et al.</i> [10], (2016)	To repair coverage hole; identifying energy hungry spots and redundant coverage areas.	Maximized network lifetime and throughput at the cost of increased delay.	DBR and the proposed technique that moves the sensor from redundant areas to scarce areas.
Chen <i>et al.</i> [13], (2013)	To overcome energy hole problem; minimizing energy dissipation of sensors and guarantee path continuity, while maximizing data collection.	Performance improvement in successful delivery rate, power dissipation and message overhead.	Apple slice technique is proposed to build multiple layers to cover the hole and to guarantee path continuity.
Rou <i>et al.</i> [14], (2018)	To enhance the energy-efficiency of acoustic sensors.	Balanced energy dissipation throughout the UWSNs and extended network lifetime.	EULC algorithm designed with unequal layering using node depth information.
Geethu <i>et al.</i> [15], (2015)	To develop energy-efficient routing protocol that relies on the optimal hop distance.	Reduced energy dissipation, prolonged lifetime of UWSN, mitigated coverage-hole, improved packet delivery ratio and increased throughput.	An algorithm exploiting forwarding nodes that are determined based on the depth, remaining energy, and distance from the optimal hop position to the source node.
Javaid <i>et al.</i> [17], (2019)	To accomplish location-free and greedy hop-by-hop packet forwarding technique.	Extended network lifetime.	BEAR algorithm operating with the remaining energy and location information.
Guan <i>et al.</i> [18], (2019)	To mitigate void region and detoured forwarding.	Increased packet delivery ratio and energy-efficiency, and average end-to-end delay.	DVOR scheme recording the smallest hop counts toward the surface sink.
Rahman <i>et al.</i> [21], (2017)	To identify the best forwarding relay by considering energy dissipation rate and packet delivery probability.	Improved packet delivery ratio, lower end-to-end delay and better network lifetime performances.	EECOR protocol to forward packets towards the surface sink.
Jin <i>et al.</i> [23], (2018)	To forward packets to buoy with an energy-efficient technique.	Prolonged lifetime of UWSNs and higher packet delivery ratio.	EBOR protocol taking both residual energy and packet delivery probability into account for determining the next optimal hop.
Yan <i>et al.</i> [24], (2018)	To mitigate unbalanced energy dissipation with the aid of AUVs.	Extended network lifetime and enhanced VoI at the cost of excessive end-to-end delay.	AUV-assisted relaying of packets that are collected at the short-range sink sensors.
Hu <i>et al.</i> [25], (2010)	To balance the energy dissipation using a machine learning-based rewarding function throughout the routing process.	About 20% longer lifetime is yielded compared to that of vector-based forwarding routing algorithm.	An adaptive and lifetime aware routing protocol based on reinforcement learning.
Zhang <i>et al.</i> [27], (2013)	To cope with the beam width and 3D direction characteristics of acoustic sensors.	Conservation of a great amount of energy.	A link detection mechanism to acquire link state and an adaptive routing feedback method.
Zhou <i>et al.</i> [29], (2011)	To guarantee a satisfactory packet error rate while maintaining a likely equally balanced energy dissipation and end-to-end packet delay.	Conservation of a great amount of energy while providing the least end-to-end packet delay.	Multipath power control transmission (MPT) scheme.
Su <i>et al.</i> [32], (2019)	To prolong the lifetime of UWSNs.	Improved network lifetime and better end-to-end delay performances.	Adaptive deep Q-network based energy and latency-aware routing protocol (DQELR).

than the conventional one-path scheme with no-retransmission, and their results demonstrated that MPT is capable of conserving a great amount of energy while providing the least end-to-end packet delay. Wang *et al.* [30] implemented an energy-efficient data transmission scheme, so-called energy-efficiency grid routing based on 3D cubes (EGRCs), tackling with the complex nature of underwater medium, such as 3D dynamic topology, high propagation delay, mobility and density, where underwater network is modeled as a 3D cube and divided into smaller cubes with a grid-based view. Each of these cubes constitutes a cluster, which is chosen for its optimal cluster-head for minimizing energy use and end-to-end delay, as well as maintaining reliability of data transmission. Li *et al.* [31] improved an existing routing protocol, so-called depth-based routing (DBR), with an extension of handshaking-based medium access control (MAC) configurations, which enable the key sensors, undertaking the majority of packet loads, to access the channel first in the case of a collision. Simulation results demonstrated that DBR-MAC outperforms existing MAC protocols in terms of throughput, energy conservation and delay efficiency performances in data collection scenarios of UWSNs. Su *et al.* [32] designed an adaptive deep Q-network-based energy and latency-aware routing protocol (DQELR) to extend the lifetime of UWSNs. In this protocol, on/off-policy methods are adopted for making globally optimal routing decisions, where forwarders with maximum Q-value are adaptively selected based on both energy and delay. Simulation results exhibit the advantages of DQELR in terms of improved network lifetime and better latency. Quality of Service (QoS) aware reliable data transmission is challenging in underwater environment. This is mainly due to the impairments of acoustic communications, such as long propagation delays, high bit error rate, low bandwidth and multipath effects. To tackle with these impairments, a novel QoS aware evolutionary cluster based routing protocol (QERP) is proposed for UWSN applications by Faheem *et al.* [33], where QERP attained low latency, high packet delivery ratio and high energy conservation. A summary of the most recent advances in UWSNs in terms of routing optimization and energy conservation techniques is provided in Table 1.

III. FUTURE RESEARCH IDEAS

In this section, we focus our attention on the recent advances of underwater communications. Therefore, rather than reporting the classical underwater communication issues, recent advances of underwater communications are amalgamated as potential research problems.

UWSN applications are drawing attentions due to a plethora of precious resources lying underwater and waiting to be explored. However, solely underwater acoustic communication is not capable of achieving higher transmission rates due to their limited bandwidth and low velocity of propagation. These drawbacks can be circumvented with the aid of underwater optical wireless communication (UOWC) [34]-[37], which can provide higher achievable rates, lower delay and improved security. On the other hand, UOWC is also constrained by the range attainability. To augment this particular shortcoming, Celik *et*

al. [35] devised a multihop communication system using underwater optical wireless sensor networks (UOWSNs) in order to share information through the intermediate nodes. These kind of systems, however, mainly rely on the centralized localization information, which may lead to excessive delay and energy waste, as well as unduly large packet loss upon the failure of the data collector (sink sensor). Indeed, distributed localization methods are required to equally likely apportion the operational load through the entire network.

Most of the literature have considered the 2-D architecture of UWSNs for the analysis of routing optimization and energy conservation techniques. It would be more realistic to consider 3-D underwater environment for UWSN applications, where the impact of currents should be considered while devising a routing protocol. This is mainly because the depth of sensors determines the intensity and the direction of currents, which directly affect the position of sensors.

Another potential research direction for delay-tolerant UWSNs would be the operational benefits of AUVs [24]. A mobile AUV can move closer to the sink sensors to gather the collected data, where a short-range large-bandwidth communication technique, such as VLC, could be exploited. AUV can periodically report the data to buoys. With this method, longer distance communication over acoustic channels could be avoided, which is often referred to as an energy-hungry event, particularly for acoustic sensors.

Light propagation in an underwater environment reacts to wavelength with relatively less signal power attenuation within the wavelength range of blue and green (~450nm-560nm). Moreover, data rate of VLC is considerably larger than that of the underwater communication. Therefore, VLC could be potentially adopted for future high-speed underwater communication networks. For further details, motivated readers are referred to [36].

IV. CONCLUSION

We reviewed the state-of-the-art techniques of routing optimization and energy conservation in UWSNs. The most recent techniques mainly consider the mitigation of the coverage and energy hole issues. Additionally, other appealing techniques for improving the performance of UWSNs, such as mobile AUVs, VLC and UOWSNs are proposed in the literature. It is evident that hybrid communication techniques including UOWC, VLC and RF could potentially form the future of high-speed underwater communication networks.

REFERENCES

- [1] M. Erol-Kantarci, H. T. Mouftah and S. Oktug, "A Survey of Architectures and Localization Techniques for Underwater Acoustic Sensor Networks," in *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 487-502, Third Quarter 2011.
- [2] Ian F. Akyildiz, Dario Pompili, Tommaso Melodia, "Underwater acoustic sensor networks: research challenges," in *Ad Hoc Networks*, vol. 3, no 3, pp. 257-279, May 2005.
- [3] Tuna G, Cagri Gungor V., "A survey on deployment techniques, localization algorithms, and research challenges for underwater acoustic

- sensor networks," in *International Journal of Communication Systems*, vol. 30, no. 17, pp. 1074-5351, November 2017.
- [4] Felemban, E., Shaikh, F. K., Qureshi, U. M., Sheikh, A. A., and Qaisar, S. B., "Underwater Sensor Network Applications: A Comprehensive Survey," in *International Journal of Distributed Sensor*, vol. 11, no. 11, pp. 257-279, November 2015.
 - [5] Han, G., Zhang, C., Shu, L., Sun, N., & Li, Q., "A Survey on Deployment Algorithms in Underwater Acoustic Sensor Networks," in *International Journal of Distributed Sensor Networks*, vol. 9, no.12, November 2013.
 - [6] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," in *IEEE Communications Magazine*, vol. 47, no. 1, pp. 84-89, January 2009.
 - [7] Mari Carmen Domingo, Rui Prior, "Energy analysis of routing protocols for underwater wireless sensor networks," in *Computer Communications*, vol. 31, no. 6, pp. 1227-1238, April 2008.
 - [8] G. Han, J. Jiang, N. Bao, L. Wan and M. Guizani, "Routing protocols for underwater wireless sensor networks," in *IEEE Communications Magazine*, vol. 53, no. 11, pp. 72-78, November 2015.
 - [9] Muhammad Ayaz, Imran Baig, Azween Abdullah, Ibrahim Faye, "A survey on routing techniques in underwater wireless sensor networks," in *Journal of Network and Computer Applications*, vol. 34, no. 6, pp.1908-1927, November 2011.
 - [10] K. Latif, N. Javaid, A. Ahmad, Z. A. Khan, N. Alrajeh and M. I. Khan, "On Energy Hole and Coverage Hole Avoidance in Underwater Wireless Sensor Networks," in *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4431-4442, June 2016.
 - [11] F. Bouabdallah, C. Zidi and R. Boutaba, "Joint Routing and Energy Management in UnderWater Acoustic Sensor Networks," in *IEEE Transactions on Network and Service Management*, vol. 14, no. 2, pp. 456-471, June 2017.
 - [12] I. Azam, N. Javaid, A. Ahmad, W. Abdul, A. Almogren and A. Alamri, "Balanced Load Distribution With Energy Hole Avoidance in Underwater WSNs," in *IEEE Access*, vol. 5, pp. 15206-15221, 2017.
 - [13] Y. Chen and Y. Lin, "Mobicast Routing Protocol for Underwater Sensor Networks," in *IEEE Sensors Journal*, vol. 13, no. 2, pp. 737-749, Feb. 2013.
 - [14] R. Hou, L. He, S. Hu and J. Luo, "Energy-Balanced Unequal Layering Clustering in Underwater Acoustic Sensor Networks," in *IEEE Access*, vol. 6, pp. 39685-39691, 2018.
 - [15] K. S. Geethu and A. V. Babu, "Optimal hop position-based minimum energy routing protocol for underwater acoustic sensor networks," in *The Journal of Engineering*, vol. 2015, no. 5, pp. 187-196, 5 2015.
 - [16] Z. Zhou, B. Yao, R. Xing, L. Shu and S. Bu, "E-CARP: An Energy Efficient Routing Protocol for UWSNs in the Internet of Underwater Things," in *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4072-4082, June1, 2016.
 - [17] N. Javaid, S. Cheema, M. Akbar, N. Alrajeh, M. S. Alabed and N. Guizani, "Balanced Energy Consumption Based Adaptive Routing for IoT Enabling Underwater WSNs," in *IEEE Access*, vol. 5, pp. 10040-10051, 2017.
 - [18] Q. Guan, F. Ji, Y. Liu, H. Yu and W. Chen, "Distance-Vector based Opportunistic Routing for Underwater Acoustic Sensor Networks," in *IEEE Internet of Things Journal*, [Early Access].
 - [19] R. Coutinho, A. Boukerche, L. F. Menezes Vieira and A. A. Loureiro, "A Joint Anypath Routing and Duty-Cycling Model for Sustainable Underwater Sensor Networks," in *IEEE Trans. on Sustainable Computing*, [Early Access].
 - [20] C. Hsu, H. Liu, J. L. García Gómez and C. Chou, "Delay-Sensitive Opportunistic Routing for Underwater Sensor Networks," in *IEEE Sensors Journal*, vol. 15, no. 11, pp. 6584-6591, Nov. 2015.
 - [21] M. A. Rahman, Y. Lee and I. Koo, "EECOR: An Energy-Efficient Cooperative Opportunistic Routing Protocol for Underwater Acoustic Sensor Networks," in *IEEE Access*, vol. 5, pp. 14119-14132, 2017.
 - [22] Y. Noh *et al.*, "HydroCast: Pressure Routing for Underwater Sensor Networks," in *IEEE Transactions on Vehicular Technology*, vol. 65, no. 1, pp. 333-347, Jan. 2016.
 - [23] Z. Jin, Z. Ji and Y. Su, "An Evidence Theory Based Opportunistic Routing Protocol for Underwater Acoustic Sensor Networks," in *IEEE Access*, vol. 6, pp. 71038-71047, 2018.
 - [24] J. Yan, X. Yang, X. Luo and C. Chen, "Energy-Efficient Data Collection Over AUV-Assisted Underwater Acoustic Sensor Network," in *IEEE Systems Journal*, vol. 12, no. 4, pp. 3519-3530, Dec. 2018.
 - [25] T. Hu and Y. Fei, "QELAR: A Machine-Learning-Based Adaptive Routing Protocol for Energy-Efficient and Lifetime-Extended Underwater Sensor Networks," in *IEEE Transactions on Mobile Computing*, vol. 9, no. 6, pp. 796-809, June 2010.
 - [26] M. Zorzi, P. Casari, N. Baldo and A. F. Harris, "Energy-Efficient Routing Schemes for Underwater Acoustic Sensor Networks," in *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 9, pp. 1754-1766, December 2008.
 - [27] S. Zhang, D. Li and J. Chen, "A Link-State Based Adaptive Feedback Routing for Underwater Acoustic Sensor Networks," in *IEEE Sensors Journal*, vol. 13, no. 11, pp. 4402-4412, Nov. 2013.
 - [28] A. K. Mohapatra, N. Gautam and R. L. Gibson, "Combined Routing and Node Replacement in Energy-Efficient Underwater Sensor Networks for Seismic Monitoring," in *IEEE Journal of Oceanic Engineering*, vol. 38, no. 1, pp. 80-90, Jan. 2013.
 - [29] Z. Zhou, Z. Peng, J. Cui and Z. Shi, "Efficient Multipath Communication for Time-Critical Applications in Underwater Acoustic Sensor Networks," in *IEEE/ACM Transactions on Networking*, vol. 19, no. 1, pp. 28-41, Feb. 2011.
 - [30] K. Wang, H. Gao, X. Xu, J. Jiang and D. Yue, "An Energy-Efficient Reliable Data Transmission Scheme for Complex Environmental Monitoring in Underwater Acoustic Sensor Networks," in *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4051-4062, June 2016.
 - [31] C. Li, Y. Xu, B. Diao, Q. Wang and Z. An, "DBR-MAC: A Depth-Based Routing Aware MAC Protocol for Data Collection in Underwater Acoustic Sensor Networks," in *IEEE Sensors Journal*, vol. 16, no. 10, pp. 3904-3913, May, 2016.
 - [32] Y. Su, R. Fan, X. Fu and Z. Jin, "DQELR: An Adaptive Deep Q-Network-Based Energy- and Latency-Aware Routing Protocol Design for Underwater Acoustic Sensor Networks," in *IEEE Access*, vol. 7, pp. 9091-9104, 2019.
 - [33] M. Faheem, G. Tuna and V. C. Gungor, "QERP: Quality-of-Service (QoS) Aware Evolutionary Routing Protocol for Underwater Wireless Sensor Networks," in *IEEE Systems Journal*, vol. 12, no. 3, pp. 2066-2073, Sept. 2018.
 - [34] N. Saeed, A. Celik, S. Alouini and T. Y. Al-Naffouri, "Performance Analysis of Connectivity and Localization in Multi-Hop Underwater Optical Wireless Sensor Networks," in *IEEE Transactions on Mobile Computing*, [Early Access].
 - [35] A. Celik, N. Saeed, T. Y. Al-Naffouri, and M. Alouini, "Modeling and performance analysis of multihop underwater optical wireless sensor networks," in *IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1-6.
 - [36] C. Wang, H. Yu and Y. Zhu, "A Long Distance Underwater Visible Light Communication System With Single Photon Avalanche Diode," in *IEEE Photonics Journal*, vol. 8, no. 5, pp. 1-11, Oct. 2016, Art no. 7906311.
 - [37] Z. Zeng, S. Fu, H. Zhang, Y. Dong and J. Cheng, "A Survey of Underwater Optical Wireless Communications," in *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 204-238, Firstquarter 2017.

An Application of Artificial Bee Colony Algorithm to Fatigue Life Estimation of Magnesium Alloy

S. KARAGÖZ¹, C.B. KALAYCI^{1,*} and Ö. KARAKAŞ²

¹ Pamukkale University, Department of Industrial Engineering, Denizli/Turkey, cbkalayci@pau.edu.tr

² Pamukkale University, Department of Mechanical Engineering, Denizli /Turkey, okarakas@pau.edu.tr

*Corresponding Author

Abstract - In this study, it is aimed to estimate the fatigue life of magnesium alloy by an artificial bee colony algorithm. Since notch factor, stress rate and stress amplitude factors influence the fatigue life, mathematical function models have been utilized for the solution of the problem and the parameters of the function optimized with artificial bee colony algorithm. The predicted results were compared with the experimental results. These results show that heuristic algorithms can be successfully applied in fatigue life estimation.

Keywords - Artificial Bee Colony Algorithm, Fatigue Life Estimation, Heuristic Algorithms, Magnesium Alloy

I. INTRODUCTION

FATIGUE life is the period of the material resistance to breakage or cracking with different periodic stresses. Fatigue means that a material becomes weary and fails at the level of stress below the nominal power of the material. The accidents caused by fatigue failure not only severely harmed the national economy but also seriously threatened the safety of human life [1]. Hence, it is important that the estimates are accurate and reliable. Therefore, reliable methodologies are needed to estimate fatigue life.

In this study, fatigue life estimation of magnesium alloy is investigated by heuristic algorithms from a parameter optimization point of view. The use of conventional estimation methods requires a large amount of data and experiments since fatigue life prediction studies are both costly and time consuming. Therefore, estimates using heuristic algorithms will save time and resources by minimizing the need for expensive and time-consuming fatigue tests.

Fatigue life estimation has long been the interest of research in the literature. Although the use of heuristic algorithms is limited in this topic, the most commonly used algorithm is the genetic algorithm [2-4]. In addition to the genetic algorithm, particle swarm optimization [5], [1], ant colony optimization [6], simulated annealing [7] algorithms are also adopted for fatigue life estimation of composite structures.

There are also hybrid frameworks where more than one method is combined to fatigue life estimation. Ma et al. [8] proposed a hybrid genetic neural network method, obtained by optimizing the parameters of artificial neural networks with a

genetic algorithm. Majidian and Saidi [9] used fuzzy logic, genetic algorithm and artificial neural network methods to determine the remaining life of the tubes of the boiler of the Neka power plant in the north of Iran. Mohanty [10] applied the Adaptive Neuro-Fuzzy Inference System approach for estimating the growth rate of fatigue crack under constant amplitude loading and fatigue life of aluminum alloy stating that can be useful as an alternative estimation method. Moghaddam et al. [11] applied support vector machines and firefly algorithm to fatigue life estimation of polyethylene terephthalate modified asphalt mixture where the results were compared with artificial neural networks and genetic programming approaches confirming the accuracy. Deveci and Artem [12] utilized a hybrid genetic algorithm for fatigue life estimation of a composite material. In recent years, evolutionary algorithms [13] are applied to fatigue life estimation while expert systems [14] are also designed. Considering the development in technology and computational power, these methods are expected to be more popular in the near future.

In this study for the fatigue life estimation of the magnesium alloy, firstly an exponential-trigonometric function model is proposed. Then, the parameters of this mathematical function are optimized with an artificial bee colony algorithm. The prediction results are compared with the original experimental results.

II. HEURISTIC ALGORITHMS

Many of the real-life optimization problems can be far too complex to be solved by exact solution methodologies and may take a long time to solve, but the desired result may not be achieved. In such situations, while trying to find solutions to problems, heuristic algorithms are preferred since heuristic algorithms do not guarantee optimal solution, yet can achieve a solution that is reasonably close to the optimal.

In this study, relatively new to the literature compared to other heuristics, artificial bee colony algorithm is adopted for fatigue life estimation of magnesium alloy.

Artificial Bee Colony Algorithm

One of the heuristic algorithms developed for the solution of optimization problems is the artificial bee colony algorithm proposed by Karaboga in 2005 [15]. The artificial bee colony algorithm, based on flock intelligence, was inspired by the behavior of honey bees in search for food.

Artificial bee colony algorithm is mostly used for solution of various optimization problems [16] such as vehicle routing problems [17], portfolio optimization problems [18], job shop scheduling problem [19], assembly line sequencing and balancing problems [20].

The operation of the algorithm begins with the leave of the bee from the hive to search for food and continues with randomized searches. If the amount of food in the source is reduced, the bees start to look for new food sources according to the information received from other bees. Bees obtain this information via waggle dances [21]. They share information about the direction, distance and nectar quality of the food source with other bees.

Artificial bee colony consists of three kinds of bees: employed bees, onlooker bee and scout bees. The employed bees investigate the food sources in which the amount of nectar is high. The number of food sources is equal to the number of employed bees. Onlooker bees utilize the information received from the employed bees about the food resources in order to further exploit the food sources where the nectar amount is at the highest. Scout bees are on the other hand, responsible for the discovery process. They search for food sources by random dispersion. When the nectar amount of a food source is fully consumed, the bee in that source turns into a scout bee and starts to search for different food sources [22].

The first step in the algorithm is the release of the employed bees onto the food sources. The nectar quality of a food source is represented by the objective function of the solution [21]. There is also a limit value, a predetermined number of attempts, that decides to unleash scout bees for randomized search.

In the first step, the quality of the food source is calculated using Equation (1):

$$x_{ij} = x_j^{min} + rand(0,1)(x_j^{max} - x_j^{min}) \quad (1)$$

Where, $i = \{1,2, \dots, NS\}$, "NS" the number of food sources, $j = \{1,2, \dots, D\}$, "D" is the parameter number of the problem, " x^{min} " and " x^{max} " are lower and upper limits in each dimension, respectively [23]. The food quality of this source, i.e. the objective function value is calculated. If the obtained solution value is better than the previous solution value, information about this food source is stored. If the solution value cannot be improved, the limit value is increased by one; otherwise the trial will be reset [22].

The employed and onlooker bees utilize the formulation given by Equation (2) to improve the solution throughout generations of the algorithm.

$$v_{ij} = x_{ij} + \Phi_{ij}(x_{ij} - x_{kj}) \quad (2)$$

Where represents, " v_i " a new solution in the adjacent of x_i , " x_i " the current solution chosen by the employed bee, " k " a random number selected in the range $[1, NS]$, " x_k " a randomly different solution selected from the food source population, " j " a randomly selected parameter in the range $[1, D]$ and " Φ_{ij} " a random value selected in the range $[-1, 1]$ [24].

All the employed bees share their knowledge about the source with onlooker bees after the food search phase is completed. The onlooker bee selects the food source in proportion to the amount of nectar, and these probability values are calculated by the Equation (3) [23]:

$$p_i = \frac{fit_i}{\sum_{i=1}^{NS} fit_i} \quad (3)$$

Where " fit_i " represents the fitness value of the source i . "NS" indicates the number of employed bees and " p_i " represents the probability values for each solution.

After all employed and onlooker bees have completed their tasks, the limit parameter value is checked. If the value of the trial reaches the limit value, the bee in that source becomes a scout bee and searches for a new food source by using Equation (1) [22]. All these steps to find the best source form one cycle of the algorithm and continue until the maximum number of cycles is reached.

III. THE PROPOSED METHODOLOGY

In order to estimate the fatigue life of the material under cyclic stress, it is very important to understand the fatigue behavior of magnesium alloys.

Fatigue is related to the micro structure, as fatigue is often caused by local stress build-up around defects in the micro structure. A material has a single micro structure condition before welding. After welding, three different micro structure conditions occur: base material, heat affected zone and weld metal. After welding, there is the base material at the edges, weld metal at the center, heat-affected zone between the weld metal and the base material.

There are three input values for the problem: notch factor, stress amplitude and stress ratio. The output value is the number of cycles corresponding to a certain stress level, i.e. fatigue life.

- Notch Factor (K_t): Notch is the sudden section change in a material. The non-notched parts show higher fatigue life. Because the fatigue life is inverse proportional with stress [25]. In the data set we utilized, there are two different K_t values: $K_t = 1$ and $K_t = 11,2$.

- Stress Ratio (R): There are two stress ratio values in the dataset. These are " $R = 0$ " and " $R = -1$ ".

- Stress Amplitude (σ): The greater the stress amplitude and mean stress, the shorter the life of the material.

The accuracy of estimation depends on both the characteristics of the data and the prediction model used. In fatigue life estimation, the parameters of the function should be optimized to minimize the error and reveal the most accurate estimation values. The proposed exponential-trigonometric function model can be written as follows:

$$f = |w_1| + |w_2|e^{w_3x_1+w_4x_2+w_5x_3+w_6}|\tanh(w_7x_1 + w_8x_2 + w_9x_3 + w_{10})| \quad (4)$$

Where " w_i " represent to the weight factors of the parameters. The normalization procedure of the input values is made according to the Equation (5).

$$X = \frac{(X - X_{min})}{(X_{max} - X_{min})} \quad (5)$$

Different results are obtained according to the values of some variables (limit value, iteration number, modification rate) used in artificial bee colony algorithm. In order to obtain the best results, full factorial set of experiments were carried out.

After determining the best parameter combination, the algorithm is executed for 10^4 iterations. Experimental data presented by Karakas [26] are used to evaluate the fatigue life of the welded joints from magnesium alloys.

22 sample data sets for the base material, 23 data sets for the weld metal, and 20 data sets for heat affected zone were considered in the algorithm runs. Fatigue life prediction results of magnesium alloy were obtained as shown in Figure 1, Figure 2, Figure 3, respectively. Error calculations have been made between the predictive and experimental results to demonstrate the accuracy and effectiveness of the proposed prediction models using Equation (6).

$$Relative\ Error = \frac{\sum_1^n (Predictive - Experimental)^2}{n} \quad (6)$$

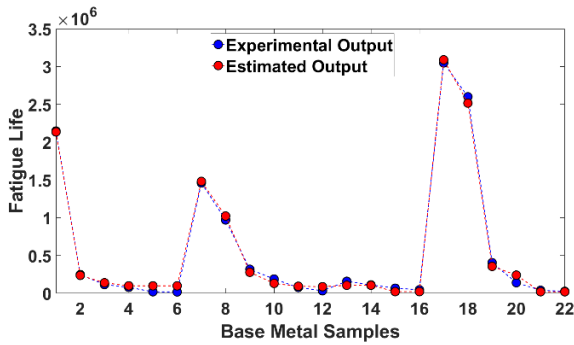


Figure1: Experimental Output vs. Estimated Output for Base Metal

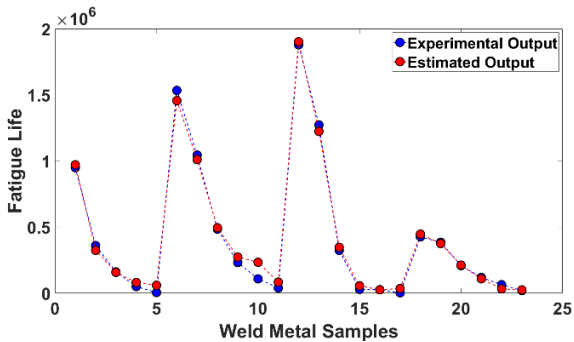


Figure2: Experimental Output vs. Estimated Output for Weld Metal

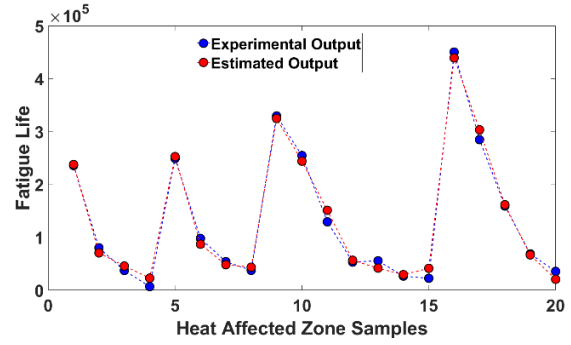


Figure3: Experimental Output vs. Estimated Output for Heat Affected Zone

The normalized mean square error is below 0.01 for all samples in just 10^4 iterations of the algorithm run while the sum of the average error is 16.37%, 33.85% and 26.41% for the base metal, weld metal and heat affected zones, respectively. The results showed that artificial bee colony algorithm is an effective tool in fatigue life estimation of the materials.

IV. CONCLUSION

In this study, fatigue life of magnesium alloy was estimated by using an application of the artificial bee colony algorithm. The results showed that the artificial bee colony algorithm was a successful method for fatigue life prediction achieved in a short time with a high estimation ability and simple operation principle. The algorithm can also be applied to the fatigue life estimation of different engineering structures and materials.

ACKNOWLEDGEMENT

This study was supported by Scientific Research Coordination Unit of Pamukkale University under the project number 2018FEBO37.

REFERENCES

- [1] Yang, X., Zou, L., Deng, W., "Fatigue Life Prediction for Welding Components Based On Hybrid Intelligent Technique", *Materials Science & Engineering A*, 642, pp. 253-261, (2015).
- [2] Canyurt, O.E., "Fatigue Strength Estimation of Adhesively Bonded Tubular Joint Using Genetic Algorithm Approach", *International Journal of Mechanical Sciences*, 46, pp. 359-370, (2004).
- [3] Karakas, O., Canyurt, O.E., Gulsoz, A., "Fatigue Strength Estimation of Butt Welded Joints in Magnesium AZ31 Alloy Using the Genetic Algorithm", *Materialwissenschaft und Werkstofftechnik*, 39 (3), pp. 234-240, (2008).
- [4] Canyurt, O. E., Meran, C., "Fatigue Strength Estimation of Adhesively Bonded Tongue and Groove Joint of Thick Woven Composite Sandwich Structures Using Genetic Algorithm Approach", *International Journal of Adhesion and Adhesives*, 33, pp. 80-88, (2012).
- [5] Ertas, A. H., "Optimization of Fiber Reinforced Laminates for Maximum Fatigue Life Using Particle Swarm Optimization", *Procedia Engineering*, 38, pp. 473-478, (2012).
- [6] Yang, X., Deng, W., Zou, L., Zhao, H., Liu, J. "Fatigue Behaviors Prediction Method of Welded Joints Based On Soft Computing Methods", *Materials Science and Engineering: A*, 559, pp. 574-582, (2013).
- [7] Ertas, A. H., Sonmez, F. O., "Design Optimization of Fiber-Reinforced Laminates for Maximum Fatigue Life", *Journal of Composite Materials*, 48 (20), pp. 2493-2503, (2014).
- [8] Ma, D., Chen, Z., Shan, X., "Forecasting of the Fatigue Life of Metal Weld Joints Based on Combined Genetic Neural Network", *Key Engineering Materials*, 439-440, 1662-9795, pp. 195-201, (2010).

- [9] Majidian, A., Saidi, M. H., "Comparison of Fuzzy Logic and Neural Network in Life Prediction of Boiler Tubes", *International Journal of Fatigue*, 29, pp. 489-498, (2007).
- [10] Mohanty, J.R., "Fatigue Crack Growth Life Prediction of 6061 Al-Alloy Under Load Ratio Effect by Using ANFIS", *International Journal of Advanced Engineering Research and Science*, 3 (11), pp. 199-204, (2016).
- [11] Moghaddam, T. B., Soltani, M., Shahraki, H. S., Shamshirband, S., Noor, N. B. M., Karim, M. R., "The Use of SVM-FFA in Estimating Fatigue Life of Polyethylene Terephthalate Modified Asphalt Mixtures", *Measurement*, 90, pp. 526-533, (2016).
- [12] Deveci, H. A., Artem, H. S., "Optimum Design of Fatigue-Resistant Composite Laminates Using Hybrid Algorithm", *Composite Structures*, 168, pp. 178-188, (2017).
- [13] Vassilopoulos, A.P., Georgopoulos, E.F., Keller, T., "Comparison of Genetic Programming with Conventional Methods for Fatigue Life Modeling of FRP Composite Materials", *International Journal of Fatigue*, 30 (9), pp. 1634-1645, (2008).
- [14] Liu, X., Xuan, F.Z., Si, J., Tu, S.T., "Expert System for Remnant Life Prediction of Defected Components Under Fatigue and Creep-Fatigue Loadings", *Expert Systems with Applications*, 34 (1), pp. 222-230, (2008).
- [15] Gao, W., Liu, S., "A Modified Artificial Bee Colony Algorithm", *Computers and Operations Research*, 39, pp. 687-697, (2012).
- [16] Xiang, W., Ma, S., An, M., "hABCDE: A Hybrid Evolutionary Algorithm Based On Artificial Bee Colony Algorithm and Differential Evolution", *Applied Mathematics and Computation*, 238, pp. 370-386, (2014*).
- [17] Yao, B., Hu, P., Zhang, M., Wang, S., "Artificial Bee Colony Algorithm with Scanning Strategy for The Periodic Vehicle Routing Problem", *Simulation*, 89(6), pp. 762-770, (2013).
- [18] Chen, W., "An Artificial Bee Colony Algorithm for Uncertain Portfolio Selection", *The Scientific World Journal*, 2014, pp. 1-12, (2014).
- [19] Zhang, R., Song, S., Wu, C., "A Hybrid Artificial Bee Colony Algorithm for The Job Shop Scheduling Problem", *International Journal of Production Economics*, 141(1), pp. 167-178, (2013).
- [20] Kalayci, C. B., Gupta, S. M., "Artificial Bee Colony Algorithm for Solving Sequence-Dependent Disassembly Line Balancing Problem", *Expert Systems with Applications*, 40 (18), pp. 7231-7241, (2013).
- [21] Kang, F., Li, J., Xu, Q., "Structural Inverse Analysis by Hybrid Simplex Artificial Bee Colony Algorithms", *Computers and Structures*, 87, pp. 861-870, (2009).
- [22] Akay, B., Karaboga, D., "A Modified Artificial Bee Colony Algorithm for Real-Parameter Optimization", *Information Sciences*, 192, pp. 120-142, (2012).
- [23] Gao, W., Liu, S., "A Modified Artificial Bee Colony Algorithm", *Computers and Operations Research*, 39, pp. 687-697, (2012).
- [24] Zhu, G., Kwong, S., "Gbest-Guided Artificial Bee Colony Algorithm for Numerical Function Optimization", *Applied Mathematics and Computation*, 217, pp. 3166-3173, (2010).
- [25] Karakas, O., Gulsoz, A., "Kaynaklı Birleştirmelerin Statik ve Yorulma Dayanımına Etki Eden Faktörler", *Mühendis ve Makine*, 48 (573), pp. 10-17, (2012).
- [26] Karakas, O., "Biçimlenebilen Magnezyum Alaşımlarından Kaynaklı Yapı Elemanlarının Yorulma Dayanımı Değerlendirmelerinde Çentik Gerilmesi Yönteminin Uygulanması", Ph. D Thesis, *Pamukkale University Institute of Science*, Denizli, (2006).

An Improved Crow Search Algorithm for Continuous Optimization Problems

Emrullah SONUÇ¹

¹ Karabuk University, Karabuk/Turkey, esonuc@karabuk.edu.tr

Abstract - This paper proposes an improved version of Crow Search Algorithm (CSA), which is inspired by behavior of crows. CSA has a two state to update the position in a search space. In this paper, second state of CSA is improved using an existing update model used in Artificial Bee Colony which is popular optimization algorithm in recent years. An improved CSA (I-CSA) is tested with five mathematical optimization benchmark problems. Optimization results show that I-CSA is very competitive compared to the CSA. Also, I-CSA has rapid convergence compared to the CSA for tested problems.

Keywords – crow search algorithm, optimization, swarm intelligence.

I. INTRODUCTION

THIS paper presents an improved version of Crow Search Algorithm for solving continuous optimization problems. Nowadays, optimization algorithms are widely inspired by behavior of animals who live or hunt together [1]. This togetherness introduces a new concept, Swarm Intelligence (SI). SI is the natural or artificial collective behavior of decentralized, self-organized system. Many optimization algorithms use concept of SI and are presented in literature. Recently presented algorithms are Crow Search Algorithm [2], Artificial Bee Colony [3], Whale Optimization Algorithm [4], Grey Wolf Optimizer [5], Squirrel Search Algorithm [6], The Sailfish Optimizer [7], Harris Hawks Optimization [8] and Selfish Herd Optimizer [9].

Crow Search Algorithm (CSA) is developed by Askarzadeh for solving constrained engineering optimization problems. CSA is obtained good quality results not only for constrained optimization problems but also unconstrained optimization problems. Main property of CSA; it has only two parameters and this causes easy to implement for solving any optimization problem. If an optimization algorithm has a good exploration and exploitation, then this algorithm is a robust algorithm. In CSA, exploration and exploitation was balanced and this causes very useful for solving optimization problems.

This paper is organized as follows: In Section II, CSA is described. An improved version of CSA for solving continuous optimization problems is presented in Section III. Section IV gives details about computational experiments on some benchmark problems and results. Finally, Section V states some conclusions and future work.

II. CROW SEARCH ALGORITHM

CSA [2] is a population-based algorithm and inspired by behavior of crows. Crows live in hives and can follow other birds and steal food from their hives. They can remember where the other birds are hiding and where are foods of others. Inspired by the behavior of crows, flowchart of CSA is shown in Figure 1.

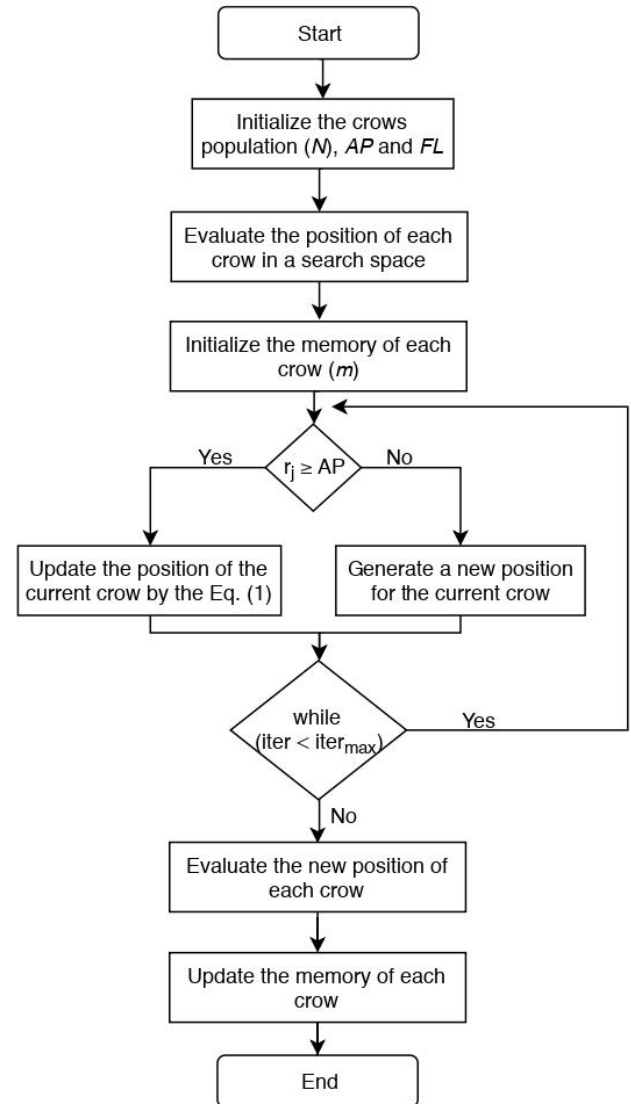


Figure 1: Flowchart of CSA.

Implementation of CSA to any optimization problem is quite easy because it has only two parameters, these are: Awareness Probability (AP) and Flight Length (FL). CSA has a two state for updating position in a search space. For the first state, each crow (*crow i*) selects a random crow (*crow j*) to steal food from its hive without being noticed. This state is determined by the parameter AP. For the first state new position is calculated using (1) following:

$$x^{i,iter+1} = x^{i,iter} + r_i \cdot fl^{i,iter} \cdot (m^{j,iter} - x^{i,iter}) \quad (1)$$

In (1), *m* is memory of *crow j* and *r* is a random number between 0 and 1. For the second state, each crow moves to a new position randomly in a search space. For the second state new position is calculated using (2) following:

$$x^{i,iter+1} = \begin{cases} x^{i,iter+1} = x^{i,iter} + r_i \cdot fl^{i,iter} \cdot (m^{j,iter} - x^{i,iter}) & r_j \geq AP^{j,iter} \\ a \text{ random position} & \text{otherwise} \end{cases} \quad (2)$$

III. AN IMPROVED CSA IMPLEMENTATION FOR OPTIMIZATION

Many metaheuristics were improved to get better position in a search space [10]–[14]. This paper aims to improve second state of CSA. In second state of CSA: Crow *j* knows that the other crow *i* is following it. As a result, crow *j* wants to protect its hive. Then crow *j* goes to another position randomly for protection of its hive. Although CSA is good in constrained optimization problems, it is not good enough in continuous optimization problems. To overcome of this problem, second state of CSA improved using formula following:

$$x^{ij,iter+1} = m^{ij,iter} + r_{ij} \cdot (m^{ij,iter} - m^{kj,iter}) \quad (3)$$

where *k* and *j* are random selected index. This formula is used in Artificial Bee Colony [3] which is presented by Karaboga and Basturk. According to this concept, a crow produces a modification on the position in its memory depending on local information and finds a food source from another crow.

Table 1: Benchmark Functions.

No. of Function	Function	Function Name	Search Space	Type
F1	$\sum_{i=1}^n x_i^2$	Sphere	$[-100, 100]^D$	Unimodal
F2	$\sum_{i=1}^{n-1} (100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2)$	Rosenbrock	$[-30, 30]^D$	Unimodal
F3	$\frac{1}{4000} \sum_{i=1}^n (x_i)^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	Griewank	$[-600, 600]^D$	Multimodal
F4	$\sum_{i=1}^n x_i + \prod_{i=1}^n x_i $	Schwefel	$[-10, 10]^D$	Unimodal
F5	$-20 \exp(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}) - \exp(\frac{1}{n} \sum_{i=1}^n \cos 2\pi x_i) + 20 + e$	Ackley	$[-32, 32]^D$	Multimodal

Table 2: The comparison results of CSA and I-CSA on the 10-dimensional benchmark functions.

	Algorithm	Best	Mean	Worst	Median	Stdev.	Time (s)
F1 (Sphere)	CSA	7.58E-11	1.36E-01	3.39E+00	3.90E-10	6.27E-01	0.640
	I-CSA	5.01E-63	4.63E-38	1.39E-36	4.79E-61	2.54E-37	0.696
F2 (Rosenbrock)	CSA	4.09E-03	5.06E+00	7.13E+00	5.49E+00	1.53E+00	0.814
	I-CSA	1.02E-02	5.05E+01	1.28E+03	3.73E+00	2.33E+02	0.887
F3 (Griewank)	CSA	1.71E-10	5.21E-02	8.01E-01	1.12E-08	1.54E-01	0.789
	I-CSA	0.00E+00	5.11E-03	8.26E-02	0.00E+00	1.95E-02	0.811
F4 (Schwefel)	CSA	1.43E-06	5.74E-03	1.72E-01	2.73E-06	3.14E-02	0.683
	I-CSA	9.25E-37	1.89E-02	3.10E-01	2.89E-35	7.15E-02	0.724
F5 (Ackley)	CSA	2.40E-06	2.70E-02	8.09E-01	7.64E-06	1.48E-01	0.702
	I-CSA	4.00E-15	7.31E-15	7.51E-14	4.00E-15	1.36E-14	0.746

IV. EXPERIMENTS FOR IMPROVED CSA

In the computational experiments, Improved CSA (I-CSA) was applied for finding the global minimum of the well-known five benchmark functions. The numeric benchmark functions used in the experiments have some characteristics. If a function has more than one local optimum, this function is called as multimodal. If a function has only one local optimum (it is also global optimum), this function is called as unimodal. Five benchmark functions are Sphere, Rosenbrock, Griewank, Schwefel and Ackley and presented in Table 1. For all benchmark functions, the global optimum values are 0. The methods are run on the 10-dimensional function for all tests. Both in CSA and improved CSA, AP and FL values were set to 0.1 and 2, respectively. Number of fitness evaluations (NFEs) is 2,000 for both algorithms. CSA and I-CSA are run 10 times for solving 10 dimensional functions. Best, mean, worst, median and standard deviation values of the results are reported in Table 2. Table 2 shows the results obtained by I-CSA in comparison with the results found by CSA over 10 independent runs. It is seen, on all the functions except F2, I-CSA outperforms the CSA in terms of the best index. For F1, I-CSA outperforms the CSA in terms of best, mean, worst, median and standard deviation indices. According to standard deviation values for F1, I-CSA is more robust than CSA. For F2, CSA outperforms the I-CSA in terms of best, mean, worst and standard deviation indices. On the other hand, the best results are not far away between I-CSA and CSA for F2. For F3, I-CSA outperforms the CSA in terms of best, mean, worst, median and standard deviation indices. According to the results, I-CSA obtained the optimum value for F3 more than one times. For F4, I-CSA outperforms the CSA in terms of best and median indices. Standard deviation results are not far away between I-CSA and CSA for F4 so that both algorithms are robust for this function. For F5, I-CSA outperforms the CSA in terms of best, mean, worst, median and standard deviation indices. According to standard deviation values for F5, I-CSA is more robust than CSA.

In Table 2, the average time of the runs is shown. It is seen that CSA has less computational time than I-CSA over equal number of fitness evaluations.

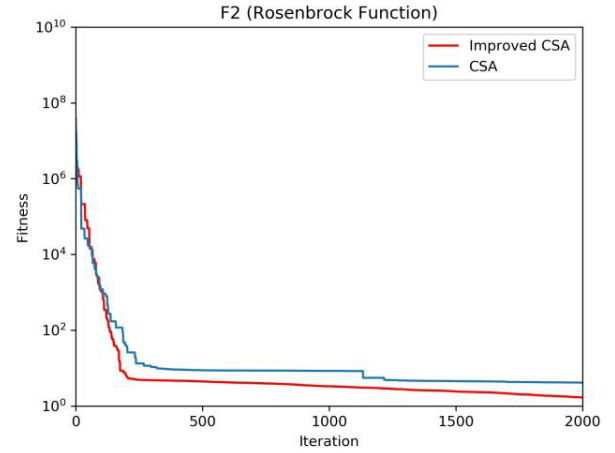


Figure 3: Comparison of convergence curves for F2.

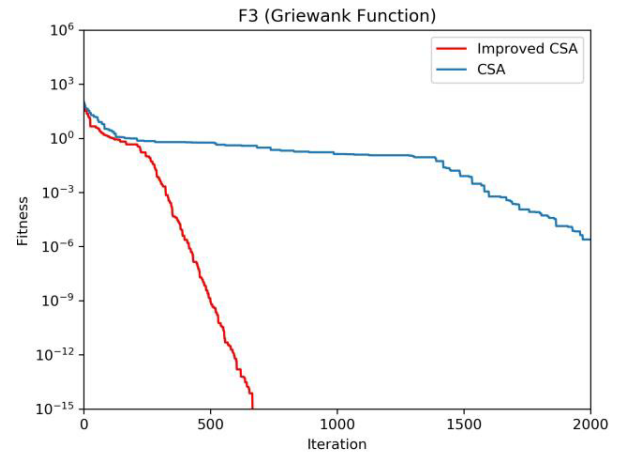


Figure 4: Comparison of convergence curves for F3.

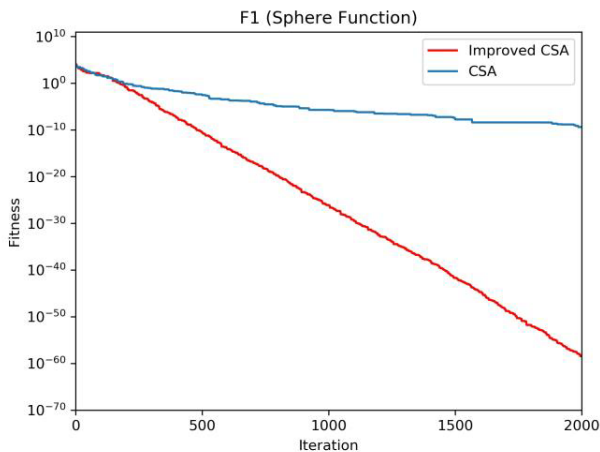


Figure 2: Comparison of convergence curves for F1.

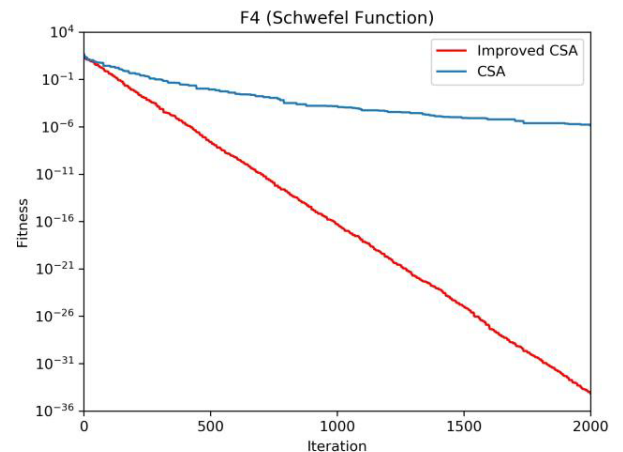


Figure 5: Comparison of convergence curves for F4.

The convergence curves of the I-CSA and CSA are provided in Figure 2 to Figure 6 to see the convergence rate of the algorithms. In general, I-CSA has rapid convergence from the initial steps of iterations as can be seen in Figure 4.

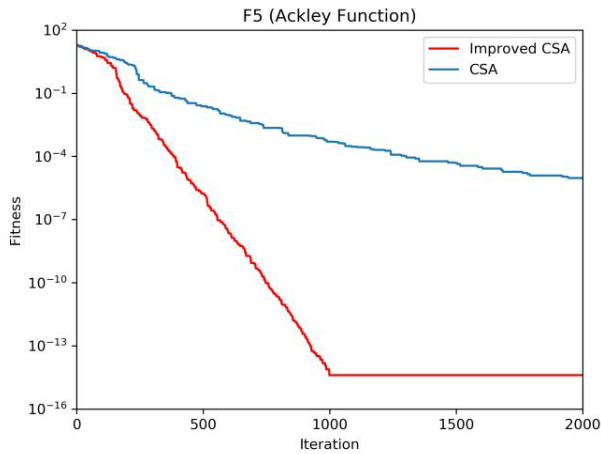


Figure 6: Comparison of convergence curves for F5.

V.CONCLUSION

This paper presented an improved version of CSA for continuous optimization problems. Experiments were conducted on five benchmark functions to analyze achieving global optimum and convergence behavior of the proposed algorithm. I-CSA was found to be enough competitive compared to the CSA. In future, I-CSA can be applied to constrained optimization problems.

REFERENCES

- [1] X.-S. Yang, *Nature-inspired Metaheuristic Algorithms*. Luniver Press, 2010.
- [2] A. Askarzadeh, "A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm," *Computers & Structures*, vol. 169, pp. 1–12, Jun. 2016.
- [3] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," *Journal of global optimization*, vol. 39, no. 3, pp. 459–471, 2007.
- [4] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, May 2016.
- [5] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, Mar. 2014.
- [6] M. Jain, V. Singh, and A. Rani, "A novel nature-inspired algorithm for optimization: Squirrel search algorithm," *Swarm and Evolutionary Computation*, vol. 44, pp. 148–175, Feb. 2019.
- [7] S. Shadravan, H. R. Naji, and V. K. Bardsiri, "The Sailfish Optimizer: A novel nature-inspired metaheuristic algorithm for solving constrained engineering optimization problems," *Engineering Applications of Artificial Intelligence*, vol. 80, pp. 20–34, Apr. 2019.
- [8] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, Aug. 2019.
- [9] F. Fausto, E. Cuevas, A. Valdivia, and A. González, "A global optimization algorithm inspired in the behavior of selfish herds," *Biosystems*, vol. 160, pp. 39–55, Oct. 2017.
- [10] P. Díaz *et al.*, "An Improved Crow Search Algorithm Applied to Energy Problems," *Energies*, vol. 11, no. 3, p. 571, Mar. 2018.
- [11] G. I. Sayed, A. Tharwat, and A. E. Hassanien, "Chaotic dragonfly algorithm: an improved metaheuristic algorithm for feature selection," *Appl Intell*, vol. 49, no. 1, pp. 188–205, Jan. 2019.
- [12] W. Xiang, X. Meng, Y. Li, R. He, and M. An, "An improved artificial bee colony algorithm based on the gravity model," *Information Sciences*, vol. 429, pp. 49–71, Mar. 2018.
- [13] T. Meng, Q.-K. Pan, J.-Q. Li, and H.-Y. Sang, "An improved migrating birds optimization for an integrated lot-streaming flow shop scheduling problem," *Swarm and Evolutionary Computation*, vol. 38, pp. 64–78, Feb. 2018.
- [14] M. Abdel-Basset, G. Manogaran, L. Abdel-Fatah, and S. Mirjalili, "An improved nature inspired meta-heuristic algorithm for 1-D bin packing problems," *Pers Ubiquit Comput*, vol. 22, no. 5–6, pp. 1117–1132, Oct. 2018.

Analysis of Electronic Countermeasure Effects on Air Defense Systems

G.B.KIRAZ¹ and A.G.PAKFILIZ¹

¹Baskent University, Ankara/Turkey, gbkiraz@gmail.com

¹Baskent University, Ankara/Turkey, apakfiliz@baskent.edu.tr

Abstract - Air defense systems are the most important obstacles in the mission path for air assault operations. Well designed and deployed air defense elements can't be overcome by small casualties which only use countermeasure capabilities. A well designed and deployed air defense system can't be overcome with small casualties by only using self-electronic countermeasure capabilities. For the task to be performed with an acceptable loss additional countermeasure capability are required. For this purpose, enemy air defense systems are to be suppressed by using additive jamming assets (soft kill) or anti-radiation missiles (hard kill).

In this study, an operation area has been modeled for analyzing the effectiveness of electronic countermeasures on an air defense system in an air assault operation.

Keywords – Stand-Off Jamming, Escort Jamming, Noise Jamming

I. INTRODUCTION

Suppression of enemy air defense (SEAD) operations are force multipliers that increases the probability of a successful air task, especially for destruction of critical targets. The SEAD has a major role in air operations, and it aims to neutralize, destruct or temporary degrade surface-based enemy air defense system or its sub-components by destructive or disruptive means. Thus, task force aircrafts can approach the target safely, fulfill the task and return to their bases with tolerable losses.

An ordinary air defense system (ADS) consists of Early Warning Radar (EW-R), Surface-to-Air Missile (SAM), and a Command and Control (C2) systems. The components of ADS have different capabilities that define the overall power of defense system as a system of systems (SoS). Therefore, an ADS should be considered as a whole. The main elements of ADS are radars, and communication means also take place in the system. The active countermeasures of ADS are SAM systems, and their guidance are provided by radar means.

Enemy air defense system can be suppressed by using active methods. Active methods ensure permanent damage on air defense systems and they are also called hard kill. In SEAD operations hard kill is applied by using Anti-Radiation Guided

Missile (ARGM) or High-speed Anti-Radiation Missile (HARM). On the other hand, enemy air defense system can be temporarily suppressed using Electronic Countermeasures (ECM) or Electronic Attack (EA) methods. ECM can be applied by jamming enemy assets or deceiving them using decoys. ECM implementation in SEAD operations results in temporary damage or operational interruptions in defense systems. Therefore, the ECM is classified as a soft-kill method.

ECM is defined as the use of electronic technology to jam and deceive enemy radars, especially those used for air defense and interception. ECM can be applied as noise or deceptive jamming. Noise jamming is the use of transmissions to saturate enemy's radar to obscure its target. This causes radar to fail to acquire target. to stop tracking target, or to output false information Apart from this, modern coherent radar systems are no longer easy to jam effectively. For this purpose, deceptive jamming methods are used. Deceptive jamming causes a radar to improperly process its return signal to indicate an incorrect range or angle to the target. Although deceptive jamming is useful for modern radar systems, it is the most important handicap of this method that most of the details about the victim radar need to be known. The most important advantage of noise jamming is that it is sufficient to know only the minimum details about enemy equipment [1]. The ECM methods used in SEAD operations are classified as follows [1-3];

- Escort Jammer (EJ), which acts with task force,
- Stand-Off Jammer (SOJ) that does not enter the lethal border of enemy SAM systems,
- Stand-In-Jammer (SIJ), which installed on an unmanned aerial vehicle (UAV) with very small radar cross-section area, operates in a short distance of a single victim radar.

For the success of jamming operations, the prior knowledge of the enemy air defense structure is essential. This prior information is composed of the location and technical characteristics of the air defense elements in the operational region, and an inseparable part of mission planning and analyzing. A literature study has been conducted and it has been determined that limited studies have been carried out in the literature [6-8]. In this paper, we mainly study the

electronic countermeasure methods in a SEAD operation and evaluate the effects of noise jamming on a radar-based air defense system. The defense system consists Early Warning (EW) radars and SAM systems. We have conducted the analysis of SEAD operation for SOJ and EJ combination.

II. MODELING BASICS OF OPERATION AREA

For modeling the operation area, the enemy defense radars, the task forces and the support jammers are considered. Figure 1 illustrates three components of Support ECM for SEAD operations, which are EJ, SOJ and SIJ. For SOJ applications the jamming unit remains just outside the range of enemy weapons, providing screening for attacking units that actually penetrate enemy defenses. Usually the jamming unit will have only that task, although it may be involved in command and control. The advantage of SOJ is that the jammer is safe from enemy home-on jamming (HOJ) weapons. The disadvantage of this geometry is that burn-through occurs earlier on the attack units because the jammer must remain at very long range, while the attack units close the enemy to very short range.

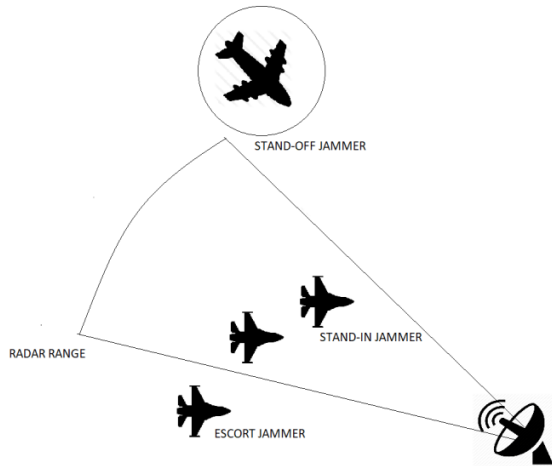


Figure 1: Support ECM Missions for SEAD Operations

For SIJ applications, a remotely piloted vehicle is orbiting very close to the victim radar. The jamming power required by the SOJ to screen a target is much greater than the jamming power required by the SIJ to screen the same target [2]. SIJ is not included in this study. EJ platform moves with the task force and can be considered as self-screening jammer.

A generic ADS has been modeled to analyze the detection range for normal operations and analyzing the reduction amount of detection range due to the effects of ECM. The radar range equation forms the basis of this modeling. Using this equation, we can compare the situation of the operation area before and after the jammer effect. In this section we will discuss the important parameters which constitute the radar range equation. In this study, the analyses are limited with

noise jamming.

First, to analyze the effect of the noise jammer on a radar, consider the power of the signal received by the radar as a single pulse without jamming [1-2].

$$S = \frac{P_T G_T G_R \lambda^2 \sigma}{(4\pi)^3 R_t^4 L_R} \quad (1)$$

In the above equation;

P_T = peak transmitted power, W
 G_T, G_R = transmitter/receiver antenna gain
 σ = RCS, m^2
 λ = wavelength, m
 R_t = target range, m
 L_R = radar system losses

Thermal noise spreads over the entire frequency spectrum uniformly. Therefore, output noise of an ideal receiver is proportional to the absolute temperature of the receiver input system (antenna etc.) which times the bandwidth of the receiver. The factor of proportionality is Boltzmann's Constant

$$N = k \cdot T_0 \cdot B \cdot F \quad (2)$$

Where;

k = Boltzmann's constant : 1.38×10^{-23}
 T_0 = noise temperature - $290^\circ K$
 B_r = noise bandwidth, Hz = $1/\tau$
 F = noise factor
 τ = radar pulse width, sec

The Signal-to-Noise Ratio (SNR) in a receiver is the signal power in the receiver divided by the mean noise power of the receiver. All receivers require the signal to exceed the noise by some amount. Usually if the signal power is less than or just equals the noise power it is not detectable. For a signal to be detected, the signal energy plus the noise energy must exceed some threshold value. S/N is a required minimum ratio, if N is increased, then S must also be increased to maintain that threshold [2]. Increasing transmitted power, increasing antenna gain/aperture area, or decreasing target range will improve the S/N ratio and enhance the probability of target detection. It also shows that decreasing the bandwidth of the radar receiver will increase the S/N ratio and enhance the probability of target detection. Reducing the effective bandwidth of the receiver results in a significant portion of the radar signal spectrum being eliminated and the probability of target detection being reduced.

$$S/N = \frac{P_T G_T G_R \lambda^2 \sigma}{(4\pi)^3 R^4 k T_0 \cdot B_t \cdot F \cdot L_R} \quad (3)$$

The maximum radar range equation [9],

$$R_{max} = \left(\frac{P_T G_T G_R \lambda^2 \sigma}{(4\pi)^3 \left(\frac{S}{N}\right)_{min} k T_0 \cdot B_t \cdot F \cdot L_R} \right)^{1/4} \quad (4)$$

SOJ power density at the radar receiver is given by (Main lobe)

$$P_{SOJ} = \frac{P_j G_j G_{sl} \lambda^2}{(4\pi)^2 R_j^2 L_j^2} \quad (5)$$

SOJ power density at the radar receiver is given by (Side lobe)

$$P_{SOJ} = \frac{P_j G_j G_r \lambda^2}{(4\pi)^2 R_j^2 L_j^2} \quad (6)$$

Where

G_j = jammer antenna gain
 G_{sl} = receiver antenna gain toward jammer
 R_j = range between jammer and radar, m
 L_j = jammer system losses

The EJ power density at the radar receiver is given by

$$P_{EJ} = \frac{P_j G_j G_r \lambda^2}{(4\pi)^2 R_j^2 L_j^2} \quad (7)$$

Self-screening (or escort) jamming (7) can be simplified by accepting $R_j = R_t$. When including the effects of the ECM in the radar equation, the most important quantities are "J-to-S" and Burn- Through Range. "J-to-S" is the ratio of the signal strength of the ECM signal (J) to the signal strength of the target return signal (S). It is expressed as "J/S", and its unit is dB. To be effective, J usually (but not always) exceeds S to an extent. Therefore, the desired result of a J/S calculation in dB will be positive. Burn-through Range is the radar to target range where the target return signal can first be detected through the ECM and is usually slightly farther than crossover range where J=S. It is usually the range where the J/S just equals the minimum effective J/S [2-4].

Main Lobe Jamming to Signal ratio for SOJ is given as,

$$\left(\frac{J}{S}\right)_{soj(ML)} = \frac{P_j G_j (4\pi) R_t^4}{P_T G_T \sigma \left[\frac{B_j}{B_R}\right] R_j^2} \quad (8)$$

Side Lobe Jamming to Signal ratio for SOJ,

$$\left(\frac{J}{S}\right)_{soj(SL)} = \frac{P_j G_j G_{R(SL)} (4\pi) R_t^4}{P_T G_T G_{R(ML)} \sigma \left[\frac{B_j}{B_R}\right] R_j^2} \quad (9)$$

Also Jamming to Signal ratio for EJ is defined as follows,

$$\left(\frac{J}{S}\right)_{esc} = \frac{P_j G_j (4\pi) R_t^2}{P_T G_T \sigma \left[\frac{B_j}{B_R}\right]} \quad (10)$$

Finally, the burn-through range equation is defined as follows,

$$R_{bt} = \left(\frac{P_T G_T G_R \lambda^2 \sigma}{(4\pi)^3 \left(\frac{S}{N}\right)_{min} \left(\frac{J}{S}\right) k T_0 \cdot B_t \cdot F \cdot L_R} \right)^{1/4} \quad (11)$$

III. SIMULATION STUDY

In this section the results of conducted simulation study is presented. For simulation study an air defense system has been modeled in MATLAB using the formulas described in the previous section and the parameters are presented in Table 1 and 2. In Table (1), radar parameters of enemy ADS are given. In Table (2) parameters of SOJ and EJ are given.

Early Warning Radar	Fire Control Radar
$P_T = 600$ MW,	$P_T = 300$ MW,
$G_{T,R} = 24$ dB	$G_{T,R} = 18$ dB
$\sigma = 1$ m ²	$\sigma = 1$ m ²
$f = 1.3$ GHz	$f = 4$ GHz (fire control) $f = 8$ GHz (track)
$L_R = 1$ dB	$L_R = 1$ dB
$G_{sl} = 5$ dB	$G_{sl} = 3$ dB
$B_T = 1$ MHz	$B_T = 1$ MHz

Table 1: Radar Parameters

Stand-Off Jammer	Escort Jammer
$P_j = 50$ kW	$P_j = 10$ kW
$G_j = 20$ dB	$G_j = 25$ dB
$B_j = 100$ GHz	$B_j = 100$ GHz

Table 2: Jammer Parameters

In Figure 2, the signal strengths of the task force aircraft, which is a target for enemy ADS radar systems, and the SOJ, which applies noise jamming into ADS radar receiver, are calculated together. For calculations, equations (1) and (5) are used. As seen in this figure, the signal from the SOJ suppresses the signal received from the target aircraft, which means a protection is provided for task force aircrafts.

In Figure 3 the effect of joint operation of EJ and SOJ is given. The signal strength from the task force aircraft to the radar receiver and the signal strength from the joint operation of EJ and SOJ are calculated together. Unlike Fig. 2, the signal strength from the jammer is not shown as a straight line but as a curved line, due to the effect of EJ. The EJ moves side by side with the mission plane and its distance from the radar is the same as the mission plane. Figure 3 shows how the signal from the target is suppressed by signals from both the SOJ and the EJ using a signal from the target using equations (6) and (7). It can be seen from the results that, the impact of SEAD ECM suppression is enhanced by superimposing the effects of EJ and SOJ. This suppression prevents the target aircraft from being detected by the radar until the task is completed.

RADARS	SOJ – Ranges (Main Lobe) Location (850,2150)	SOJ – Ranges (Side Lobe) Location (1250,2150)	Escort Jammer/AC Ranges Location (1184,1868)
FC1	570km	917km	168km
FC2	850km	335km	432km
FC3	1570km	1250km	1134km
FC4	1011km	901km	741km
EW1	728km	626km	438km
EW2	538km	509km	185km
EW3	1320km	528km	923km
EW4	1320km	1208km	895km

Table 3: Ranges Between Jammers/ACs and Radars

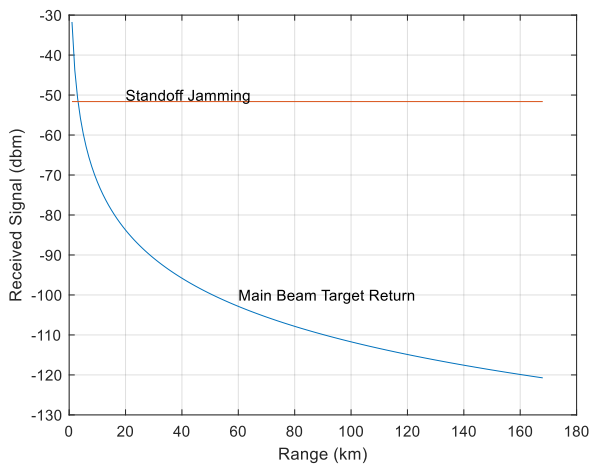


Figure 2: Combined SOJ and Target Detection Effect in Radar Receiver

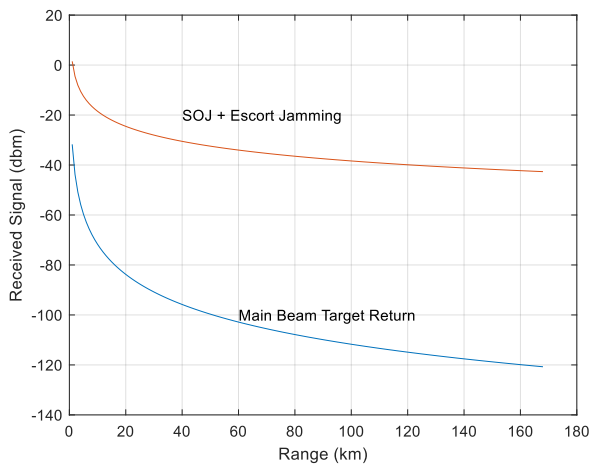


Figure 3: Combined EJ+SOJ and Target Detection Effect in Radar Receiver

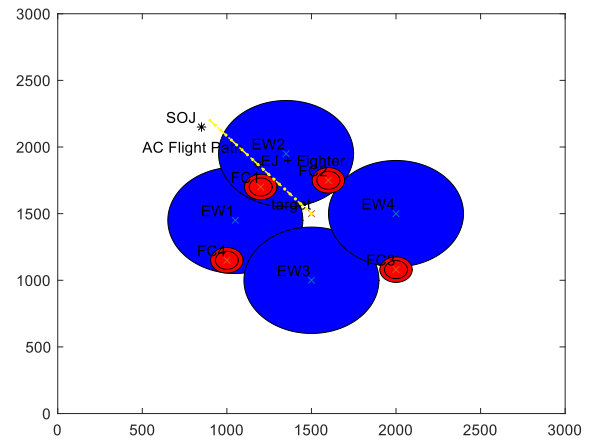


Figure 5: Early Warning and Fire Control Radar Zones (Before Jamming)

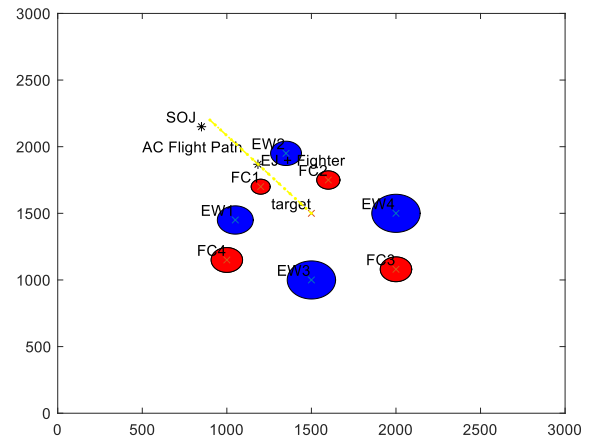


Figure 6 : The radar ranges after the effect of SOJ on the main lobe on FC1/FC2 and all of the EW radars.

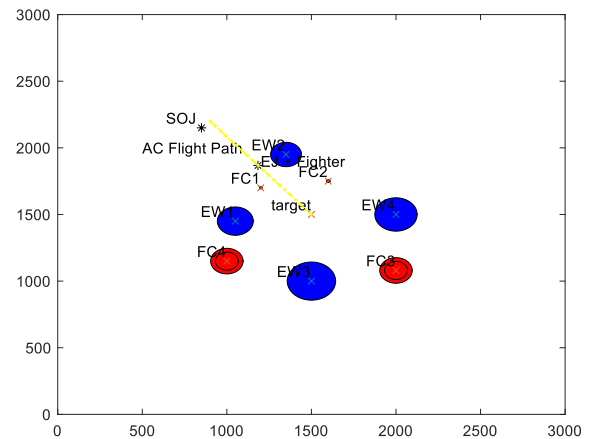


Figure 7: Joint operation of SOJ and EJ effects on FC1 and FC2 (main lobe) / SOJ effects on EWs

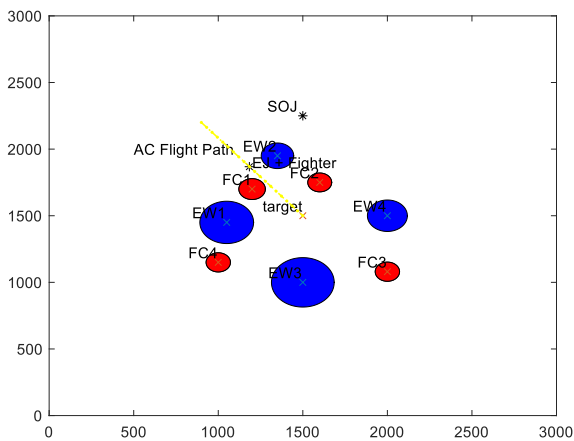


Figure 8 : SOJ effects on FC Radars (side lobe) / SOJ effect on EW Radars

In Figure 5, the previous positions of the radar field without jamming and the positions of EJ and SOJ are given. And the following other figures are modeled on the jammer effects that will occur on this radar zone with different scenarios. The effectiveness of the jammer depends upon its distance from the radar. In figure 8 The effect of SOJ on side lobe was observed on FC radar. And when we look at the new radar range, we observe that the side lobe effect is quite low compared to the main lobe.

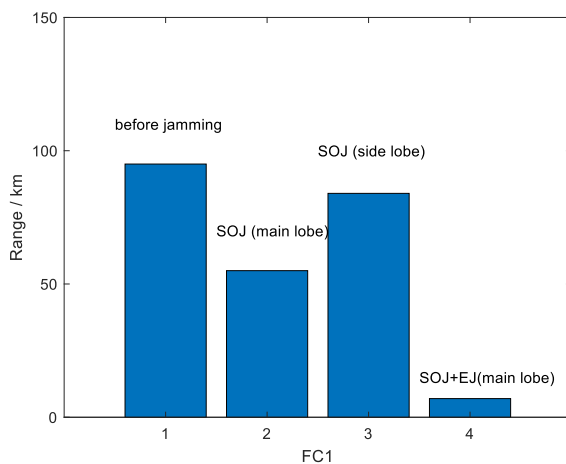


Figure 9 : FC1 Ranges

As seen in the Figure 9, the effect of SOJ on the radar range from FC was 42% from the main lobe and its effect was reduced to 11% from the side lobe. In addition, the effect of the joint operation of SOJ and EJ was calculated as 92%.

IV. CONCLUSION

In this study, we have analyzed how air operations will work with different scenarios using SOJ and Escort Jammer in an air defense zone consisting of Early Warning and Fire Control radars. The results show that it is better to get support from an SOJ aircraft than to be the only an EJ aircraft accompanying a

pack of aircraft in an air assault operation. Also, the effects of SOJ jamming on the side lobes and main lobes of fire control radars have been analyzed. At the end of this analysis, the amount of the superiority of the main lobe jamming over sidelobe jamming has been evaluated.

REFERENCES

- [1] D. CurrisScheler, *Electronic Warfare in the Information Age*. Boston: London, 1999.
- [2] Scott O'Neill, *Electronic Warfare and Radar System Engineering Handbook*, 2013
- [3] David Adamy, *EW 101A First Course in Electronic Warfare*, Boston: London, 2001.
- [4] Bassem R. Mahafza, *Radar System Analysis and Design Using Matlab*, Third Edition, Hunstville, Alabama, USA
- [5] Thomas Withingtoni Wild Weasel Fighter Attack, South Yorkshire, 2008
- [6] Wang Jundong, *Complex Environment Noise Barrage Jamming Effects On Airborne Warning Radar*, China, 2018
- [7] KeerthanaK, Shanmugha Sundaram G A, "Effect of jammer signal on stepped frequency PAM4 radar waveforms", ICACC-2018
- [8] Sulaiman H. M. Al Sadoon ,Badal H. Elias, *Radar theoretical study: minimum detection range and maximum signal to noise ratio (SNR) equation by using MATLAB simulation program*, Duhok, Iraq, 2013
- [9] Skolnik, M. I., *Introduction to Radar Systems*, 2nd Ed., New York: McGraw-Hill, 1980.

Performance Evaluations of Meta-Heuristic Algorithms On Solving Group Elevator Control Problem

A. USTA¹, U. ATILA¹, K. AKYOL² and Y.Y. BAYDILLI¹

¹ Karabuk University, Karabuk/Turkey, ayseglusta@gmail.com

¹Karabuk University, Karabuk/Turkey, umitatila@karabuk.edu.tr

²Kastamonu University, Kastamonu/Turkey, kakyol@kastamonu.edu.tr

¹Karabuk University, Karabuk/Turkey, yusufbaydilli@karabuk.edu.tr

Abstract - Elevator is a factor that makes people life easier for years. Both mechanical and software experiments are made on elevators to achieve efficient systems. For this purpose, group elevator control algorithms were developed in order to transfer people efficiently. With these algorithms, it is aimed to select suitable cabins to respond to a floor call. In this study, the group elevator control problem was solved by using meta-heuristic optimization algorithms and average waiting time for the people was minimized. A scenario with 10 different floor calls in a building with 16 floors and 7 elevators was considered. For the sake of comparison, the algorithms were run with 10000, 2000 and 30000 maximum fitness function evaluation values. All the algorithms except GSA performed the most ideal elevator and floor call matching and responded to 10 floor calls in 22.4 seconds. According to results, ABC algorithm was the most robust algorithm with the lowest standard deviation value in all cases.

Keywords - Group elevator control, optimization, meta-heuristic algorithms, discrete optimization

I. INTRODUCTION

The elevator uses to people transport in multi-floors buildings. Elevators are used in many service areas. The number of elevator can vary according to number of floors and population. In buildings with high number of elevators, the elevators can move independently or they can be controlled as a group. Developments in the field of control and computer sciences have enabled the enhancement of the elevator group control systems. Elevator group control systems regulate operations according to the floor calls. Whenever a new call arises, they examine the current status of cabins and select the appropriate elevator and direct it to the corresponding floor where the call is located. The calls should not be left answered and an instant decision must be made for the new call.

The main objective is to transport people to the destination floor as quickly as possible. The main issue here is to minimize the waiting time of people. This can be accomplished by a well-planned timing provided by

group control evaluator systems. In literature, different optimization algorithms such as fuzzy logic, genetic algorithms, artificial neural networks, particle swarm optimization and artificial atom algorithm were used to achieve optimal performance from these systems [1-6].

In this study, a group elevator control problem with 10 calls in a 16-storey building with 7 elevators was solved by using PSO, ABC, GSA, GWO, WOA algorithms and performances of these algorithms were evaluated.

In Section 3 of this study group elevator control problem was introduced. In Section 4, Meta-Heuristic Optimization Algorithms used in this study were presented. Experimental results were given and the performances of the algorithms were compared in Section 5. Finally, we conclude the study with evaluations and discussions in Section 6.

II. GROUP ELEVATOR CONTROL PROBLEM (GECp)

Group elevator control systems are such systems that control the cabins to efficiently transport people waiting on the floors in high-floors buildings having three or more elevators. The purpose of these systems is to assign the appropriate cabin for the calls generated by people and reduce waiting and transport time [7].

An elevator system has two different call types. First one is the floor call made from floors stating that a person wants up or down. Second one is the cabin call made by a person by pressing the floor number in the cabin. Group elevator control system operates according to floor calls. Each time a new call arrives, it examines current status of the cabin and selects the appropriate elevator and directs it to floor where the call is located. In this study, floor calls were taken into consideration and existence of 10 floor calls were assumed. Starting positions of elevators and floor calls made by people are given in Figure 1.

The number of floors and elevators in the building are the physical characteristics of the buildings. In the considered group elevator problem, the physical properties of the evaluators and the building were

considered to be constant since they do not affect the success of optimization algorithms.

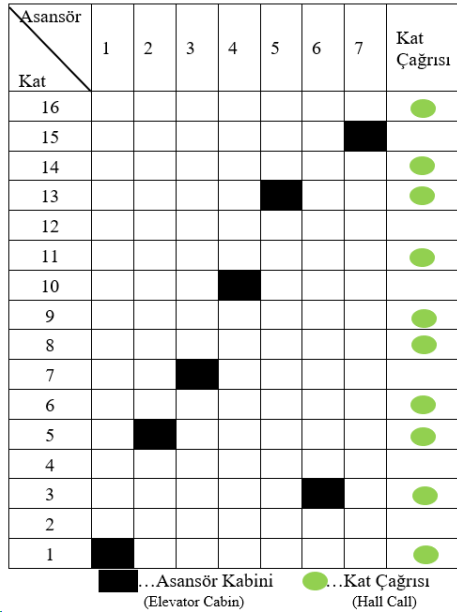


Figure 1: Starting positions of elevator cabins and floor calls

For this problem, the transition time between floors and opening/closing time of elevator door were determined as physical properties of the elevator. These properties were demonstrated in Table 1.

Table 1: Physical properties of buildings and elevators

Property	Value
Floor number of the building	16
Number of elevators in the building	7
Average opening/closing time of elevator door (sec)	4
Average transition time between floors (sec)	3

The same objective function is used for each algorithm in this study. The calculation of objective function value is given in Equation 1.

$$M[B(j)] = [|k_p(i) - k_h(i)|] * s_g * a_1 + s_{ak} * a_2 \quad (1)$$

Objective function:

$$\min \sum_{i=1}^m (M[B(j)] = [|k_p(i) - k_h(i)|] * s_g * a_1 + s_{ak} * a_2) \quad (2)$$

$$i = \{1, 2, \dots, m\}$$

$$j = \{1, 2, \dots, n\}$$

k_p : Cabin position

k_h : Destination floor of cabin

s_g : Average transition time of cabin between floors (sec)

s_{ak} : Average opening/closing time of elevator door

a_1 : Weight coefficient-1 (0.8)

a_2 : Weight coefficient-2 (0.2)

An evaluator can respond to multiple floor calls. In group elevator control problem, the value of objective function is the sum of the time taken by all elevators to answer floor calls. Therefore, if an elevator is answering more than one floor call, the starting position of the elevator needs to be updated at every step in order to calculate the value of objective function correctly.

III. META-HEURISTIC OPTIMIZATION ALGORITHMS

Most of the optimization problems are difficult to solve with traditional optimization methods since they contain specific constraints. High number of parameters in the optimization problems and the non-linear characteristics of the effect of these parameters on the objective function increase the problem complexity and many local optimum points occurs in the objective function. For this reason, the researchers who want to achieve the global optimum in optimization problems prefer to use evolutionary optimization methods which offer more effective and efficient solutions instead of classical methods. Evolutionary algorithms are population based and nature-inspired meta-heuristic algorithms and can be based on swarm intelligence, biological systems, physics principal of chemistry depending on the source of inspiration [8]. In this study, group elevator control problem was solved by using Genetic Algorithm, Particle Swarm Optimization, Artificial Bee Colony, Whale Optimization Algorithm, Gravitational Search Algorithm and Grey Wolf Optimization which are among the popular evolutionary algorithms and experimental results were compared.

A. Particle Swarm Optimization (PSO)

PSO is a population based evolutionary optimization method based on swarm intelligence developed by Kennedy and Eberhart in 1995 [9]. The method is modeled based on the collective movements of bird and fish flocks during food search and escape from dangers.

The algorithm is superior to many other optimization algorithms since it finds results rapidly without sticking into local optimums using less number of parameters.

In the algorithm, an individual that represents solution to the problem is called particle. Each particle has two vectors, velocity and position. In each generation, the particles update their own speeds and positions according to the best particle. When particles come closer to the best particle, they can achieve more quality solutions than the best particle of the previous generation with the help of random operators. This rapid improvement between generations allows the PSO converge faster [10].

B. Artificial Bee Colony (ABC)

ABC algorithm was developed based on the behaviors of bees when they search for food sources [11]. In this process, work sharing of bees among themselves, having different roles according to changing conditions and self-

organization of swarm are some of these behaviors [12].

There are three kind of bees which are worker bees, explorer bees and discoverer bees in ABC algorithm. Worker bees are responsible from food resource and the number of the worker bee equals to the number of observer bees. When food resources are exhausted, worker bees turn into explorer bees and begin to look for new resources. The worker bees investigate their food sources and share information about food sources with observer bees. In the algorithm, the location of food source corresponds to a solution in the problem. The amount of nectar in food indicates objective function value of the solution [13].

C. Whale Optimization Algorithm (WOA)

WOA was inspired by the Bubble hunting strategy of humpback whales [14]. Humpback whales feeding with small fish swarms create breath bubbles underwater. So, they gather their preys together. They move towards to surface of the water by bubbling in these bubbles. In this way, they provide their prey to remain in the bubbles and hide themselves [15].

Gravitational Search Algorithm (GSA): GSA was developed by Rashedi et al. inspired from Newton's laws of motion and gravity. In the algorithm, each individual in the search space is considered to be a mass. All masses attract one another and force each other with the gravity force. The masses exposed to gravitational force try to reach the optimum result by moving in the search space. In this algorithm, the mass amount of each individual indicates the performance of that individual. When the algorithm is executed, the heaviest mass will move slowly and apply a gravitational force to attract others [16].

D. Gray Wolf Optimization (GWO)

GWO has been developed by imitating hunting and swarm behaviors of grey wolves in nature [17]. In the social hierarchy, grey wolves are classified as alpha, beta, delta and omega. Alpha wolves represent the most dominant wolves. Beta wolves are the helper of alpha wolves and provide communication between alpha wolves and other wolves. Omega wolves are chosen by the alpha wolves and take part at the lowest level in the hierarchy. These wolves feed at last during the hunting. Delta wolves, that are not included in the alpha, beta and omega wolves, dominate omega wolves and chosen by alpha and beta wolves [18]. In GWO algorithm, candidate solutions are formed by taking into account the social hierarchy of wolves. Solutions with optimal objective values are considered as alpha, beta and delta wolves, respectively. The main parts of the hunting consist of the stages of searching for prey, encircling prey, and attacking prey [19].

having 7 elevators and the performances of these algorithms were evaluated. Experiments were carried out on a computer with Intel(R) Core(TM) i3-4010U processor, 1.70 GHz speed and 8 GB RAM. We used MATLAB R2016a on 64 bits Windows operating system for coding algorithms. Algorithms were run under following conditions:

- Maximum number of function evaluations (MaxFEVs) were fixed as 10000, 20000 and 30000 for all algorithms.
- Population size were selected as 40.

Dimension of the problem is 10 as we have 10 floor calls in the building and in the initial population each dimension was assigned with values between 1 and 7 as we have 7 elevators. Internal parameters of algorithms and their values were assigned as given in Table 2. GWO and WOA does not have any internal parameter.

Table 2: Internal parameters for algorithms

	Parameter
ABC	limit =400 Colony size = 40x2 = 80 Food number: 40
PSO	wMax = 0.9 wMin = 0.2 c1 = 2, c2 = 2; vMax = 1.2, vMin = -1.2
GSA	Gravity constant $G_0 = 100$ $\alpha = 20$ $\varepsilon = 10^{-12}$

Convergence curves demonstrating mean of best objective function values of 30 runs for algorithms with MaxFEVs values fixed to 10000, 20000 and 30000 were given in Figure 2-4. As seen in these figures, ABC algorithm performed better than other algorithms while increasing the value of MaxFEVs. Besides, the worst performing algorithm was the GSA.

Performances of the algorithms were evaluated considering statistical results obtained from 30 run of algorithms and given in Table 3. Considering cases with 10000, 20000 and 30000 MaxFEVs values as case-1, case-2 and case-3 respectively, all algorithms found best solution by 22.40 except GSA and GWO for case-1 and except GSA for case-2 and case-3.

On the other hand, ABC algorithm had lowest mean objective and standard deviation values for all cases. Likewise, ABC algorithm found best solution 26, 29 and 30 times out of 30 runs showing that ABC was the most robust algorithm for this problem. Looking at the convergence speed, PSO was the fastest by 687, 1109 and 485 function evaluations respectively for all cases.

IV. EXPERIMENTAL STUDIES

In this study GA, PSO, ABC, GSA, GWO, WOA algorithms were used in order to solve the group elevator control problem with 10 calls in a 16-storey building

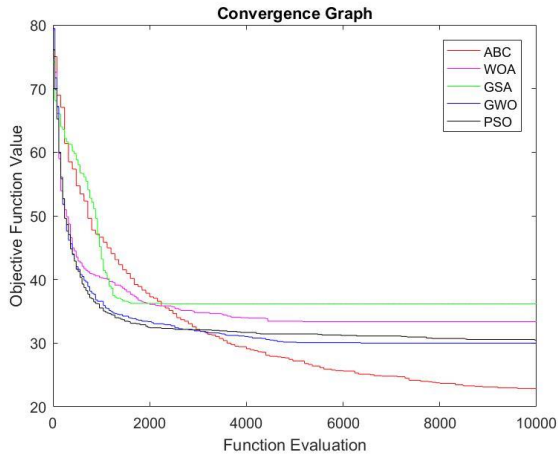


Figure 2: Mean objective function values of algorithms
(Run=30, MaxFEVs =10000)

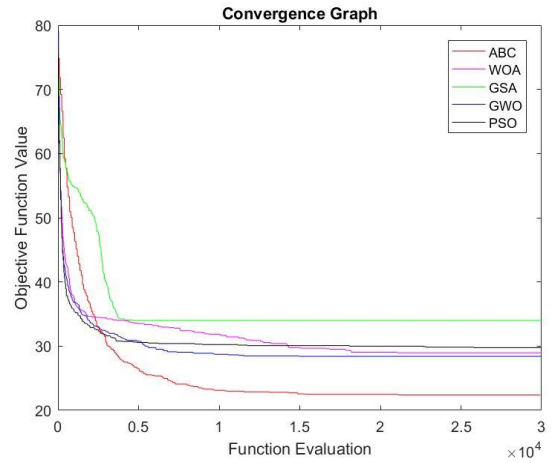


Figure 4: Mean objective function values of algorithms
(Run=30, MaxFEVs =30000)

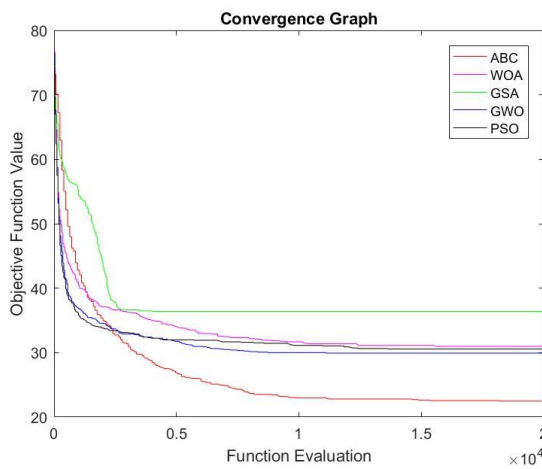


Figure 3: Mean objective function values of algorithms
(Run=30, MaxFEVs =20000)

Table 3: Performance results obtained by GA, PSO, ABC, GSA, GWO and WOA algorithms as a result of 30 independent runs for GECP problem.

MaxFEVs	Algorithm	Best Objective Value	Mean Objective Value	Worst Objective Value	Std. Deviation	Best Function Evaluation	Number of Best Result
10.000	PSO	22.4	30.48	44	5.33	687	4
	ABC	22.4	22.88	27.2	1.32	1240	26
	GSA	32	36.16	41.6	2.95	960	--
	GWO	27.2	30	44	5.46	3585	--
	WOA	22.4	33.36	44	5.77	1489	1
20.000	PSO	22.4	30.56	44	5.83	1109	5
	ABC	22.4	22.48	24.8	0.44	2360	29
	GSA	29.6	36.4	48.8	4.17	4040	--
	GWO	22.4	29.92	44	5.69	11139	1
	WOA	22.4	30.96	44	6.51	4908	3
30.000	PSO	22.4	29.76	46.4	6.01	485	6
	ABC	22.4	22.4	22.4	1.08E-14	2920	30
	GSA	29.6	34	44	3.68	2640	--
	GWO	22.4	28.4	41.6	3.32	12875	4
	WOA	22.4	28.96	44	5.67	3377	7

Table 4: Hall calls and response sequences answered by elevators based on best results in 30000 maxFEVs)

Elevator	PSO	ABC	GSA	GWO	WOA
1	Floor 1	Floor 1	Floor 1	Floor 1	Floor 1
2	Floor 3-4	Floor 3-4	Floor 3-4	Floor 3-4	Floor 3-4
3	Floor 5-6	Floor 5-6	Floor 5-6	Floor 5-6	Floor 5-6
4	Floor 7	Floor 7	Floor 7-10	Floor 7	Floor 7
5	Floor 8-9	Floor 8-9	--	Floor 8-9	Floor 8-9
6	Floor 2	Floor 2	Floor 2	Floor 2	Floor 2
7	Floor 10	Floor 10	--	Floor 10	Floor 10

According to the results obtained from 30 run with 30000 MaxFEVs, all algorithms except GSA found the best solution considering the scenario given in Figure 1. Elevators-call matchings and paths to be followed by elevators in this solution were given in Table 4. According to best solution, floor calls and responding elevators were as follows:

- Elevator 1 responds to call from floor 1.
- Elevator 2 responds to call from floor 3 and 4.
- Elevator 3 responds to call from floor 5 and 6.
- Elevator 4 responds to call from floor 7.
- Elevator 5 responds to call from floor 8 and 9.
- Elevator 6 responds to call from floor 2.
- Elevator 7 responds to call from floor 10.

I. CONCLUSION

Elevators are used for transporting people in multi-floor buildings. Considering today's busy work life, people do not tolerate waiting for elevator. Therefore, various improvements are made both in mechanics and software, and the average waiting time is tried to be minimized.

In this study GECP which is a discrete optimization problem was solved using meta-heuristic algorithms including GA, PSO, ABC, GWO and GSA. All algorithms except GSA found optimal solution for all cases. According to results, ABC algorithm was the most robust algorithm with the lowest standard deviation value in all cases. Looking at the convergence speed, PSO had the lowest value. However, considering the standard deviation value of PSO, the robustness of this algorithm was worse than ABC. All the algorithms except GSA performed the most ideal elevator and floor call matching and responded to 10 floor calls in 22.4 seconds.

As a conclusion, considering the performances of all algorithms it is recommended to use ABC algorithm to

achieve accurate and stable results in the design of group control elevator system.

REFERENCES

- [1] Y. Umeda, K. Ueton, H. Ujihara, and S. Tsuji, "Fuzzy theory and intelligent options," *Elevator World*, vol. 37, pp. 86-91, 1989.
- [2] R. Gudwin and F. Gomide, "Genetic algorithms and discrete event systems: An application," *IEEE World Congress on Computational Intelligence*, pp. 742-745, June 1994.
- [3] C. E. İmrak and G. C. Barney, "Application of neural networks on traffic control," *ELEVCON'98, Elevator Technology*, vol. 9, pp. 140-148, 1998.
- [4] F. Luo, Z. Xiaocui, and Y. Xu, "A new hybrid elevator group control scheduling strategy based on particle swarm simulated annealing optimization algorithm," *Proceedings of the 8th World Congress on Intelligent Control and Automation*, pp. 5121-5124, July 2010.
- [5] B. Bolat, O. Altun, and P. Cortés, "A particle swarm optimization algorithm for optimal car-call allocation in elevator group control systems," *Applied Soft Computing*, vol. 13, pp. 2633-2642, 2013.
- [6] A. Yıldırım Erdoğan, "Yapay atom algoritması ve ayrık problemlere uygulanması," Doktora Tezi, İnönü Üniversitesi, Fen Bilimleri Enstitüsü, 2018.
- [7] H. Lee Kwang and C. B. Kim, "Techniques and applications of fuzzy theory to an elevator group control system," *Fuzzy Theory Systems: Techniques Applications*, vol. 1, pp. 461-481, 1999.
- [8] I. Fister jr, X. S. Yang, I. Fister, J. Brest, and D. Fister, "A Brief Review of Nature-Inspired Algorithms for Optimization," *Elektrotehniski Vestnik/Electrotechnical Review*, vol. 80, pp. 1-7, 2013.
- [9] J. Kennedy, and R. C. Eberhart, "Particle swarm optimization," *Proceeding of IEEE International Conference on Neural Network*, Perth, pp. 1942-1948, 1995.
- [10] Y. Shi, and R. Eberhart, "Empirical study of particle swarm optimization," *Proc. of the Congress on Evolutionary Computation*, Washington, vol. 3, pp. 1945-1950, 1999.
- [11] D. Karaboğa, "An idea based on honey bee swarm for numerical optimization," Technical Report-TR06, Erciyes University Engineering Faculty Computer Engineering Department, 2005.
- [12] C. Öztürk, "Yapay sinir ağlarının yapay arı koloni algoritması ile eğitilmesi," Doktora Tezi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü, 2011.
- [13] B. Akay, "Nümerik optimizasyon problemlerinde yapay arı kolonisi algoritmasının performans analizi," Doktora Tezi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü, 2009.
- [14] S. Mirjalili and S. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51-67, 2016.
- [15] J. A. Goldbogen, A. S. Friedlaender, F. Calambokidis, M. F. Mckenna, M. Simon, and D. P. Nowacek, "Integrative approaches to the study of Baleen whale diving behavior, feeding performance and foraging ecology," *BioScience*, vol. 63, pp. 90-100, 2013.
- [16] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, "GSA: A new gravitational search algorithm," *Information Science*, vol. 179, pp. 2232-2248, 2009.
- [17] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46-61, 2014.
- [18] A. Tunç, "Finans sektörü için yapay öğrenme teknikleri kullanarak kredi kullanılabilirliğin tespiti," Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, 2016.
- [19] T. Koç, O. K. Baykan, and İ. Babaoğlu, "Gri kurt optimizasyon algoritmasına dayanan çok seviyeli imge eşik seçimi," *Politeknik Dergisi*, vol. 21, pp. 841-847, 2018.

Performance Comparison between Genetic Algorithm and Ant Colony Optimization on Traveling Salesman Problem

A. DUMLUPINAR¹, F. CELIK¹, T. USLU¹, D. EKMEKCI¹

¹ Karabuk University, Karabuk/Turkey, ahmet.taha.dumlupinar@gmail.com

¹ Karabuk University, Karabuk/Turkey, farukcelik.661@gmail.com

¹ Karabuk University, Karabuk/Turkey, tuncayuslu1997@gmail.com

¹ Karabuk University, Karabuk/Turkey, dekmekeci@karabuk.edu.tr

Abstract - Traveling Salesman Problem since the day it was created, it is one of the most important optimization problems in different areas. Within the scope of the study, solutions for TSP have been developed by using Genetic Algorithm (GA) and Ant Colony Optimization (ACO) metaheuristic algorithms. Proposed method were tested on test problems which has 51 nodes with parameter values close to each other for algorithm control parameters. When the results obtained from ACO and GA were compared, it was observed that 25% of iterations required for GA was enough for ACO. Also, it was clearly observed that ACO achieved more successful results in each test.

Keywords – Genetic Algorithm, Ant Colony Optimization, Traveling Salesman Problem

I. INTRODUCTION

Traveling Salesman Problem (TSP) emerged as a mathematical expression by Karl Menger at the beginning of the years 1930 [1]. Nowadays, this problem has many application areas [2]. GSM Operators location determination for base stations, flight routes for airplanes, vehicle routing problems and electronic circuit design, examples of application areas of the traveler salesman problem.

The goal of the TSP is to minimize the cost of the sales person visiting each city and returning to the starting city. TSP is one of the hardest problems to solve in optimization problems [3]. Because it is necessary to review all the cities in order to find the path that has the minimum cost. To summarize, there are $n!$ solutions for network consisting of n cities. The problem is due to the excessive time of the solution in linear methods, for the solution, the meta heuristics algorithms were developed inspired by the lifestyles of living creatures and the ability to find foods in nature. The Genetic Algorithm (GA) and Ant Colony Optimization (ACO) methods, which are Developed from these algorithms, are the successful algorithms that approach the optimum outcome [4][5].

In this study, we have developed solutions for TSP, using GA and ACO methods. To assess the performance of the

algorithms, we have tested the applications, Dethloff benchmark problems, set close values to each other for control parameters. (Dethloff problems developed for vehicle routing problems. In our study, we transformed the problems into a TSP model by ignoring the delivery and pickup requests on the nodes.) The results show that ACO can find more successful solutions than GA.

II. GENETIC ALGORITHM

The fundamentals of the GA were first introduced by John Holland [6]. Holland gathered his study in this direction, in 1975, in the book named 'Adaptation in Natural Artificial Systems'.

In Genetic algorithms, usually the initial population is randomly generated based on certain criteria. After the initial population is created, the crossover process is performed. There are multiple crossover methods. These are one point, two point and uniform crossover methods [7] as shown in Figure 1.

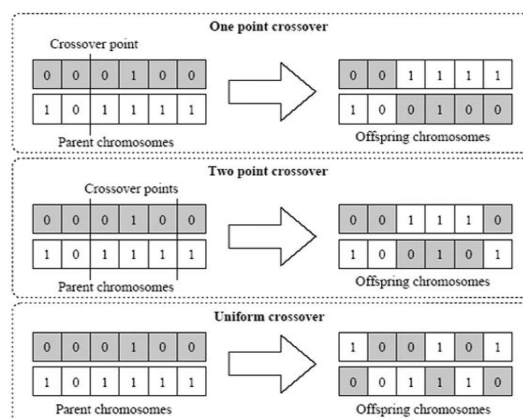


Figure 1: Crossover examples of GA [7]

The mutation process is performed on the new chromosomes after the crossover process. The reason for the mutation process is to be able to move to different areas of the search space. There are several methods that are also used in the mutation process. These methods are referred to as insertion, reversal, placement, and mutual change method as shown in Figure 2.

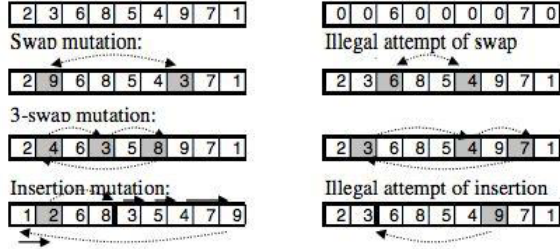


Figure 2: Mutation examples of GA

General flow chart of GA is presented in Figure 3.

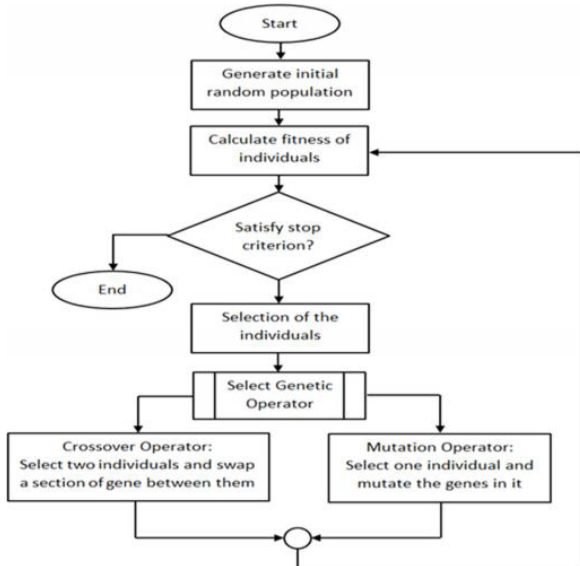


Figure 3: Flow chart of GA. [8]

III. ANT COLONY OPTIMIZATION

The other algorithm used to resolve the TSP is the ant colony optimization. ACO was introduced as an ant system by Marco Dorigo in 1991 [9][10]. The algorithm is mainly based on the foraging behavior of ants. In 1999, it was described as a meta-heuristics ant colony optimization by Dorigo, Di Caro and Gambardella. [11]. Real ants have the ability to find the shortest route to the food source without any visual clues [12] [13]. Real ants use pheromone to find the shortest path. Ants store a certain amount of pheromone when starting their journey. Randomly disperse into their environment and begin the process of food search. Ants reaching food sources produce pheromone as they return to their nest in the food source.

There are two alternatives to choose the point s for the k -ant at point r during a round in the ACO [5]. The first alternative is the selection of the path where the amount of pheromone is high. Necessary formulation is given in (1).

$$s = \begin{cases} \arg \max_{u \in M_k} \{ [\tau(r, u)] \cdot [\eta(r, u)]^\beta \} & \text{if } q \leq q_0 \\ S & \text{otherwise} \end{cases} \quad (1)$$

In equation (1) $\tau(r, u)$ is the amount of the pheromone value between r and u . $\delta(r, u)$ is a distance between r and u and $\eta(r, u)$ is inverse of the distance. α and β are two parameters that control the relative weight of pheromone trail and heuristic value [5].

The second alternative is the choice made according to the probability distribution created by pheromone quantities given (2) [11].

$$p_k(r, s) = \begin{cases} \frac{[\tau(r, u)] \cdot [\eta(r, u)]^\beta}{\sum_{u \in M_k} [\tau(r, u)] \cdot [\eta(r, u)]^\beta} & \text{if } s \notin M_k \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The ants have completed their tours, the pheromones they secrete are subjected to evaporation and pheromone updates are made. After the ants have completed their tours, the pheromones they secrete are subjected to evaporation and pheromone updates are made (3) [5].

$$\tau(r, s) \leftarrow (1 - \alpha) \cdot \tau(r, s) + \alpha \cdot \tau_0 \quad (3)$$

In the global pheromone update, the amount of pheromone in the path followed by the least-cost ant is increased (4) [5].

$$\tau(r, s) \leftarrow (1 - \alpha) \cdot \tau(r, s) + \alpha \cdot \Delta\tau(r, s) \quad (4)$$



Figure 4: The path between food source and nest [5]

At this point, the pheromone, which is secreting the ant that returns to the nest in a shorter way, is exposed to less evaporation. The ants have found the shortest path because they choose the path with more pheromone. (Figure 4)

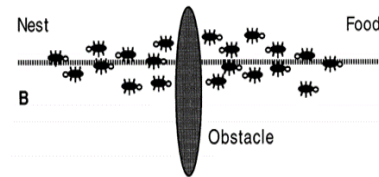


Figure 5: The obstacle between the Ants food supply[5]

When the shortest possible path is deterioration due to environmental factors, it adapts to this situation and begins to seek a new path. (Figure 5)

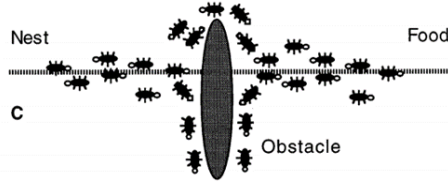


Figure 6: The choice of new ways of Ants[5]

If any obstacle is encountered on the road, the ants must make a choice and the probability of these elections is equal (Figure 6).

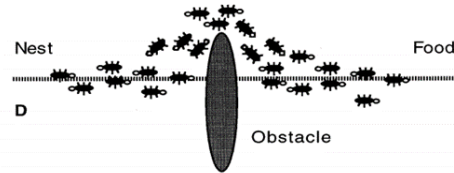


Figure 7: The new way the Ants create [5]

As a result of the election, the pheromones of the ants who prefer the shortest path will be more, the other ants begin to use this path (Figure 7) [14].

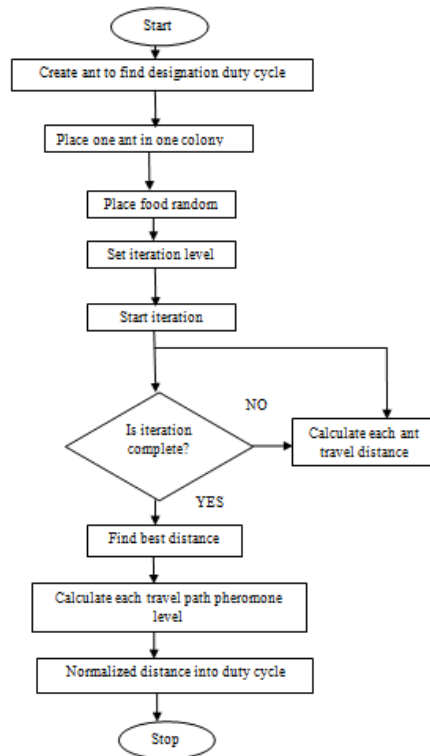


Figure 8: Flow chart of ACO [15]

IV. COMPUTATIONAL STUDY

Developed solution methods with GA and ACO were tested on Dethloff benchmark problems. When Dethloff test problems are examined, it is observed that these problems are divided into four groups in the names CON3, CON8, SCA3 and SCA8. Because Dethloff problems contain different values in terms of distances between nodes and vehicle capacity, there are different sets of solutions for each problem. Each test problem consists of a starting point and 50 different visiting nodes. The starting and ending point for solving Dethloff problems is the node that is designated as the center. The points to navigate were created using two different methods. In SCA test problems, all the nodes are randomly distributed in an area of [100x100]. In the CON test problems, half of the nodes are distributed randomly in [100x100] and the other half is distributed in $[(100/3) \times (200/3)]$ [16].

For Dethloff problems, the solution spaces will be shaped by the distance between the nodes and the load ratio parameters, and in cases where the standard deviation values for these parameters are low, the solutions will occur at closer levels. CON3.1 and SCA3.1 problems are the examples where the standard deviations of distances between nodes and delivery/pickup amount values are low. Therefore, the CON3.1 and SCA3.1 problems can be considered as the most difficult Dethloff examples [17]. Differences between the results can be easily seen by changing the parameters in the experiment and differences between the GA and the ACO become more obvious.

In the study; for crossover process, double point order crossover (OX) method and for mutation, swap method was used. By selecting the starting population 51 in the experiments, equal initial conditions are provided between the algorithms.

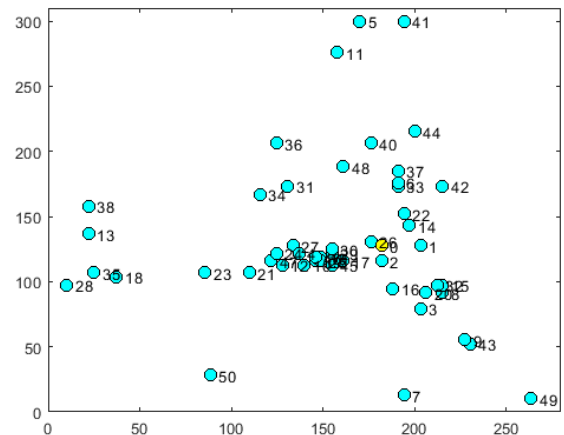


Figure 9: The locations of 51 nodes in CON 3.1 problem

Table 1: Parameter setting for GA

Test No	Start Pop.	Control Parameters	
		Crossover Rate (%)	Mut. Rate (%)
1	51	70	5
2	51	70	10
3	51	70	15
4	51	80	5
5	51	80	10
6	51	80	15
7	51	90	5
8	51	90	10
9	51	90	15
10	51	100	5
11	51	100	10
12	51	100	15

In the experiments, crossing and mutation rates were changed on GA and 30 repeats were made for each different parameter. Genetic algorithm was coded on the Matlab. In the experiments, the crossing rate was started from %70 and continued for up to %100 with a %10 increase, mutation rate, starting at %5 and increasing by %5, continued up to %15. The system with a starting population of 51 was run at 10000 iterations with a crossover ratio of %70 and a mutation rate of %5, and this experiment repeated 30 times. The starting parameters for the Dethloff CON 3.1 and SCA3.1 problems are given in Table 1.

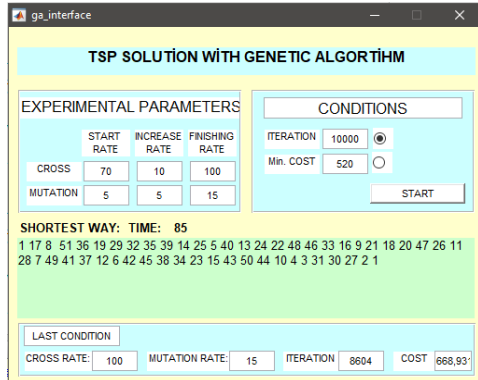


Figure 10: GA user interface on Matlab

Table 2: Parameter setting for ACO

Test No	Control Parameters			
	α	β	q_0	ρ
1	1	3,5	0,7	0,1
2	1	4	0,7	0,1
3	1	4,5	0,7	0,1
4	1	5	0,7	0,1
5	1	3,5	0,8	0,1
6	1	4	0,8	0,1
7	1	4,5	0,8	0,1
8	1	5	0,8	0,1
9	1	3,5	0,9	0,1
10	1	4	0,9	0,1
11	1	4,5	0,9	0,1
12	1	5	0,9	0,1

ACO is coded with C # through Visual Studio. For the initial condition alpha values 1 and P values were taken as 0.1, β is increased from 3,5 to 5 with 0.5 increments, q_0 is increased from 0,7 to 0,9 with increments of 0,1. Because of comparison, 30 repeats and 10000 iterations were performed in each test. The starting parameters for the CON 3.1 and SCA3.1 problems are given in Table 2.

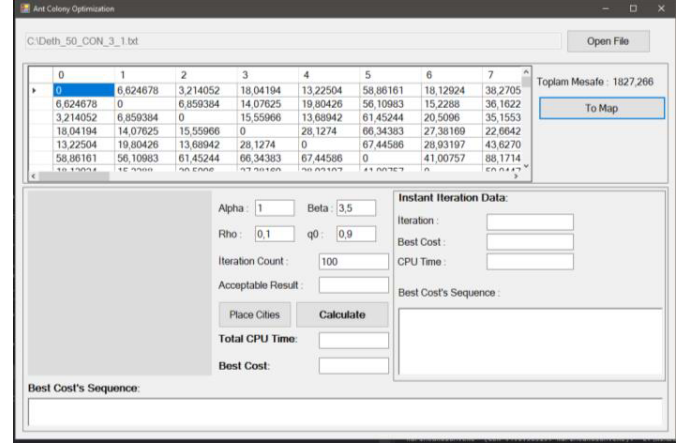


Figure 11: ACO user interface on Visual C#

V. COMPUTATIONAL RESULTS AND DISCUS

Table 3: Results comparisons between GA and ACO on CON3.1

Test No	GA					ACO				
	Best	CPU Time for Best (s)	Mean	Mean CPU Time	Standard Deviation	Best	CPU Time for Best (s)	Mean	Mean CPU Time	Standard Deviation
1	597,28	75	656,73	74	35,02	514,76	124	530,29	124	8,50
2	600,94	96	667,94	95	45,57	517,26	122	534,64	124	6,60
3	596,04	129	668,73	128	45,15	524,92	124	536,12	124	5,76
4	588,88	77	665,96	77	36,87	524,57	125	534,20	124	5,47
5	589,28	99	660,85	99	45,96	517,26	111	526,80	118	8,44
6	595,18	131	668,98	132	43,99	522,51	110	537,17	111	6,44
7	584,27	82	669,56	81	42,86	529,02	118	538,77	113	4,78
8	570,00	105	659,87	104	45,95	517,26	115	540,19	119	6,72
9	589,81	135	666,50	137	49,93	519,36	102	539,31	102	7,57
10	603,27	86	670,28	86	31,08	516,89	101	541,34	102	6,68
11	579,37	107	654,73	108	46,28	534,78	102	542,18	105	3,95
12	634,18	141	673,80	141	29,37	536,48	104	544,94	104	3,63

Table 4: Results comparisons between GA and ACO on SCA3.1

Test No	GA					ACO				
	Best	CPU Time for Best (s)	Mean	Mean CPU Time	Standard Deviation	Best	CPU Time for Best (s)	Mean	Mean CPU Time	Standard Deviation
1	721,16	73	845,18	73	72,22	608,08	122	616,05	102	3,55
2	749,96	96	863,19	95	61,17	605,68	125	614,92	106	4,41
3	758,20	128	854,86	128	58,85	609,01	128	616,85	123	3,33
4	725,29	78	863,40	77	83,45	608,08	121	616,29	106	4,18
5	695,76	98	842,65	99	64,56	607,64	111	615,88	96	4,36
6	703,29	132	836,76	132	63,98	612,88	111	617,55	100	2,26
7	738,41	82	858,39	82	62,64	611,59	111	617,88	101	3,85
8	740,13	102	843,02	103	54,47	610,76	111	618,90	100	5,39
9	713,12	136	846,65	136	70,25	608,61	101	619,99	91	5,11
10	707,52	85	832,11	86	74,82	615,01	101	620,46	94	5,19
11	747,19	107	865,50	108	59,87	609,87	101	620,28	88	4,69
12	708,08	142	846,80	141	72,40	612,88	106	620,28	94	4,69

The results are shown in Table 3 and Table 4. The table shows the CPU time of the path with the least cost for each parameter, the average results of 30 tests for each parameter, the average CPU times and the standard deviation values

For GA CON 3.1 problem, while overall mutation was 10%, low costs were obtained. The minimum cost was obtained with 100% crossover and 10% mutation (Figure 12, Test No 11).

As shown in Figure 12 and Test 12, the mean cost was found as the maximum, while the standard deviation was the minimum value.

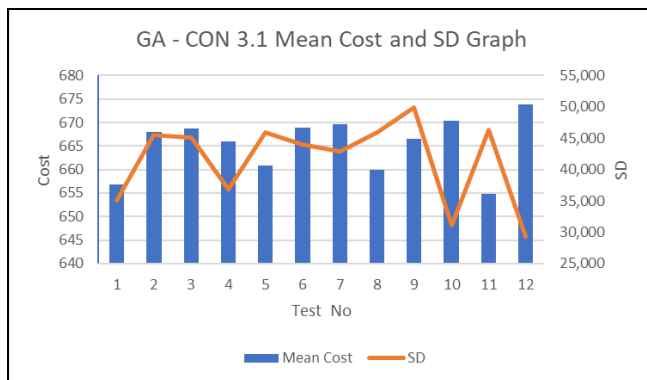


Figure 12: GA CON 3.1 Mean Cost and SD Graph

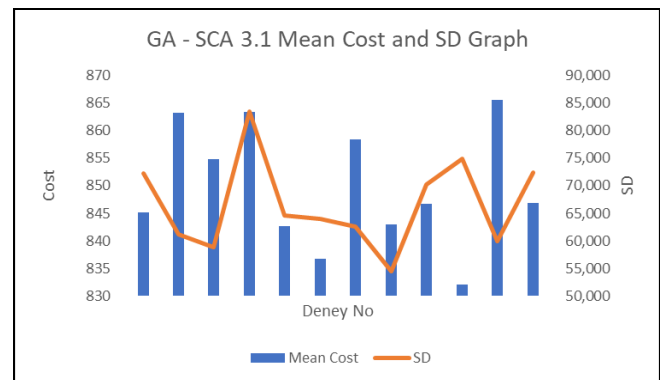


Figure 13: GA SCA 3.1 Mean Cost and SD Graph

In GA, the minimum average cost for the SCA problem solution (Figure 13) was obtained in Test No 10, while the highest average cost was obtained in Test No 11. As can be seen from Figure 13, it was determined that the mutation, which observed that the standard deviation values were high, performed stronger searches in the solution space and the results showed heterogeneous distribution.

Figure 16 shows that SD values are between 30 and 50. In Figure 17 these values, it was found between 55 and 85. Based on these data, more of premature convergence was observed in CON 3.1 problem solved with GA.

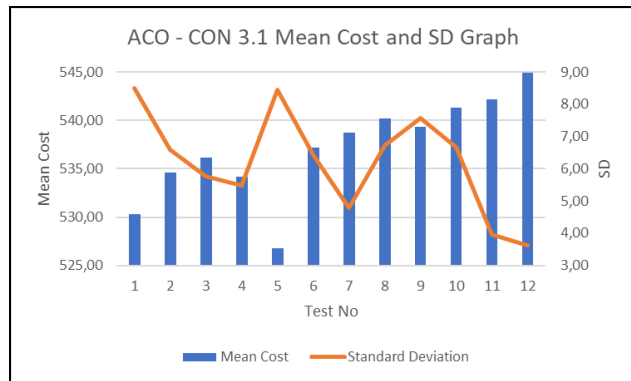


Figure 14: ACO CON 3.1 Mean Cost and SD Graph

Based on the distribution of SD values, ACO results show homogeneous distribution. The best result of the tests is observed at Test No 5. Figure 14 shows that β and q_0 are inversely proportional to average cost.

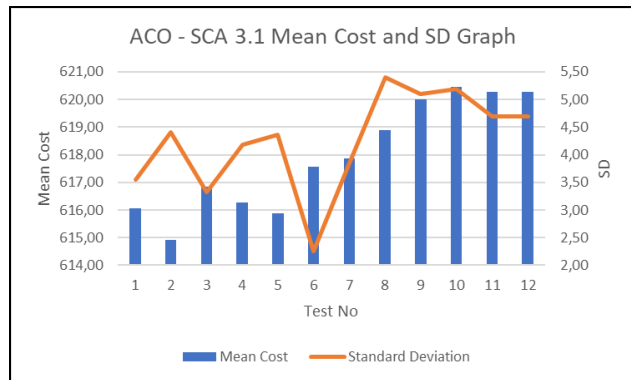


Figure 15: ACO SCA 3.1 Mean Cost and SD Graph

In the SCA 3.1 problems for ACO, it can be seen from Figure 15 that q_0 0,8 and β 4 produce a lower cost. The minimum SD value is in the β 4 and q_0 0.8 parameter values, the minimum average result is clearly seen in β 4 and q_0 0.7. For ACO, the general difference between CON and SCA is that the SD value decreases on average when the parameter values increase as shown in Figure 14.

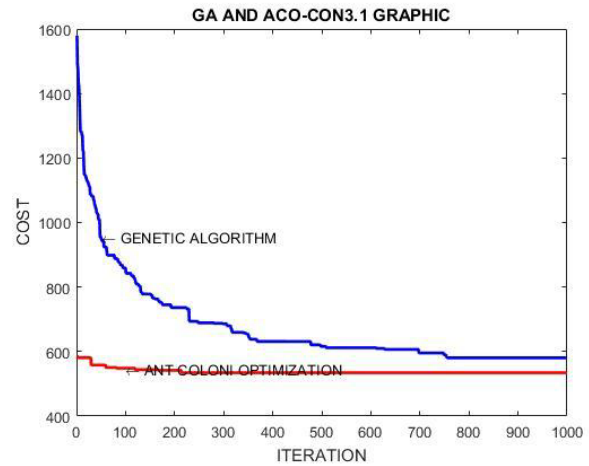


Figure 16: CON 3.1 Cost – Iteration Graph between GA and ACO

ACO iterations in search space seek a solution based on the distance between the nodes. The effect of distance is seen on Figure 20, initial cost is less than 600. GA can generate close to ACO with more iterations by random search. In Figure 20, it is seen that GA has more breaking point than ACO. The breakpoints in the graphs indicate that the algorithms shift from a local optimum point to another local optimum point.

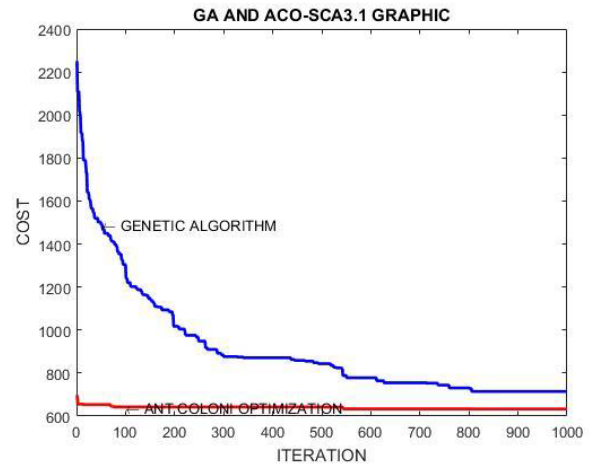


Figure 17 : SCA 3.1 Cost – Iteration Graph between GA and ACO

The performance progress in ACO and GA in Figure 16 is also seen in Figure 17. In terms of the solution approach to TSP problems, there is a clear difference between ACO and GA under equal conditions. ACO achieves the best result from 500 to 600 iterations, starting at a very close to the minimum result, Reduction of up to 10000 iterations continues in GA.

VI. CONCLUSION

In this study we presented the performance analysis between GA and ACO. As a result of comparison of GA and ACO, it has been observed that ACO is more successful as meta heuristic solution algorithm in optimization problems. The best

results in GA were 90% crossover and 10% mutation ratio in the CON problem, 80% crossover and 10% mutation rate in SCA problem. The best results in ACO are β 3.5 and q_0 0.7 for the CON problem, β 4 and q_0 0.7 values in SCA problem. Best results for 51-node TSP in GA and ACO were found in approximately 100 seconds. In the case of larger TSP problems, the most effective result can only be achieved with a proportionally increased dwell time. It can be predicted that this situation can be overcome by parallel processing.

REFERENCES

- [1] S. Colak, Genetik Algoritmalar Yardimi ile Gezgin Satirci Probleminin Cözümlü Üzerine Uygulamalar, vol. 19, no. 3. 2010.
- [2] G. Laporte and U. Palekar, "Some applications of the clustered travelling salesman problem," J. Oper. Res. Soc., 2002.
- [3] G. J. Woeginger, "Exact Algorithms for NP-Hard Problems: A Survey," 2007.
- [4] Razali N M and Geraghty J, "Genetic algorithm performance with different selection strategies in solving TSP," Proc. world Congr. Eng., 2011.
- [5] M. Dorigo and L. M. Gambardella, "Ant colonies for the travelling salesman problem," BioSystems, 1997.
- [6] J. H. Holland, Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. 1975.
- [7] D. E. Goldberg and K. Sastry, "Analysis of Mixing in Genetic Algorithms: A Survey," Building, 2002.
- [8] M. N. Ab Wahab, S. Nefti-Meziani, and A. Atyabi, "A comprehensive review of swarm optimization algorithms," PLoS One, vol. 10, no. 5, 2015.
- [9] M. Dorigo, V. Maniezzo, A. Colomi, and M. Dorigo, "Positive Feedback as a Search Strategy," Tech. Rep. 91-016, 1991.
- [10] M. Dorigo, "Optimization, learning and natural algorithms," PhD thesis, Dip. di Elettron. Politec. di Milano, 1992.
- [11] M. Dorigo, G. Di Caro, and L. M. Gambardella, "Ant algorithms for discrete optimization," Artif. Life, 1999.
- [12] B. Hölldobler and E. O. Wilson, The ants. 1990.
- [13] R. Beckers, J. L. Deneubourg, and S. Goss, "Trail laying behaviour during food recruitment in the ant *Lasius niger* (L.)," Insectes Soc., 1992.
- [14] R. Beckers, J. L. Deneubourg, and S. Goss, "Trails and U-turns in the selection of a path by the ant *Lasius niger*," J. Theor. Biol., 1992.
- [15] M. Rajasekaran and A. C. Vaithlingam, "Maximum power point tracking for PV array based on ant colony optimization under uniform and non-uniform irradiance," Int. J. Intellect. Adv. Res. Eng. Comput., vol. 5, no. 2, 2017.
- [16] J. Dethloff, "Relation between vehicle routing problems: An insertion heuristic for the vehicle routing problem with simultaneous delivery and pick-up applied to the vehicle routing problem with backhauls," J. Oper. Res. Soc., 2002.
- [17] F. Simsir and D. Ekmekci, "A metaheuristic solution approach to capacitated vehicle routing and network optimization," Engineering Science and Technology, an International Journal, 2019.

On forecasting ability of a logistic model for prostate cancer under intermittent androgen suppression therapy

O. OZTURK MIZRAK¹ and C. MIZRAK¹

¹ Karabuk University, Karabuk/Turkey, cihanmizrak@karabuk.edu.tr

¹ Karabuk University, Karabuk/Turkey, ozlemozturk@karabuk.edu.tr

Abstract – In this work, we handle some logistic models generated as an alternative against to cell quota models in the prostate cancer modelling literature. We demonstrate that even if these models serve nearly better data fitting with respect to their cell-quota counterparts, it doesn't mean they always serve better forecasting results with numerical simulations applying ensemble Kalman Filter for this prediction process.

Keywords – logistic models, prostate cancer modelling, data fitting, forecasting, ensemble Kalman Filter

I. INTRODUCTION

In the 1930s and 40s Charles Huggins and his colleagues demonstrated that surgical castration often leads to a significant regression of prostate cancer and in 1966 he was awarded the Nobel Prize in Medicine and Physiology [1]. Today, androgen deprivation therapy (ADT) performs the same goal without surgery; however, the treatment is expensive and has many negative side effects such as sexual dysfunction and dementia [2].

Continuous androgen deprivation (CAD) treatment is a standard treatment method applied after the first radiation therapy that failed for localized advanced prostate cancer patients [2-5]. On the other hand, most patients eventually develop resistance to treatment, and then the disease becomes more aggressive and the prognosis is poor at this stage [6-7]. Therefore, it is very important to estimate when resistance will occur in a patient, to improve the quality of life and to prevent treatment in vain. Intermittent androgen deprivation (IAD) aims to reduce side effects and delay development of resistance; however, the delay in resistance with this treatment is still controversial and CAD remains the standard treatment method [4].

A solid understanding of the treatment of prostate cancer and the lack of standards trigger the need for mathematical models. Especially in the last 15 years, many mathematical models have been proposed to help explain the functioning dynamics of prostate cancer in the hope of answering the above mentioned questions.

Jackson [8] developed the first prostate cancer model in

2004, based on experimental data, under CAD treatment, which would lead to modeling efforts to be shown later. In 2008, Ideta et al. [9] demonstrated a mathematical model under IAD therapy, including the mutation of androgen-independent (AI) cells from androgen-dependent (AD) cells. Shimada and Aihara [10] investigated the competition between different prostate cell populations. Another approach to competition modeling is based on Ideta's model. [9]. Eikenberry et al. [11] developed Ideta's model to investigate the evolutionary role of androgens on prostate cancer, taking into account intracellular signals. This model was later described by Portz et al. [12] (PKN) to ensure compliance with clinical PSA (prostate-specific antigen) data, and the PKN model was then simplified by Baez and Kuang (BK) [13] to ensure compliance with both PSA and androgen data. As an another expansion of the models of Ideta [9] and Eikenberry [11], Jain et al. [14] provided that the model they generate captures the biochemical dynamic of prostate cancer in detail. On the other hand, Hirata et al. [15] developed a fragmented linear model taking into account three populations of cells to ensure compliance with clinical PSA data. Many researchers have studied the parameter estimation [16-17], optimum change times for IAD and control [16]; Hirata et al. [18] and the prediction of progression of resistant prostate cancer [19] using the model of Hirata et al. [15].

Here we especially focus on the BK model as it is one of the latest version of PKN model and has a good data fitting and forecasting ability. On the other hand, although cell quota phenomenon used in both PKN and BK models is an original approach to reflect the limiting effect of nutrient content in cancer cell populations, the minimum androgen amounts (cell quotas) required for survival of cell populations of castration sensitive (CS) and resistant (CR) are not easily measurable, while serum androgen data are easily available in the clinic. Therefore, this situation poses a problem in moving the model to clinical practice. In addition, in the BK model, the authors made data fitting and prediction under the hypothesis that the amount of androgen in the serum is equal to the amount of androgen in the cancer cells. However, this assumption is contrary to the biological fact that a portion of the serum androgen reaches the cell by diffusion, so it has misleading

results, such as at the time of the discontinuation of IAD treatment, the intracellular androgen reaches its maximum level instantly.

At this point, our question is: "Can we create a new model that allows us to use androgen data directly by representing the growth of cancerous cells with well-known logistic growth terms rather than cell quota?". As an answer to this question, we developed two logistic models below:

One population logistic model

$$\frac{dx}{dt} = r(A, x)x \quad (1)$$

$$\frac{dv}{dt} = -dv \quad (2)$$

$$\frac{dA}{dt} = \underbrace{\gamma(a_0 - A)}_{diffusion} - \underbrace{\mu \frac{A^\omega}{A^\omega + \rho^\omega} x}_{uptake} \quad (3)$$

$$\frac{dP}{dt} = \underbrace{bA}_{baseline} + \underbrace{\sigma Ax}_{tumor} - \underbrace{\epsilon P}_{clearance} \quad (4)$$

production

where

$$r_1(A, x) = \underbrace{\mu \frac{A^\omega}{A^\omega + k^\omega} (1 - \frac{x}{\theta})}_{growth} - \underbrace{(d_1 \frac{R_1}{A + R_1} + \delta_1 x_1)}_{death} \quad (5)$$

$$\gamma = \gamma_1 u(t) + \gamma_2, \quad u(t) = \begin{cases} 1, & \text{on treatment} \\ 0, & \text{off treatment} \end{cases} \quad (6)$$

Two population logistic model

$$\frac{dx_1}{dt} = r_1(A, x_1, x_2)x_1 - m(A)x_1 \quad (7)$$

$$\frac{dx_2}{dt} = r_2(A, x_1, x_2)x_2 + m(A)x_1 \quad (8)$$

$$\frac{dA}{dt} = \underbrace{\gamma(a_0 - A)}_{diffusion} - \underbrace{\mu \frac{A^\omega}{A^\omega + \rho^\omega} (x_1 + x_2)}_{uptake} \quad (9)$$

$$\frac{dP}{dt} = \underbrace{bA}_{baseline} + \underbrace{\sigma A(x_1 + x_2)}_{tumor} - \underbrace{\epsilon P}_{clearance} \quad (10)$$

production

where

$$r_1(A, x_1, x_2) = \underbrace{\mu \frac{A^m}{A^m + k^m} (1 - \frac{x_1 + x_2}{\theta})}_{growth} - \underbrace{(d_1 \frac{R_1}{A + R_1} + \delta_1 x_1)}_{death} \quad (11)$$

$$r_2(A, x_1, x_2) = \underbrace{\mu \frac{A^n}{A^n + k^n} (1 - \frac{x_1 + x_2}{\theta})}_{growth} - \underbrace{(d_2 \frac{R_2}{A + R_2} + \delta_2 x_2)}_{death} \quad (12)$$

$$m(A) = \underbrace{\frac{cK}{A + K}}_{CS \text{ to } CR} \quad (13)$$

Since the terms other than growth and uptake are similar to those in the BK models, we avoid explaining them repeatedly. Here, we hypothesise that proliferative advantage of cells over each other changes as to androgen levels that are represented by A in the model formulations. In other words, while below a threshold level of androgen, CR cells have greater capacity for proliferation compared to CS cells, above this threshold value CS cells gain proliferative advantage over CR reversely. To be able to capture this assumption, we prefer to use Hill functions with constraint $n < m$. Thus while we vary n between 0.01 and 2.99, m is assumed as $\ln n + 1$. Due to the fact that we desire to have no significantly change in Hill functions for androgen rich environments we vary k between 0.1 and 1. We also avoid to choose far less quantities for lower bounds of n and k not to encounter with any mathematical uncertainty in case of low androgen levels approaching zero. Likewise as androgen within the cells is used for growth, we formulate the uptake term using a Hill function that allows the uptake rate approaches its maximum, μ , when the serum androgen concentration increases. In this term, we expect $\omega \in [1, 3]$ and $\rho \in [(a/2) - 2, (a/2)]$ where a is the maximum androgen level in patient's serum. Other parameter values are nearly same as parameter values in BK models. Differently, carrying capacity of cancer cells is taken as estimated by Rutter and Kuang [20].

For the analysis and calibration of these proposed logistic models during the research, [2] data from 109 patients treated at the Vancouver Prostate Center (Vancouver, BC, Canada) are used. In the first one-and-a-half cycle of treatment, at least 20 points for both androgens and PSAs are selected. Of the 109 patients, 62 are selected, and the proposed models provide nearly better data fitting results with less mean squared errors (please see Table 1).

Model	PSA			Androgen		
	Min	Mean	Max	Min	Mean	Max
One pop-logistic	0,173931	2,445644	36,51354	0,968212	10,34476	49,41232
One pop-BK	0,058844	2,46837	25,04435	0,773269	8,656484	51,09163
Two pop-logistic	0,088222	2,438892	44,1696	0,801594	10,16664	45,70097
Two pop-BK	0,353765	3,766446	69,38941	0,998116	12,53949	55,10578

Table 1: Comparison of Mean Squared Error (MSE) for Androgen and prostate-specific-antigen (PSA) for the first 1.5 cycles

However, does better data fitting to a particular set of data mean better forecasting for a model when compared with its predecessor? Or can a model succeed in predicting the results of future observations with an acceptable degree of reliability if it is able to produce the results of observation in a limited time? If not, what are the reasons for this?

II. METHOD AND NUMERICAL SIMULATIONS

For the above questions, we utilize from the ensemble Kalman filter (EnKF) which is used in the estimation of the BK model or in the data assimilation, weather forecasting and many other forecasting processes. If we briefly summarize this approach here, starting with a set of K initial conditions at t_n , we integrate the model to time t_{n+1} to produce a background ensemble ($\{\mathbf{x}(t_{n+1})_{b(k)}\}_{k=1}^K$). Let $\mathbf{y}^{\text{obs}}(t_{n+1})$ denote the vector of measurements at t_{n+1} . Here we assume that the observables (i.e., predicted observations) are related to the model state vector by the function \mathbf{H} , sometimes called the forward operator, as

$$\mathbf{y}(t_{n+1}) = \mathbf{H}(\mathbf{x}(t_{n+1}))$$

The objective of the ensemble transform Kalman filter is to find a linear combination of the elements of the background ensemble that best match the data, in the sense that the weight vector \mathbf{w} minimizes

$$J(\mathbf{w}) = (K-1)\mathbf{w}^T\mathbf{w} + [\mathbf{y}_{\text{obs}} - \mathbf{H}(\bar{\mathbf{x}}_b + \mathbf{X}_b\mathbf{w})]^T \mathbf{R}^{-1} [\mathbf{y}_{\text{obs}} - \mathbf{H}(\bar{\mathbf{x}}_b + \mathbf{X}_b\mathbf{w})]$$

where

$$\bar{\mathbf{x}}_b = K^{-1} \sum_{k=1}^K \mathbf{x}(t_{n+1})_{b(k)}$$

is the ensemble mean. \mathbf{X}_b is the $m \times K$ ensemble perturbation matrix, \mathbf{R} is the covariance matrix of the errors in the observations, and \mathbf{H} is the forward operator associated with the model. The filter produces the analysis ensemble $\{\mathbf{w}_{a(k)}\}_{k=1}^K$, which in model space becomes

$$\mathbf{x}(t_{n+1})_{a(k)} = \bar{\mathbf{x}}_b + \mathbf{X}_b\mathbf{w}_{a(k)}, \quad i = 1, \dots, K$$

Please notice that we give numerical simulations as an application of EnKF just for two population logistic model that has slightly better fitting results with respect to one population form.

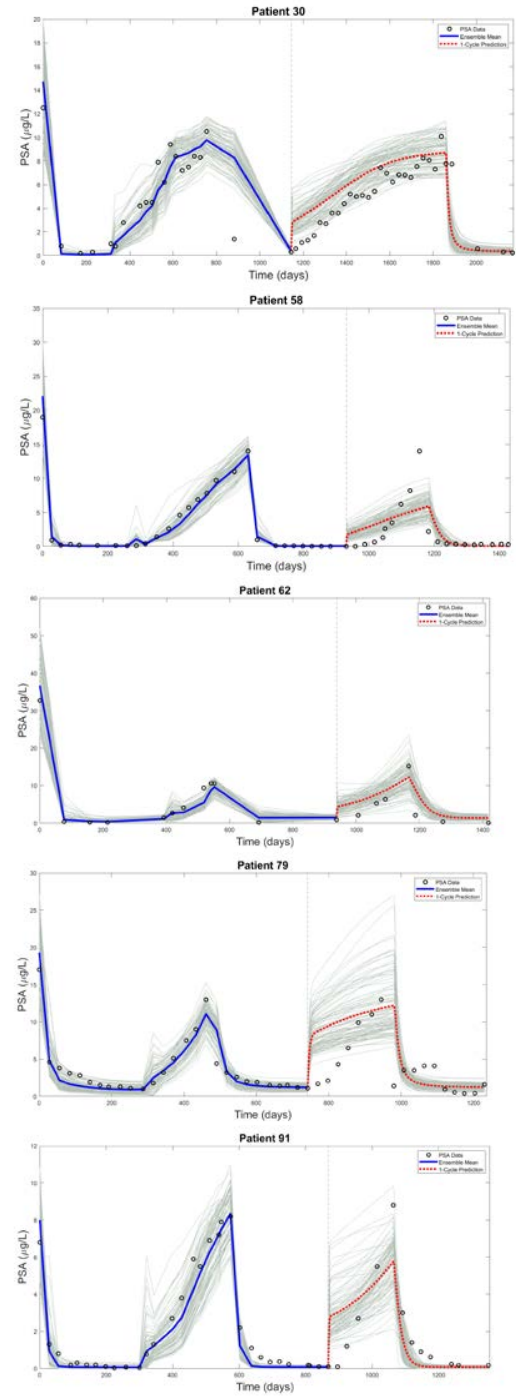


Figure 1: Simulations of fittings for every model for 1.5 cycles of treatment (left of dashed gray line) and one cycle of forecast (right of dashed gray line). For these five patients, we can see that two population logistic model provides more or less fine forecasting results.

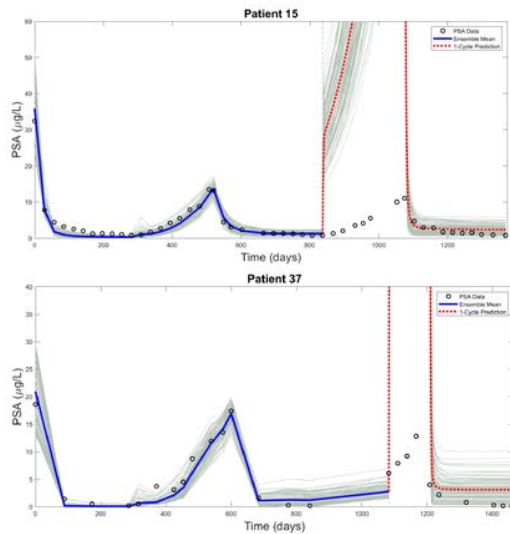


Figure 2: Simulations of fittings for every model for 1.5 cycles of treatment (left of dashed gray line) and one cycle of forecast (right of dashed gray line). For these two patients, we can see that two population logistic model provides much worse forecasting results with respect to patients in Figure 1.

III. CONCLUSION

Our main conclusion for this research is that good data fitting results do not guarantee good forecasting results. Even if some simulations belong to different patients' data give prospective prediction results, some of them fails. For future work, identifiability of the model parameters must be handled to see which factors affect forecasting results. Sensitivity analysis of the parameters also should be done to determine the most and least sensitive parameters. Fixing the less sensitive ones and varying the others, model uncertainty may be reduced.

ACKNOWLEDGMENT

We are grateful to the late Dr. Nicholas Bruchovsky for the clinical data set.

REFERENCES

- [1] S. R. Denmeade and J. T. Isaacs, "A history of prostate cancer treatment", *Nature Reviews Cancer* 2, 5, 389-396 (2002).
- [2] N. Bruchovsky, L. Klotz, J. Crook, S. Malone, C. Ludgate, W. J. Morris and M. E. Gleave, S. L. Goldenberg, "Final results of the Canadian prospective Phase II trial of intermittent androgen suppression for men in biochemical recurrence after radiotherapy for locally advanced prostate cancer: Clinical parameters", *Cancer* 107, 2, 389-395.
- [3] T. Nishiyama, "Serum testosterone levels after medical or surgical androgen deprivation: a comprehensive review of the literature.", *Urologic oncology* 32, 1, 38.e17-28 (2013).
- [4] Crook, J. M., C. J. O'Callaghan, G. Duncan, D. P. Dearnaley, C. S. Higano, E. M. Horwitz, E. Frymire, S. Malone, J. Chin, A. Nabid, P. Warde, T. Corbett, S. Angyal, S. L. Goldenberg, M. K. Gospodarowicz, F. Saad, J. P. Logue, E. Hall, P. F. Schellhammer, K. Ding and L. Klotz, "Intermittent Androgen Suppression for Rising PSA Level after Radiotherapy", *New England Journal of Medicine* 367, 10, 895-903 (2012).
- [5] Bryce, A. H. and E. S. Antonarakis, "Androgen receptor splice variant 7 in castration resistant prostate cancer: Clinical considerations", *International Journal of Urology* 23, 8, 646-653 (2016)

- [6] Feldman, B. and D. Feldman, "The development of androgen-independent prostate cancer", *Nature Reviews Cancer* 1, 1, 34-45 (2001).
- [7] Deaths, P. C., "Cancer Statistics, 2011 The Impact of Eliminating Socioeconomic and Racial Disparities on Premature Cancer Deaths", (2011).
- [8] T. L. Jackson, "A mathematical model of prostate tumor growth and androgen-independent relapse", *Discrete & Continuous Dynamical Systems-Series B*, 4 (2004), 187-202.
- [9] A. M. Ideta, G. Tanaka, T. Takeuchi and K. Aihara, "A mathematical model of intermittent androgen suppression for prostate cancer", *Journal of nonlinear science*, 18 (2008), 593-614.
- [10] T. Shimada and K. Aihara, "A nonlinear model with competition between prostate tumor cells and its application to intermittent androgen suppression therapy of prostate cancer", *Mathematical biosciences*, 214 (2008), 134-139.
- [11] S. E. Eikenberry, J. D. Nagy and Y. Kuang, "The evolutionary impact of androgen levels on prostate cancer in a multi-scale mathematical model", *Biology direct*, 5 (2010), 24.
- [12] T. Portz, Y. Kuang and J. D. Nagy, "A clinical data validated mathematical model of prostate cancer growth under intermittent androgen suppression therapy", *Aip Advances*, 2 (2012), 011002.
- [13] J. Baez and Y. Kuang, "Mathematical models of androgen resistance in prostate cancer patients under intermittent androgen suppression therapy", *Applied Sciences*, 6 (2016), 352.
- [14] H. Vardhan Jain and A. Friedman, "Modeling prostate cancer response to continuous versus intermittent androgen ablation therapy", *Discrete & Continuous Dynamical Systems-Series B*, 18.
- [15] Y. Hirata, N. Bruchovsky and K. Aihara, "Development of a mathematical model that predicts the outcome of hormone therapy for prostate cancer", *Journal of theoretical biology*, 264 (2010), 517-527.
- [16] Guo, Q., Z. Lu, Y. Hirata and K. Aihara, "Parameter estimation and optimal scheduling algorithm for a mathematical model of intermittent androgen suppression therapy for prostate cancer", *Chaos* 23, 4, 43125 (2013).
- [17] Tao, Y., Q. Guo and K. Aihara, "A partial differential equation model and its reduction to an ordinary differential equation model for prostate tumor growth under intermittent hormone therapy", *Journal of Mathematical Biology* pp. 1-22 (2013).
- [18] Hirata, Y., K. Akakura, C. S. Higano, N. Bruchovsky and K. Aihara, "Quantitative mathematical modeling of PSA dynamics of prostate cancer patients treated with intermittent androgen suppression", *Journal of Molecular Cell Biology* 4, 3, 127-132 (2012).
- [19] Hirata, Y., S.-i. Azuma and K. Aihara, "Model predictive control for optimally scheduling intermittent androgen suppression of prostate cancer", *Methods* 67, 3, 278-281 (2014).
- [20] E.M. Rutter and Y. Kuang, 2017. Global dynamics of a model of joint hormone treatment with dendritic cell vaccine for prostate cancer, *DCDS-B*, 22, 1001-1021.

Comparison of Classification Algorithms on Heart Disease Data

Y.ORTAKCI¹ and B.MADEN¹

¹ Karabuk University, Karabuk/Turkey, yaasinortakci@gmail.com

¹Karabuk University, Karabuk/Turkey, burakmadenn@gmail.com

Abstract - Today, early diagnosis in the treatment of diseases is of great importance. In this study, the classification algorithms of machine learning, k-Nearest Neighbor, Support Vector Machine, Decision Tree, and Naive Bayes were used for the early diagnosis of heart diseases. Firstly, the software we developed for these algorithms were tested on benchmark datasets well known in the literature such as Iris, Wine and Glass. Our results showed that the classification success of our software was similar to the results obtained at other studies in the literature. Then, these algorithms were tested on the Heart Disease dataset and the best estimation was made by the Support Vector Machine algorithm. The Decision Tree, and Naive Bayes algorithms, respectively, showed a good estimation success. The performance of k-Nearest Neighbor algorithm was lower than the other three algorithms. The results indicate that Support Vector Machine, Decision Tree, and Naive Bayes algorithms can be used to determine whether people have heart disease.

Keywords – Classification, algorithms, KNN, SVM, DT, NB.

I. INTRODUCTION

TODAY classification methods are used in the early diagnosis of diseases in the medical. Early diagnosis, which is of great importance in the treatment of heart diseases, enables doctors to start treatment to the patients early on. Thus, the risk of serious heart diseases and heart attacks can be reduced. The data used in the classification are one of the most important factors determining the accuracy of the classification results. Another factor affecting accuracy is the selection of appropriate classification algorithm and its parameters.

In this study, the classification of the Heart Disease dataset was performed with different algorithms and the results were compared. As classification algorithms, k-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT) and Naive Bayes (NB) algorithms were performed and their performance on Iris, Wine and Glass such as well-known classification datasets were evaluated. Then, the classification success of these algorithms on the Heart Disease dataset was measured and compared. The results have shown that these algorithms can be used in the diagnosis of heart diseases.

II. RELATED WORKS

Classification algorithms are widely used in prediction studies. Deepti and Dilip estimated the probability of people becoming diabetic using the SVM, DT and NB algorithms.

They used accuracy, precision, f-measure and recall values as comparison criteria in the classification process. They mentioned that Naive Bayes gives the best results with 76.30% estimation success [1].

Karakoyun ve Hacibeyoglu run KNN, NB, CN2, SVM, Random Forest (RF) and Artificial Neural Network (ANN) classification algorithms on nine different datasets and compared the algorithms in terms of classification success and time. While the best classification success was obtained by Artificial Neural Network, the minimum time was spent by KNN [2].

Ozgur and Erdem have classified the KDD99 dataset used in intrusion detection system with DT, ANN, SVM and, Adaboost algorithms by performing a performance analyzes among these algorithms [3].

Naik and Samant run DT, NB, KNN, and Decision Stumps algorithms on Rapid Minner, WEKA, Tanagra, Orange, and Knime data mining software for classification purpose of Indian people with and without Liver disorder and Knime tool gave the best results [4].

Sina and Mandara compared SVM, DT, KNN, NB, Logistic Regression, and Discriminant Analysis classification algorithms on ten different clinical datasets. As a result, they have determined that the SVM has reached the best result [5].

As a result, the classification algorithms used may show different performance on different datasets. In this study, we will measure the classification performance of KNN, SVM, DT and, NB algorithms on heart diseases dataset.

III. CLASSIFICATION ALGORITHMS

A. k-Nearest Neighbors (KNN)

KNN is one of the nonparametric and instance based basic algorithms used in classification [6]. It has been also used in many fields such as pattern recognition, statistical estimation etc. since 1970. KNN is a simple, effective and fast classifier, which groups the samples in the dataset based on their nearest neighbor classes by taking similarity measurements such as Euclidian, Minkowski, and Chebyshev distances into account [7]. Firstly, the dataset is divided into two parts as training and test data. In the training phase of the algorithm, it is provided to determine the classes using this training set in the n-dimensional space. Then, in the test phase, the sample data is scattered to the appropriate classes. To find the appropriate class, the distances of the sample data to the points in the

training set are calculated and these distances are ordered ascendingly and the k neighbor samples with the lowest distance are selected. Then the sample data is assigned to the class which most of these selected neighbors belong to.

In this study, we used Euclidian distance [8] as a similarity measurement.

$X_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}$ and $X_j = \{x_{j1}, x_{j2}, \dots, x_{jn}\}$
where,
 X_i, X_j : i^{th} and j^{th} sample points
 n : the number of attributes

$$distance(X_i, X_j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (1)$$

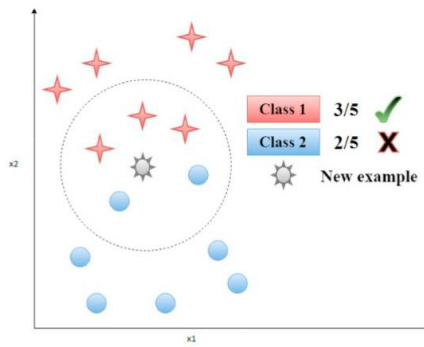


Figure 1: Classification with KNN [9].

Figure 1 shows how KNN algorithm classifies data visually. Five of the neighbor data adjacent to the new example data are shown in the circle. Three of these data belong to Class 1, while the other two belong to Class 2. Since three of the five neighbors of the sample data belong to Class 1, the new example data will be assigned to Class 1.

B. Support Vector Machine (SVM)

SVM is a widely used, supervised, and nonparametric classification algorithm running based on a statistical learning theory [10, 11]. Its mathematical formulation fits quadratic programming but needs more computational time to solve large datasets [12]. The SVM determines the classes of sample data in an n -dimensional space using the training data. For this, it defines a hyperplane passing through two classes by keeping the distance maximum between these classes in order to crisply separate them [13]. In the hyperplane determination, the SVM determines the most extreme data points in the dataset to separate the classes. Many hyper-planes can be drawn to separate classes, but hyperplanes should be drawn to keep the distance maximum between them.

Mathematically, a hyperplane is formed to make the distance maximum between the hyperplane indicated by $w \cdot xi + b = -1$ and the hyperplane indicated by $w \cdot xi + b = 1$. This hyperplane is represented by $w \cdot xi + b = 0$. The margin is

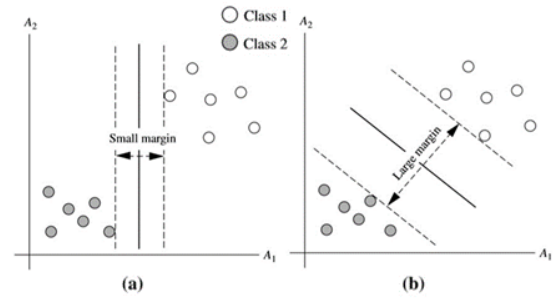


Figure 2: Hyperplane of SVM [8].

indicated by $\frac{2}{\|w\|}$, so in order to get maximum margin, $\|w\|$ must be minimum [1].

Figure 2 shows hyperplanes separating the two classes. The margin between the two classes is narrower in Figure 2.a while it is wider in Figure 2.b, that means the classification in Figure 2.b gives more accurate results than the classification in Figure 2.a.

C. Decision Tree (DT)

Decision Tree is one of the algorithms used for classification problems in machine learning. The decision tree creates a decision rule by using the data in the training set, thus allowing the target classes to be estimated [1]. The decision tree make classification using tree-structure node model which includes root and additional nodes on the tree [14]. Each test data is expected to be assigned to the appropriate class by exploiting of the previously established decision rule [10]. The sample data proceeds from the root node to the chance nodes and, finally, to the end node that will determine the class. Figure 2 shows an example tree structure of DT algorithm.

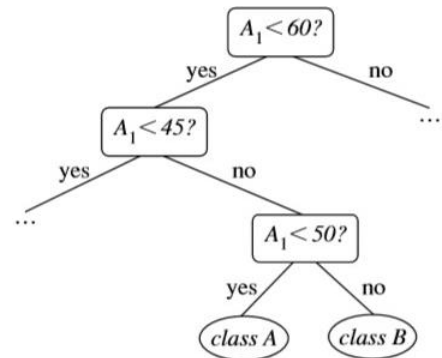


Figure 3: A sample of DT model [8].

D. Naive Bayes (NB)

NB algorithm, which is one of the machine learning algorithms used for classification, uses Bayes' probability theorem [15]. Its variables are independent of the classes and attributes [16]. It is powerful against missing data or imbalances in the dataset [1]. In class determination, the probability of belonging to each class for the sample data is calculated and the class with the highest probability value is selected [17]. $P(X|Y)$ must be calculated to find the probability that a

sample data belongs to a class.

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \quad (2)$$

where,

$P(X|Y)$: Possibility of occurrence of event X when event Y occurs

$P(Y|X)$: Possibility of occurrence of event Y when event X occurs

$P(A), P(B)$: Prior probability of A and B events, respectively

IV. EXPERIMENTS AND RESULTS

In this study, the classification performances of the KNN, SVM, DT and NB machine learning algorithms on Iris, Wine and Glass datasets were tested. Subsequently, the performance of these algorithms was measured on the Heart Disease dataset.

A. Definition of Datasets

Iris Dataset: Iris dataset [18] is one of the benchmark test data commonly used to measure the performance of the classification algorithms in pattern recognition at machine learning. It consists of three classes, each containing 50 sample data, and each sample has four attributes. Table 1 gives an overview of the iris dataset.

Wine Dataset: Wine dataset [18], which keeps the chemical properties of wines made by different cultures in the same region in Italy, is one of the benchmark test data used to measure the performance of classification algorithms. It contains 178 sample data, each of which has 13 chemical attributes, is distributed to three classes. Table 1 gives an overview of the wine dataset.

Glass Dataset: Glass dataset [18], which indicates the list of chemical elements in the glass, is one of the benchmark test data used to measure the performance of classification algorithms. It contains 214 sample data, each of which has 9 chemical element attributes, is distributed to six classes. Table 1 gives an overview of the glass dataset.

Heart Disease Dataset: Although the Heart Disease [18] dataset contains 76 attributes in total, 14 of them, which effect the diagnosis of heart disease, are used in most of the classification studies. This dataset contains a total of 303 samples scattered to two classes that shows whether a person has a heart disease or not. The attributes of Heart Disease dataset are given below:

- 1) Age: Age of patient
- 2) Sex: Gender of patient (1 = male, 0 = female)
- 3) Cp: Chest paint type
- 4) Trestbps: Resting blood pressure (in mm Hg on admission to the hospital)
- 5) Chol: serum cholesterol in mg/dl
- 6) Fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
- 7) Restecg: resting electrocardiographic results
- 8) Thalac: maximum heart rate achieved
- 9) Exang: exercise induced angina (1 = yes; 0 = no)
- 10) Oldpeak: ST depression induced by exercise relative to rest
- 11) Slope: the slope of the peak exercise ST segment
- 12) Ca: number of major vessels (0-3) colored by fluoroscopy
- 13) Thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
- 14) Target: 1 or 0

Table 1: Properties of Datasets.

Dataset	Number of Attributes	Number of Sample	Classes	Attribute Types
Iris	4	150	3	Real
Wine	13	178	3	Integer, Real
Glass	9	214	6	Real
Heart Disease	14	303	2	Categorical, Real, Integer

B. Performance Measurement Criteria

In this study, Accuracy, Precision, Recall, and F-Measure values are used as criteria for measuring the performance of classification algorithms [19]. We utilized Confusion Matrix to calculate these criteria. Confusion Matrix is used to evaluate the performance of classification methods in machine learning. It is a matrix that shows the correct and incorrectly classified numbers of sample data after a classification process. Each row of the matrix indicates the sample of the actual class while each column indicates the sample of the predicted class (Figure 4). Calculations of Accuracy, Precision, Recall, and F-Measure values are shown in the following formulas:

		Predicted class	
		1	0
Actual Class	1	True Positives (TP)	False Negatives (FN)
	0	False Positives (FP)	True Negatives (TN)

Figure 4: Confusion matrix [20].

$$Accuracy = \frac{TP+TN}{P+N} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F - Measure = \frac{2*precision*recall}{precision+recall} \quad (6)$$

In this study, 60% of each dataset was used for training and 40% for testing, and same samples in the datasets were used in training and test phases of all algorithms. The optimum parameters were chosen for each algorithm in the classification of all datasets.

The applications developed within the scope of this study are coded in Python programming language utilizing the Scikit-Learn artificial intelligence library. Developed applications were run on a notebook with i7 4.1 GHz CPU and 8 GB 2666 MHz RAM.

Table 2: Classification results for Iris data.

Algorithm	Accuracy (%)	Precision	Recall	F-Measure
KNN	93.33	0.94	0.93	0.93
SVM	100	1.00	1.00	1.00
DT	100	1.00	1.00	1.00
NB	98.33	0.98	0.98	0.98

In Table 2, the classification results of the applications we developed are given for Iris data. According to these results, the best results belong to the SVM and DT with 100% estimate accuracy. The performance order of other algorithms was NB and KNN respectively. The parameters of the algorithms for Iris dataset:

- KNN: $p=2$, $n_neighbors=9$, $weights='uniform'$
- SVM: $kernel='poly'$, $gamma=0.01$, $C=100$
- DT: $criterion='gini'$, $max_features=2$, $min_samples_leaf=2$
- NB: Gaussian classifier, $priors='none'$

Table 3: Classification results for Wine data.

Algorithm	Accuracy (%)	Precision	Recall	F-Measure
KNN	76.39	0.75	0.76	0.76
SVM	97.22	0.97	0.97	0.97
DT	97.22	0.97	0.97	0.97
NB	100	1.00	1.00	1.00

In Table 3, the classification results of the applications we developed are given for Wine data. According to these results, the best results belong to the NB with 100% estimate accuracy. The performance order of other algorithms was DT, SVM, and KNN respectively. The parameters of the algorithms for Wine dataset:

- KNN: $p=2$, $n_neighbors=5$, $weights='uniform'$
- SVM: $kernel='poly'$, $gamma=0.0001$, $C=100$
- DT: $criterion='entropy'$, $max_features=11$, $min_samples_leaf=2$
- NB: Gaussian classifier, $priors='none'$

Table 4: Classification results for Glass data.

Algorithm	Accuracy (%)	Precision	Recall	F-Measure
KNN	70.93	0.71	0.71	0.70
SVM	70.93	0.69	0.71	0.67
DT	73.26	0.73	0.73	0.71
NB	32.56	0.39	0.33	0.32

In Table 4, the classification results of the applications we developed are given for Glass data. According to these results, the best results belong to the DT with 73.26% estimate accuracy. The performance order of other algorithms was KNN, SVM, and NB respectively. The parameters of the algorithms for Glass dataset:

- KNN: $p=2$, $n_neighbors=4$, $weights='distance'$
- SVM: $kernel='rbf'$, $gamma=0.8$, $C=0.8$
- DT: $criterion='gini'$, $max_features=5$, $min_samples_leaf=4$
- NB: Gaussian classifier, $priors='none'$

Table 5: Classification results for Heart Disease data.

Algorithm	Accuracy (%)	Precision	Recall	F-Measure
KNN	68.03	0.68	0.68	0.68
SVM	87.71	0.88	0.88	0.88
DT	84.43	0.84	0.84	0.84
NB	81.97	0.82	0.82	0.82

In Table 5, the classification results of the applications we developed are given for Heart Disease data. According to these results, the best results belong to the SVM with 87.71% estimate accuracy. The performance order of other algorithms was DT, NB, and KNN respectively. The parameters of the algorithms for Iris dataset:

- KNN: $p=2$, $n_neighbors=3$, $weights='uniform'$
- SVM: $kernel='linear'$, $C=1$
- DT: $criterion='entropy'$, $max_features=4$, $min_samples_leaf=3$
- NB: Gaussian classifier, $priors='none'$

V. CONCLUSION

In this study, first, we tested KNN, SVM, DT, and NB applications, which we had developed, on Iris, Wine and Glass data. The results we obtained were sufficient to confirm the accuracy of our applications' results when compared to the results in the literature. Then, we run these applications on the Heart Disease data resulting that SVM gave the best classification performance. In addition, DT and NB succeeded to produce results that can be used to classify these data. The performance of KNN in classifying these data is lower than the others. As a result, it is seen that the possibility of a person having heart disease can be determined by SVM, DT and NB algorithms.

The results indicate that these classification algorithms can be used in diagnosing of many diseases. Thus, treatments of many diseases can be started earlier, and human health can be better protected.

REFERENCES

- [1] Sisodia, D. and D.S. Sisodia, Prediction of diabetes using classification algorithms. *Procedia computer science*, 2018. 132: p. 1578-1585.
- [2] KARAKOYUN, M. and M. HACIBEYOĞLU, Biyomedikal Veri Kümeleri İle Makine Öğrenmesi Sınıflandırma Algoritmalarının İstatistiksel Olarak Karşılaştırılması. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 2014. 16(48): p. 30-42.
- [3] Özgür, A. and H. Erdem, Saldırı Tespit Sistemlerinde Kullanılan Kolay Erişilen Makine Öğrenme Algoritmalarının Karşılaştırılması. *Bilişim Teknolojileri Dergisi*, 2012. 5(2): p. 41-48.
- [4] Naik, A. and L. Samant, Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime. *Procedia Computer Science*, 2016. 85: p. 662-668.
- [5] Khanmohammadi, S. and M. Rezaeiahari, AHP based classification algorithm selection for clinical decision support system development. *Procedia Computer Science*, 2014. 36: p. 328-334.
- [6] Bhuvanawari, P. and A.B. Therese, Detection of cancer in lung with k-nn classification using genetic algorithm. *Procedia Materials Science*, 2015. 10: p. 433-440.
- [7] Deekshatulu, B. and P. Chandra, Classification of heart disease using k-nearest neighbor and genetic algorithm. *Procedia technology*, 2013. 10: p. 85-94.

- [8] Han, J., J. Pei, and M. Kamber, Data mining: concepts and techniques. 2011: Elsevier.
- [9] Bablani, A., D.R. Edla, and S. Dodia, Classification of EEG Data using k-Nearest Neighbor approach for Concealed Information Test. *Procedia computer science*, 2018. 143: p. 242-249.
- [10] Jung, M., O. Niculita, and Z. Skaf, Comparison of different classification algorithms for fault detection and fault isolation in complex systems. *Procedia Manufacturing*, 2018. 19: p. 111-118.
- [11] Anzid, H., et al., Multimodal Images Classification using Dense SURF, Spectral Information and Support Vector Machine. *Procedia computer science*, 2019. 148: p. 107-115.
- [12] Santosa, B., Multiclass Classification with Cross Entropy-Support Vector Machines. *Procedia Computer Science*, 2015. 72: p. 345-352.
- [13] Markovic, R., et al., Comparison of different classification algorithms for the detection of user's interaction with windows in office buildings. *Energy Procedia*, 2017. 122: p. 337-342.
- [14] Anyanwu, M.N. and S.G. Shiva, Comparative analysis of serial decision tree classification algorithms. *International Journal of Computer Science and Security*, 2009. 3(3): p. 230-240.
- [15] Kalıpsız, O. and P. Cihan, Öğrenci Proje Anketlerini Sınıflandırmada En İyi Algoritmanın Belirlenmesi. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 2015. 8(1): p. 41-49.
- [16] Hussain, J. and S. Lalmuanawma, Feature analysis, evaluation and comparisons of classification algorithms based on noisy intrusion dataset. *Procedia Computer Science*, 2016. 92: p. 188-198.
- [17] Rasjid, Z.E. and R. Setiawan, Performance comparison and optimization of text document classification using k-NN and naïve bayes classification techniques. *Procedia computer science*, 2017. 116: p. 107-112.
- [18] Dua, D.a.G., C. . UCI Machine Learning Repository. 2019; Available from: <http://archive.ics.uci.edu/ml>.
- [19] Markham, K. Simple guide to confusion matrix terminology. 2014; Available from: <https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/>.
- [20] Raschka, S., Confusion Matrix. 2018; Available from: http://rasbt.github.io/mlxtend/user_guide/evaluate/confusion_matrix/

ECG Signal Classification with Neural Network Ensemble

¹A.WANYENZE¹ and G. TEZEL¹

¹ Konya Technical University, Konya/Turkey, masabaa239@gmail.com

¹ Konya Technical University, Konya/Turkey, gtezel@ktun.edu.tr

Abstract - Electrocardiography being one of the most important diagnostic tools in the assessment of the heart functionality in medicine has attracted a lot of attention from various researchers and scholars. Neural networks have been and are still a booming area of research where machines are being trained to think like humans. In this study we employ an average weighted neural network ensemble in the classification of ECG signal beats into normal and abnormal. The individual networks in the ensemble are trained on the two different feature sets extracted from the same dataset with different training parameters.

Keywords - Ensemble learning, Ecg Signals, Classification, Ensemble neural network.

I. INTRODUCTION

AN ELECTROCARDIOGRAPHIC (ECG) signal is a biomedical signal that reflects the electrical manifestation of the contractile activity of the heart and can be recorded easily with surface electrodes placed on the limbs or chest of the specimen. It is one of the most well-known and used biomedical signals. It defines the graphical representation of the functioning of the human heart over a given period of time [1]. The electrocardiogram is used to investigate some types of abnormal heart functions including arrhythmias and conduction disturbances, as well as heart morphology (i.e., the orientation of the heart in the chest cavity, hypertrophy, and evolving myocardial ischemia or infarction). It is also useful for assessing performance of pacemakers [2]. Automatic arrhythmia detection and classification is so important in clinical cardiology when carried out in real time. This can be achieved through analyzing the ECG single through the extracted features [3].

Different methods and algorithms for automated analysis and classification of heart beats have been proposed. The common methods for ECG beat classification is Artificial Neural Networks (ANNs) that have shown accurate performance in different classification tasks. Multi-Layer Perceptions and Radial Basis Function networks have been of great interest[4, 5].

Combining neural classifiers is one way to improving the classification performance for complex problems [6, 7]. The neural ensemble classifiers are classification methods which train multiple neural classifiers and combine their prediction results to produce the final output [8]. An ensemble contains a number of classifiers normally called base classifiers. The

generalization ability of an ensemble is usually much stronger than that of base classifiers.

Ensemble learning is fascinating because it is able to boost weak classifiers which are slightly better than random guesses to strong classifiers which can make accurate predictions. Base classifiers are generated from training data by a base learning algorithm which can be decision tree, neural network or other kinds of machine learning algorithms. Most ensemble methods use a single base learning algorithm to produce homogeneous base classifiers, but some methods use multiple learning algorithms to produce heterogeneous classifiers [9, 10].

Machine Learning supports the idea that bringing together decisions of different estimators/ classifiers with a unique combinational structure can lead to substantial improvements in both training accuracy and learning generalization [11]. The ensemble neural network is a learning structure, where a collection of finite number of artificial neural networks is trained for the same task and put together taking the average of their predictions[12]. Xie et al developed a novel method for classifying melanocytic tumors as benign or malignant by the analysis of digital dermoscopy images. A neural network ensemble model was developed by combining BP neural networks with fuzzy neural networks to increase individual net diversity. In the experiments, feature extraction and reduction were verified and classification performance was tested using Feedforward Neural Network (FNN), Random forests, K-Nearest Neighbors (KNN), Gentle Adaboost, two Support Vector Machine (SVM) methods, and two systems using the BoF model, and the proposed meta-ensemble model on two datasets that respectively include xanthous race data and Caucasian race data and results showed that classification accuracy was greatly improved by the use the proposed classifier model[13].

In this study the use of ensemble neural classifiers for the classification of ECG signals is proposed with the aim of getting better and improved classification results with higher accuracy, reliability and generalization ability.

II. DATA AND FEATURE EXTRACTION

A. Dataset

For experiments and evaluation, the MIT-BIH ECG arrhythmia database available at MIT medical storage Physionet is used. This database consists of a 48, 30 minute ECG recordings each containing two (2) ECG lead signals

digitalized at 360 sampling rate with 11 bit resolution over 10 mV range. 23 of the recordings were randomly selected from a mixed population of in-patients and out-patients at the medical center. The remaining 25 recordings were selected from the same set but included less common clinically symbolic arrhythmias [14]. ECG data contains several kinds of irregularities such as power line interference, muscle artifacts, baseline wandering and motion artifacts and electromagnetic interferences and in order to solve this ,some preprocessing stages have to be undergone for better results[15]. ECG signal processing refers to the analysis, sampling, encoding and general manipulation of the ECG signals to access important information [16].

B. Feature Extraction

The most common features mentioned in literature related to ECG classification are the morphological and spectral features. Morphological features are peak values that is; P, Q, R, S, T and U. These are the main features of an ECG signal with the QRS complex being the highest peak (R peak). Different methods and algorithms for R peak detection such as Pan Tompkins algorithm, Hermit functions, wavelet functions, and frequency based functions, Hilbert transform, empirical mode decomposition etc. This ECG feature is easier to detect as compared to other portions of the signal due to its structural form and high amplitude. The R peak is very essential in the diagnosis of heart rhythm irregularities (arrhythmias) and also in determining Heart Rate Variability (HRV) [17].

Pan Tompkins algorithm has relatively simple filters, non-linear transformations and decision methods. The Pan Tompkins algorithm is sufficiently precise and possesses adequate sensibility values. The algorithm uses a band pass filter which sets a low pass filter (LPF) and a high pass filter (HPF) to reduce noise. This provides interference against signals outside the frequency band in which the QRS operates. Then a quadratic function is applied to the resulting signal of the derivative function to further enhance the high frequency characteristics of the QRS complex. Lastly, the energy estimate is made with a mobile window size of the longer QRS complex [18].

The low-pass filter is described by the formula:

$$y(n) = 2y(n-1) - y(n-2) + x(n) - 2x(n-6) + x(n-12) \quad (1)$$

The high-pass filter is given by:

$$y(n) = y(n-1) - 1/32 (x(n) + (n-16) - x(n-17) + 1/32 x(n-32)) \quad (2)$$

After filtering, the signal is differentiated to provide the QRS slope information using the following formula:

$$y(n) = 1/8 [2x(n) + x(n-1) - x(n-3) - 2x(n-4)] \quad (3)$$

The signal is squared point by point making all data points positive and emphasizing the higher frequencies:

$$y(n) = x^2(n) \quad (4)$$

The algorithm performs sliding window integration in order to obtain waveform feature information.

$$y(n) = 1/N [x(n-(N-1)) - (x(n-(N-2)) + \dots + x(n))] \quad (5)$$

where N is the size of the sliding window and depends on the sampling rate.[19]

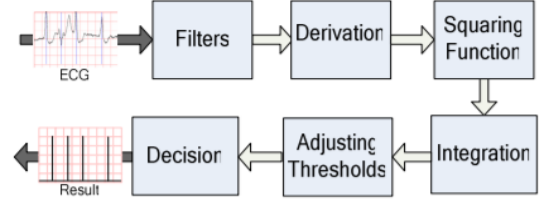


Figure 1: Pan Tompkins algorithm

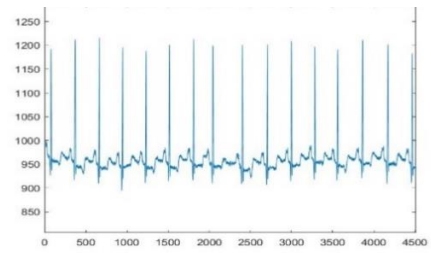


Figure 2: Raw ECG signal

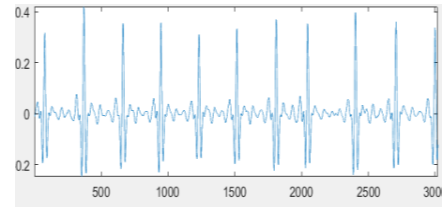


Figure 3: Band filtered ECG signal

Using the R peaks location results from the Pan Tompkins algorithm, the other features are extracted using the time domain approach.

R-R interval is calculated from: [20]

$$R-R \text{ int} = ((Rloc(n+1) - Rloc(n)) \div fs) \text{ sec} \quad (6)$$

fs = Sampling frequency

$Rloc$ = location of R peak

For P- wave

$$a = Rloc(i) - floor(0.2xfs) \quad (7)$$

$$b = Rloc(i) - floor(0.07xfs), \quad (8)$$

Where a and b are the p wave time domain window intervals. The maximum value between these values is considered to represent p wave.

For Q wave

$$a = Rloc(i) - floor(0.069xfs) \quad (9)$$

$$b = Rloc(i) \quad (10)$$

Where a and b are the Q wave time domain window intervals. The minimum value between these values is considered to represent Q wave.

For S wave

$$a = Rloc(i) \quad (11)$$

$$b = Rloc(i) - floor(0.07xfs) \quad (12)$$

The minimum value in this window is considered to represent S wave.

For T wave

$$a = Rloc(i) + floor(0.08xfs) \quad (13)$$

$$b = Rloc(i) + floor(0.4xfs) \quad (14)$$

The maximum value in this window represents the T wave.

The amplitudes of the waves were also calculated based on the mean value of the signal that is to say,

$$P_wave_amp = P_peak_value - mean_value$$

$$Q_wave_amp = Q_peak_value - mean_value$$

$$R_wave_amp = R_peak_value - mean_value$$

$$S_wave_amp = S_peak_value - mean_value$$

$$Twave_amp = T_peak_value - mean_value$$

III. NEURAL NETWORK ENSEMBLES

Neural networks are widely used for classification purposes and many factors influence the classification performance of these networks. Such factors are parameter settings, extracted features and the quality of experimental samples. Neural network ensembles where proposed by Hansen and Salamon as learning techniques that combine several individual neural networks that are trained for the same task with the purpose of improving the generalization performance of the Neural network ensemble as compared to single neural networks [9]. The use of ensemble neural network model resulted into improved classification performance and high confidence in the classification results while predictions from a single network classifier were “not confident”. Generally, network ensemble models achieve more reliable classification on non-representative objects [21]. This study covers homogeneous ensemble by taking a number of individual multilayer perception (MLPs) learners and their predications with an aim of improving the generalization performance.

Combining the predictions from multiple neural networks adds a bias that in turn counters the variance of a single trained neural network model [22]. A number of homogeneous multilayer perceptions are independently generated with the same architecture but different initial weights and biases trained using backpropagation algorithm with the same training samples and accuracy requirements. The designing of ensemble neural networks involves two processes which are developing an ensemble and determining the ensemble combiner. There are different kinds of ensembling models such as average weighted, boosting, bagging stacking etc. In stacking, an algorithm takes the outputs of sub-models as input and attempts to learn how to best combine the input predictions to make a better output prediction[23]. In a similar way, several models are trained separately and the average weights from each model contributes to the subsequent ensemble mode in this study. Each model of the ensemble is compiled and trained using the same dataset but different parameters as the base learners for the ensemble network.

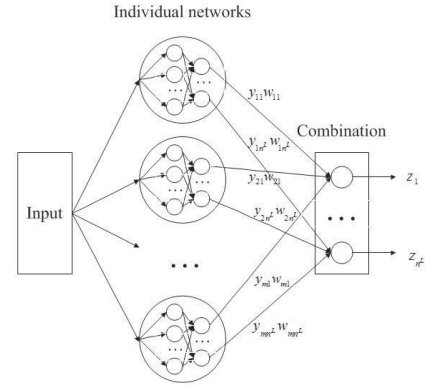


Figure 4. Ensemble neural network structure

The first model (model1) is made of three hidden layers with [64, 64, 32] neurons, ReLu activation function in two hidden layers and the third hidden layer employs the hyperbolic tangent activation function. The output layer employs softmax activation function. The second model (model2) and the third model (model3) consist of 2 hidden layers with [32, 16] [32, 9] neurons respectively. The activation functions used for the second model are ReLu in the first and second hidden layers of model 2. The third layer uses the ReLu in the first layer and the sigmoid function in the second hidden layer.

IV. RESULTS

In this study, two different feature sets were extracted from the ECG dataset, that is; morphological and spectral features. For both sets 75% of the data was used as training and 25% used for testing. The performance analysis of the ensemble network was estimated using the mean absolute error (MAE) for accuracy and the root mean square error (RMSE) as a loss function. The training and validation results of individual models in this study on the morphological and spectral feature sets are as shown in the tables below.

Table 1: Morphological feature set (MAE)

Model	Training Acc%	Testing Acc%
Model1	87.6	88.6
Model2	86.6	88.3
Model3	86.5	88.0
Ensemble		88.9

Table 2: Spectral feature set (MAE)

Model	Training Acc%	Testing Acc%
Model1	73.6	73.8
Model2	73.0	75.0
Model3	72.0	73.0
Ensemble		73.32

Table 3: Morphological feature set (RMSE)

Model	Training loss	Testing loss
Model1	0.108	0.099
Model2	0.106	0.099
Model3	0.099	0.097
Ensemble		0.099

Table 4: Spectral feature set (RMSE)

Model	Training loss	Testing loss
Model1	0.190	0.180
Model2	0.191	0.183
Model3	0.195	0.185
Ensemble		0.183

Results obtained here are part of an ongoing ECG signal classification using ensemble neural networks research. In future studies, we look forward to combining two or more models trained on different feature sets into a single ensemble model and analyse its classification performance.

REFERENCES

- [1] B. Subramanian, "ECG signal classification and parameter estimation using multiwavelet transform," 2017.
- [2] A. Gacek and W. Pedrycz, *ECG signal processing, classification and interpretation: a comprehensive framework of computational intelligence*. Springer Science & Business Media, 2011.
- [3] S. Osowski and T. H. Linh, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 11, pp. 1265-1271, 2001.
- [4] U. R. Acharya, P. S. Bhat, S. S. Iyengar, A. Rao, and S. Dua, "Classification of heart rate data using artificial neural network and fuzzy equivalence relation," *Pattern recognition*, vol. 36, no. 1, pp. 61-68, 2003.
- [5] Z. Wang, Z. He, and J. Z. Chen, "Blind EGG separation using ICA neural networks," in *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Magnificent Milestones and Emerging Opportunities in Medical Engineering* (Cat. No. 97CH36136), 1997, vol. 3, pp. 1351-1354: IEEE.
- [6] R. Ebrahimpour, E. Kabir, H. Esteky, and M. R. Yousefi, "A mixture of multilayer perceptron experts network for modeling face/nonface recognition in cortical face processing regions," *Intelligent Automation & Soft Computing*, vol. 14, no. 2, pp. 151-162, 2008.
- [7] R. Ebrahimpour, E. Kabir, H. Esteky, and M. R. Yousefi, "View-independent face recognition with mixture of experts," *Neurocomputing*, vol. 71, no. 4-6, pp. 1103-1107, 2008.
- [8] C.-Y. Chiu, B. Verma, and M. Li, "Impact of variability in data on accuracy and diversity of neural network based ensemble classifiers," in *The 2013 International Joint Conference on Neural Networks (IJCNN)*, 2013, pp. 1-5: IEEE.
- [9] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 10, pp. 993-1001, 1990.
- [10] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," in *Advances in neural information processing systems*, 1995, pp. 231-238.
- [11] D. S. Nascimento and A. L. Coelho, "Ensembling heterogeneous learning models with boosting," in *International Conference on Neural Information Processing*, 2009, pp. 512-519: Springer.
- [12] M. Madhwarasan and S. Deepa, "Application of ensemble neural networks for different time scale wind speed prediction," *neural networks*, vol. 4, no. 5, 2016.
- [13] F. Xie, H. Fan, Y. Li, Z. Jiang, R. Meng, and A. Bovik, "Melanoma classification on dermoscopy images using a neural network ensemble model," *IEEE transactions on medical imaging*, vol. 36, no. 3, pp. 849-858, 2017.
- [14] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45-50, 2001.
- [15] R. Acharya, S. M. Krishnan, J. A. Spaan, and J. S. Suri, *Advances in cardiac signal processing*. Springer, 2007.
- [16] A. Andreas, "Digital signal processing: Signals, systems, and filters," ed: McGraw-Hill, New York, 2006.
- [17] R. Mabrouki, B. Khaddoumi, and M. Sayadi, "R peak detection in electrocardiogram signal based on a combination between empirical mode decomposition and Hilbert transform," in *Advanced Technologies for Signal and Image Processing (ATSIP), 2014 1st International Conference on*, 2014, pp. 183-187: IEEE.
- [18] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. 32, no. 3, pp. 230-236, 1985.
- [19] C. Pavlatos, A. Dimopoulos, G. Manis, and G. Papakonstantinou, "Hardware implementation of Pan & Tompkins QRS detection algorithm," in *Proceedings of the EMBEC05 Conference*, 2005: Citeseer.
- [20] K. K. Patro and P. R. Kumar, "Effective Feature Extraction of ECG for Biometric Application," *Procedia Computer Science*, vol. 115, pp. 296-306, 2017.
- [21] V. Bukhtoyarov and E. Semenkin, "Neural networks ensemble approach for detecting attacks in computer networks," in *2012 IEEE Congress on Evolutionary Computation*, 2012, pp. 1-6: IEEE.
- [22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [23] H. Trevor, T. Robert, and F. JH, "The elements of statistical learning: data mining, inference, and prediction," ed: New York, NY: Springer, 2009.

Detection of Wart Treatment Method Using Machine Learning Algorithms

M. E. TENEKECI¹ and K. KARADAĞ²

¹ Harran University, Şanlıurfa/Turkey, etenekeci@harran.edu.tr

² Harran University, Şanlıurfa /Turkey, k.karadag@harran.edu.tr

Abstract - The warts seen in all body of the human are benign tumours. These warts are made and spread in humans' body by Human Papillomavirus (HPV). Since the treatment methods used differ from person to person, an exact method cannot be determined. The effectiveness of these methods varies depending on the characteristics of the patients and the warts. In this study, one of the methods of immunotherapy and cryotherapy will be determined as appropriate for the treatment of two common wart types, plantar and common. The data set used in the study was collected from 180 patients in the dermatology clinic of Ghaem Hospital in Iran. 90 of the patients were treated with immunotherapy and the remaining 90 patients were treated with cryotherapy. The results were recorded with patient and wart information. Artificial Neural Network (ANN), k-Nearest Neighbours (KNN) and Naïve Bayes (NB), which are methods of the machine learning algorithms, were used for determining the treatment method. Encouraging results have been achieved in the study. Thanks to this work, depending on the characteristics of the patient and wart, the appropriate treatment method for the specialist physician can be suggested. Thus, there will be no loss of time and money due to faulty treatment.

Keywords – Wart, Determination of treatment method, Machine Learning Algorithms.

I. INTRODUCTION

Wart disease is one of the skin health problems, which is usually characterized by the appearance of small lumps on the surface of the skin. This disease is caused by a virus called human papilloma virus (HPV) [1]. HPV viruses that cause warts can easily spread through direct contact. The immune system is very important for the infection of HPV. People with a weak immune system will be vulnerable to wart disease.

There are many different methods in the treatment of warts. However, none of these treatments is a definite solution. It varies according to the type of wart and the physiology of the patient. There are various methods such as surgical intervention, immunotherapy, laser therapy, cryotherapy, and intralesional injection [2]. In the treatment of warts, where the treatment method is found to be better, it is still unclear why even the most appropriate treatment for which patient will be better. [3]. Cryotherapy and immunotherapy are the most common methods for the treatment of warts. [4].

In this study, machine-learning algorithms will be used to

determine the efficiency of immunotherapy and cryotherapy methods in wart treatment. Immunotherapy, a technique used for the treatment of wart disease, strengthens the immune system to cope with wart disease [5]. Cryotherapy is a chemical burn for removing warts [6]. The application of these two treatment methods is shown in Figure 1.



(a)



(b)

Figure 1: Image of (a) Cryotherapy and (b) Immunotherapy Application [7,8].

Machine learning algorithms are used to analyze large datasets, find and extract knowledge from the data. The use of machine learning algorithms in the health industry is usually applied in diagnosing various diseases. Clinical predictions developed very rapidly by adopting computer science and information technology in processing health data.

In the study [9] used data mining algorithms for the first time in the selection of wart treatment. The characteristics that are important in the choice of treatment have been determined for both types of treatment. Apriori algorithm for determining the treatment method of the warts has issued rules. The obtained rules are optimized with fuzzy logic and the results are improved. In [10], the efficacy of immunotherapy in wart

treatment was determined. Fuzzy Logic and Artificial Neural Networks were used to estimate the outcome of the treatment. In the trials, the response of the patients to treatment was determined as 83.33% correctly. In study [11] the success of cryotherapy and immunotherapy treatment methods were estimated based on machine learning. Naive Bayes and KNN used classifiers as classifiers. The success of the treatment using the KNN has achieved an accurate estimate of 80% success rate. Study [12] compared the effect of cryotherapy and immunotherapy methods with Candida treatment. In [13], the determination of the performance of wart treatment methods using decision trees has been determined from fuzzy images. Accurate estimation rate was determined as 90% - 95%. This is quite high compared to previous studies.

In this work, Artificial Neural Network (ANN), k-Nearest Neighbor (KNN) and Naïve Bayes (NB) are data mining algorithm methods used to classify dataset and estimate new samples. The aim of this study was to compare the success of machine learning algorithms in determining the effectiveness of treatment methods for according to patient and wart characteristics, cryotherapy and immunotherapy methods are more appropriate. The flow diagram for the implementation of the proposed method is shown in Figure 2.

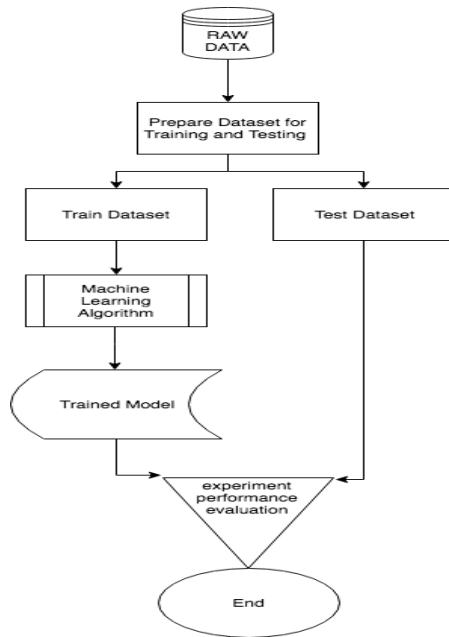


Figure 2: Blok diagram of the proposed model.

II. MATERIAL AND METHODS

2.1. Data Set

This study will be carried out effectiveness of treatment of the cryotherapy and immunotherapy method, which is the two most commonly, used wart treatment methods. Dataset Mentioned in [4, 9], the published UCI database [14] was taken. Data were collected from plantar and common warts

patients who admitted to the dermatology clinic of Ghaem Hospital in Mashhad, Iran from January 2013 to February 2015. Of the 180 patients, 90 were treated with cryotherapy and the remaining 90 were treated with immunotherapy. As shown in Table 1 and Table 2, there are 6 features in patients with cryotherapy method and 7 in patients with immunotherapy. It is also given last value as the result of whether or not the treatment responds.

Table 1: Cryotherapy Dataset Feature Explanation.

Attributes	Values	Comment
Gender (Man and Woman)	0-1	0 is woman – 47 1 is man - 43
Age(Year)	15-56	Age of Patient
Time elapsed before treatment (month)	0-12	
The number of warts	1-19	Count of Warts
Types of wart (count)	1,2,3	1-Common 2-Planatar 3-Both
Surface area of the warts (mm2)	4-750	Total surface area
Result	0-1	0 is Negative Response Treatment – 42 1 is Positive Response Treatment – 48

Table 2: Immunotherapy Dataset Feature Explanation.

Attributes	Values	Comment
Gender (Man and Woman)	0-1	0 is woman – 47 1 is man - 43
Age(Year)	15-56	Age of Patient
Time elapsed before treatment (month)	0-12	
The number of warts	1-19	Count of Warts
Types of wart (count)	1,2,3	1-Common 2-Planatar 3-Both
Surface area of the warts (mm2)	4-750	Total surface area
Result	0-1	0 is Negative Response Treatment – 42 1 is Positive Response Treatment – 48

2.2. Methods

2.2.1. Artificial Neural Network

Learning in biological systems is accomplished by the adjustment of synaptic connections between neurons. The learning process is done by using training samples for ANN. In the training process, weights of neurons are changed by processing input and output. This process is repeated until the error reaches a specified threshold value. AANs are mathematical systems consisting of many processing units (neurons) connected together in a weighted nodes. In general, the processing units correspond roughly to real neurons and

are interconnected in a network; this structure forms artificial neural networks. There are many different artificial neural networks as models [15].

2.2.2. K-Nearest Neighbor

K Nearest neighboring (KNN) algorithm performs classification according to the distance of the inter-sample properties. The distance between the properties of tested samples and the properties of the other samples in the data set is calculated. The class of the sample tested is determined by the classes to which the nearest k instance belongs [16]. This algorithm is very popular because it has a simple structure and does not require a learning phase and it gives successful results. [17].

2.2.3. Naïve Bayes

Naive Bayes Classifier is a probability-based classification process that uses the assumption that each attribute is statistically independent [18]. Probability theory uses in the classification of data. The assumption of a situation does not affect the occurrence of other conditions.

2.3. Evaluation Metrics

In the study, the performance of the proposed methods should be determined. For this reason, the experiments were performed by cross-validation method. The results are evaluated according to Precision, Recall and F-Measure values, which are expressed in Equations 1, 2 and 3. It uses the confusion matrix shown in Table 3 to obtain these values.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (1)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (2)$$

$$F - \text{Measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

Table 3: Confusion Matrix.

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

III. RESULTS

In the study, the results of cryotherapy and immunotherapy treatment methods were used. The success of the method for the treatment of warts depends on the physical conditions of the patient and the characteristics of the wart. The patient characteristics are very close to each other for selection of cryotherapy and immunotherapy. The aim of the study is to

analyze the patient characteristics with the help of these algorithms and to determine which treatment method is more appropriate. For this purpose, machine learning algorithms have been used. ANN, NB and KNN methods were used to estimate whether the treatment was successful or not.

The data sets used in the study were modeled with machine learning algorithms and the treatment method was determined according to patient characteristics. Data sets were performed for all data with 10-fold cross validation. The results were evaluated according to the performance metrics. In Table 4, the contrast matrix obtained in the KNN classification method for immunotherapy treatment is shown as an example.

Table 4: Sample Confusion Matrix for KNN (K=3) Classifier on Immunotherapy Treatment.

		Actual	
		Positive	Negative
Predicted	Positive	7	12
	Negative	11	60

In Table 5, the results of the experiments to determine the treatment with KNN are given. K values of 1, 3 and 5 were tried. As shown in Table 5, the most successful results were obtained by selecting K value 5.

Table 5: KNN Performance for Estimating Treatment Method.

	KNN								
	K=1			K=3			K=5		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Imm.	68,2	70	69	74	74,4	74,2	72,8	72,2	72,5
Cryo.	88,9	88,9	88,9	87	86,7	86,7	88,9	88,9	88,9
Mean	78,6	79,5	79,0	80,5	80,6	80,5	80,9	80,6	80,7

The results obtained with the KNN, ANN and NB algorithms are shown in Table 6. The most successful result for KNN is K = 5. According to the obtained results, the most successful prediction was obtained by ANN method. In ANN, a single secret layer with 5 neurons was used. Back propagation learning was performed with 500 iterations.

Table 6: Evaluation Performance of Machine Learning Algorithms for Estimating Treatment Method.

	KNN								
	K=1			K=3			K=5		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Imm.	72,8	72,2	72,5	71,9	74,4	72,9	72,9	74,4	73,6
Cryo.	88,9	88,9	88,9	83,6	83,3	83,2	91,2	91,1	91,1
Mean	80,9	80,6	80,7	77,8	78,9	78,1	82,1	82,8	82,4

IV. CONCLUSION

For the treatment of warts, a type of disease caused by viruses, a suitable method cannot be determined for all patients. The success of the treatment depends on the characteristics of the patients. In our study, machine learning algorithms were used to determine the treatment method according to patient characteristics. For the treatment methods used, the results of classification algorithms were analysed. It is observed that the highest success was achieved with ANN. 10-fold cross-validation was applied to 72.9% for Immunotherapy and 91.2% for Cryotherapy. The results show that these algorithms can help experts for selection method of wart treatment.

REFERENCES

- [1] Kardani K, Bolhassani A. Types of benign or malignant diseases associated with HPV infections. *HPV Infect Diagn Prev Treatment* 2018;30.].
- [2] M.M. Lipke, "An armamentarium of wart treatments", *Clin Med Res*, 4, 273-293, 2006.]
- [3] D. McGibbon, "Rook's Textbook of Dermatology (7th edition)", *Clin Exp Dermatol*, 31, 178-179, 2006.
- [4] Khozeimeh, F., Azad, F.J., Oskoue, Y.M., Jafari, M., Tehranian, S., Alizadehsani, R., Layegh, P., (2017). Intralesional immunotherapy compared to cryotherapy in the treatment of warts, *International Journal of Dermatology* 56: 474-478.
- [5] M.M. Clifton, S.M. Johnson, P.K. Roberson, J. Kincannon, T.D. Horn, "Immunotherapy for Recalcitrant Warts in Children Using Intralesional Mumps or Candida Antigens", *Pediatr Dermatol*, 20, 268-271, 2003.
- [6] Walczuk I, Eertmans F, Rossel B, Cegielska A, Stockfleth E, Antunes A, et al. Efficacy and safety of three cryotherapy devices for wart treatment: a randomized, control, investigator-blinded, comparative study. *Dermatol Therapy* 2017;1-14.
- [7] Nasser, N., "Treatment of common warts with the immune stimulant *Propionium bacterium parvum*". *An. Bras. Dermatol.* [online]. 2012, vol.87, pp.585-589
- [8] <https://www.scidacreview.org/cryotherapy-for-warts> (Access Date: 10/10/2018)
- [9] Khozeimeh, F., Alizadehsani, R., Roshanzamir, M., Khosravi, A., Layegh, P., Nahavandi, S. (2017). An expert system for selecting wart treatment method, *Computers in Biology and Medicine* 81: 167-175
- [10] Guimarães A., Araujo V., Campos S. P., Araujo V., Silva R. T., "Using Fuzzy Neural Networks to the Prediction of Improvement in Expert Systems for Treatment of Immunotherapy", *Advances in Artificial Intelligence*, 229 - 240, 2018, Springer
- [11] R. Uzun, Y. İşler and M. Toksan, "Choose of wart treatment method using Naive Bayes and k-nearest neighbors classifiers," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4.].
- [12] Majid, I., Imran, S.: Immunotherapy with intralesional *Candida albicans* antigen in resistant or recurrent warts: a study. *Indian J. Dermatol.* 58(5), 360 (2013)
- [13] Akben S. B., Predicting the success of wart treatment methods using decision tree based fuzzy informative images, *Biocybernetics and Biomedical Engineering*, Volume 38, Issue 4, 2018, 819-827, ISSN:0208-5216
- [14] UCI KDD Archive. [Online]. Available: <http://archive.ics.uci.edu/ml/>
- [15] Wu, X., V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A. Gb. Liu & S.T. Philip, 2016. Top 10 Algorithms in Data Mining, *Knowl inf. Syst.* 14(1)(2016)108-118.
- [16] T.M. Cover, P.E. Hart, "Nearest neighbor pattern classification", *IEEE Trans Inf Theory*, IT13(1):21-27, 1967
- [17] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, 2nd Edition, 2000.
- [18] A.Y. Ng, M.I. Jordan, "On discriminative versus generative classifiers: a comparison of logistic regression and naive Bayes", *Adv. Neural Inform. Process. Syst.*, 14 (2001), pp. 605-610

Detecting Student Engagement in e-learning Environment Based on Head Pose Estimation and Eye Aspect Ratio

M. U. UÇAR¹ and E. ÖZDEMİR²

¹ Havelsan A.Ş., Konya/Turkey, mucar@havelsan.com.tr

² Iskenderun Technical University, Hatay/Turkey, ersin.ozdemir@iste.edu.tr

Abstract – In this paper, we explored the feasibility of detecting student engagement in e-learning environment based on Head Pose Estimation and Eye Aspect Ratio of the student. 1000 video frames were extracted from UPNA Head Pose Database and used as a dataset for training and testing. Euler angles (pitch, yaw and roll) of the head were calculated based on Levenberg-Marquardt optimization. We constructed our feature vector by utilizing Euler angles and Eye Aspect Ratio for each frame. In addition to these behavioral features, each frame was annotated as “Engaged” or “Not Engaged” by five labelers. Engagement classifiers were trained with the labelled dataset. Experiments were made with Random Forest, Decision Tree, SVM and KNN machine learning algorithms to classify student engagement. SVM classifier achieved 72.4% accuracy on the labelled dataset. Our study shows that Head Pose and Eye Aspect Ratio are significant contributors to the student’s Visual Focus of Attention and engagement.

Keywords – engagement detection, head pose estimation, eye aspect ratio, computer vision, machine learning.

I. INTRODUCTION

Along with the technological developments, it has become feasible to process big data such as videos in real-time.

Software libraries that consists of the state-of-the-art computer vision and machine learning algorithms such as OpenCV [1] and dlib [2] has accelerated the studies in image processing field.

The growing importance of Human-Computer Interaction and the introduction of smart devices in every aspect of our lives have brought the necessity of gaining humanlike skills to the smart devices.

Computer vision is one of these leading skills. One of the most important and common application of computer vision is face recognition. In contrast to face recognition, Head Pose Estimation has recently started to get more attention from researchers.

Although the studies related with image processing in the field of education are not very common [3], it is observed that they have increased in recent years.

Since engagement has a vital role for efficiency and success of a course regardless of whether it is a traditional classroom or e-learning course, learner engagement detection has been one of the main research topics in this field.

In traditional classrooms, teachers observe their students and adjust the teaching activity according to the behavioral and emotional state of the students. However, it is not feasible for a teacher to track each one of his/her students at the same time. Therefore, providing a personalized learning environment for each student is not a realistic expectation.

In order to automate personalization process for each student and improve learning and teaching quality, computer vision and machine learning techniques are mostly used in recent days.

In this study, we try to address engagement detection problem and uncover the contribution of Head Pose Estimation and Eye Aspect Ratio to the student’s engagement. Our main goal is to find out a way for automatically detecting engagement by capturing and processing student’s video in real time with the aid of a web camera, a computing device and advanced computer vision and machine learning technologies.

In the next section, we will review related studies in this field. Then we will talk about our methodology and come up with a cost-effective and simple engagement classification model. Later, we will present the results of our experiments to reveal the performance of the proposed solution. Finally, we will sum up our findings and propose some improvements for future work.

II. RELATED WORK

Australian Council of Educational Research defined student engagement as “*students’ involvement with activities and conditions likely to generate high quality learning*” [4]. According to Kuh (2008):

“*Student engagement represents both the time and energy students invest in educationally purposeful activities and the effort institutions devote to using effective educational practices.*” [5].

Student engagement is a significant contributor to the student success. Several researches made with students reveal the strong relation between student engagement and academic achievement [6, 7]. Similarly, Trowler who has published a comprehensive literature review on student engagement states that a remarkable amount of literature has established robust correlations between student involvement and positive outcomes of student success and development [8].

As cited in Fredricks et al., student engagement has three

dimensions: *behavioral*, *cognitive* and *emotional* engagement [9]. Behavioral engagement refers to student's participation and involvement in learning activities. Cognitive engagement corresponds to student's psychological investment on the learning process such as being thoughtful and focusing on achieving goals. On the other hand, emotional engagement can be understood as student's relationship with his/her teacher and friends; feelings about learning process such as being happy, sad, bored, angry, interested and disappointed.

In this study, we focus on behavioral engagement, since its symptoms, such as drowsiness, looking at an irrelevant place and focusing on the screen can be detected explicitly.

Head pose, gaze and facial expressions are main contributors to behavioral and emotional engagement. Automated engagement detection studies mostly based on these features.

Vertegaal et al. presented an experiment with eye-trackers attached to the subjects during a conversation and concluded that gaze is an excellent predictor of conversational attention [10].

Stiefelhagen and Zhu defined gaze as the direction where the eyes are pointing in the space. They also point out that this direction has a close relation with the Visual Focus of Attention (VFoA) and it is sum of head and eye orientation. In their study, they made an experiment to discover the contribution levels of head and eye orientation to the VFoA. As a result, they found that head pose contributes 68.9% to the overall gaze and 88.7% to VFoA estimation [11]. The results show us head pose is a good indicator for VFoA and student engagement detection.

Aslan et al. measured student engagement in 1:1 learning environment with 85% accuracy. They used multiple sensors: (1) 2D camera for facial feature detection and analyzing learner's emotional state; (2) external eye tracking module to detect student's gaze and region of interest on the screen; (3) 3D camera (Intel's RealSense™ 3D camera and RealSense™ SDK) to record depth data corresponding to each pixel of 2D camera. Utilizing a 3D camera, they could extract more precise facial landmark data resulting more accurate emotion recognition, head and body pose detection. They collected data from high school students. Three human labelers annotated collected data as "Engaged" or "NotEngaged". They used labelled data to train and test their Decision Tree, Random Forest and Naïve Bayes classifiers [12].

Nezami et al. proposed a deep learning model to improve engagement detection for the face images in the Facial Expression Recognition-2013 (FER-2013) dataset. Their model is trained with 4627 engaged and disengaged samples. On the test set, proposed model achieves 72.38% classification accuracy [13].

Bosch described results from several prior works that utilize facial features to detect student engagement. He proposed a new method based on Mind Wandering detection using face-based techniques to improve engagement accuracies [14].

Whitehill et al. explored methods for automatic student engagement recognition from facial expressions. They used machine learning to automate process and found that engagement detection of 10-second videos can be reliably predicted by using average label of the frames in video. Finally,

they showed the correlation between human and automatic engagement results [15].

In addition to behavioral engagement studies, some researchers focused on student's emotional engagement. Ayvaz and Gürüler developed an application to detect student's emotional state based on facial expressions. They created a dataset from 11680 emotional images of 12 student (5 female and 7 male) and achieved 97.15% accuracy with SVM to classify student's emotional state [16].

Asteriadis et al. aimed a user-independent, non-intrusive engagement detection method, which is also resistant to different lighting conditions, and needs only a computing device, and a web-camera. They tracked head pose and eye gaze to classify user's attention and frustration. They achieved above 80% accuracy with Neuro-Fuzzy system [17].

III. METHODS

A. Dataset and Labelling

In order to classify the student engagement, we employed several machine-learning algorithms that required labelled data in training dataset for supervised learning.

We created student engagement dataset by using *UPNA Head Pose Database* [18] that is suitable for our needs. This database is publicly available for research purposes and created specifically for head pose estimation. There are totally 120 videos in the database for 10 different persons (4 female and 6 male) and 12 videos for each person. The videos are at 30 fps in MPEG-4 format with a resolution of 1280×720 pixels and each video is 10 seconds long. In each video, subjects moved their head combining translations and rotations along the three spatial axes of head. Actual head pose angles and head position data for each frame were measured with a sensor attached to the subject's head. These ground-truth measurements are shared as a text file with the database.

We extracted random 100 frames for each person from UPNA Head Pose Database and created a dataset that consists of totally 1000 images. Five human labelers annotated each image as (0)-"Not Engaged" or (1)-"Engaged". We measured consistency between the labelers based on *Fleiss' kappa measure* [19], which is a statistical approach for assessing the inter-rater reliability of more than two raters. Percent agreement and Fleiss' kappa values of our five labelers are 0.95 and 0.85 respectively. These values indicate there is strong agreement between labelers. Then, we labelled each image as majority decision of our labelers.

Sample images in our dataset are shown below in Figure 1.



Figure 1. Sample images in the dataset.
(First row: Engaged, Second row: Not Engaged)

B. Face and Facial Landmark Detection

We detected the face in the image by utilizing the Histogram of Oriented Gradients algorithm of Dalal and Triggs [20], which is already implemented in Dlib library.

After face detection, facial landmarks are extracted by using Dlib's implementation of the Kazemi and Sullivan's paper [21]. Detected 68 facial landmarks are illustrated in Figure 2.

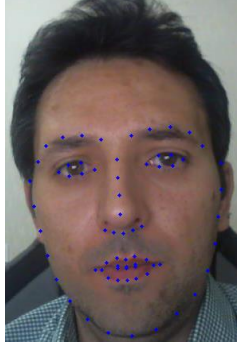


Figure 2. Facial landmarks

C. Head Pose Estimation

We used built-in *solvePnP* function in OpenCV library to estimate head pose. This function solved Perspective-n-Point problem by using Levenberg-Marquardt optimization [22]. As cited in Zheng et al., Perspective-n-Point is a problem of estimating the pose of a calibrated camera given a set of n 3D ($n \geq 3$) point coordinates in the world and their corresponding 2D projections in the image [23]. We used 2D facial landmark coordinates for 6 points (*nose tip, chin, left corner of left eye, right corner of right eye, mouth's left corner and mouth's right corner*) and their corresponding approximate 3D point coordinates as Mallick suggested [24]. Then, we got corresponding rotation and translation vector from *solvePnP* function. We transformed the rotation vector to a rotation matrix by using Rodrigues' rotation formula [25]. Finally we computed Euler angles (pitch, yaw and roll) from the rotation matrix as described in Slabaugh's paper [26]. Euler angles are illustrated in Figure 3.

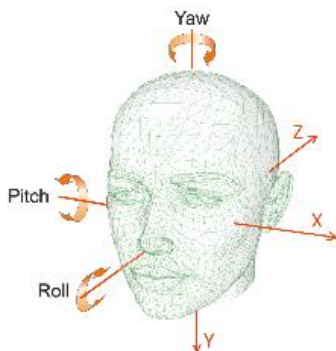


Figure 3. Euler angles

D. Eye Aspect Ratio (EAR)

In addition to Euler angles, another significant contributor to the student engagement is EAR. Soukupová & Cech used EAR value for real-time eye blink detection successfully [27]. Similar to eye blink detection, student's drowsiness can be

detected by using EAR value. When student is sleepy or closed his/her eye, EAR value is minimized. Facial landmarks used to calculate EAR value are shown below in Figure 4.

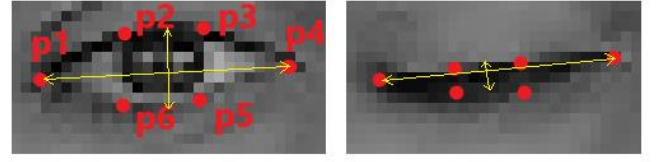


Figure 4. Opened and closed eyes with related facial landmarks

We calculated EAR value based on Euclidean distance formula by using facial landmark coordinates in the eye region as shown in Equation 1 below:

$$EAR = \frac{\|p2 - p6\| + \|p3 - p5\|}{2 \times \|p1 - p4\|} \quad (1)$$

E. Classification

We constructed our feature vector with four elements that consists of Pitch, Yaw, Roll and EAR values for each frame in our previously labelled dataset. We used half of our dataset for training purpose and the other part for testing. We trained SVM, KNN, Random Forest and Decision Tree classifiers with training set. Thanks to these classifiers, we were able to classify a student image captured from camera as "Engaged" or "Not Engaged" in real-time. Our student engagement classification software runs as a standalone application in student's computer during he/she takes an e-learning course. Screenshot of our application is shown in Figure 5.

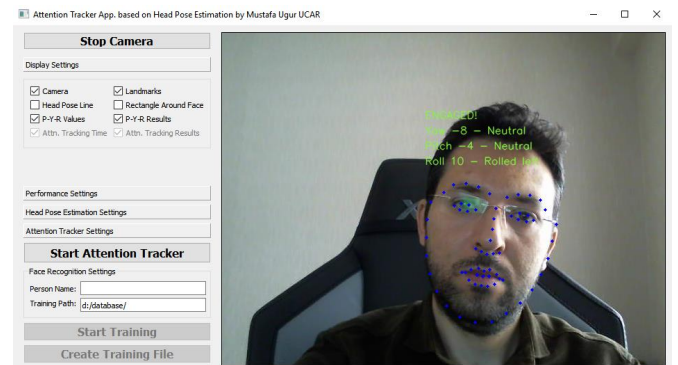


Figure 5. Screenshot of our application

IV. RESULTS

In order to analyze achievement of the student engagement detection system, we performed two experiments.

In the first experiment, we aimed to measure the mean absolute deviation of Euler angles. For this purpose, we processed 120 videos of 10 users in UPNA Head Pose Database. Pitch, yaw and roll angles for each frame is calculated and compared with the ground-truth pitch, yaw and roll angles, which are provided by UPNA Head Pose Database. The mean absolute deviation from ground-truth data in pitch, yaw and roll angles are 1.3°, 4.9° and 4.3° respectively and details are provided in Table 1.

Table 1. Mean Absolute Deviations of Euler Angles

	Pitch (°)	Yaw (°)	Roll (°)
User_01	1.656645	2.685429	1.94371
User_02	1.109542	2.396021	1.527393
User_03	1.095765	5.770226	2.941756
User_04	1.402484	3.083149	1.954714
User_05	1.245349	6.498999	4.362588
User_06	1.768564	4.561221	4.521068
User_07	1.247303	4.029759	2.855128
User_08	1.33414	9.503025	6.241764
User_09	0.968213	6.234931	5.708336
User_10	1.590701	4.939992	11.44409
Deviation	1.341871	4.970275	4.350055

In the second experiment, we aimed to measure the overall success of student engagement classification. SVM, KNN, Random Forest and Decision Tree student engagement classifiers, which we trained previously, are tested with the training part of our dataset. 72.4%, 71.6%, 70.6% and 70% accuracies are achieved using SVM, KNN, Random Forest and Decision Tree classifiers respectively as shown in Table 2.

Table 2. Overall Success of the Classifiers

	Accuracy (%)
SVM	72.4
KNN	71.6
Random Forest	70.6
Decision Tree	70.0

Additionally, we conducted second experiment with a new dataset that is created by using ground-truth Euler angles, and calculated EAR value. 78.6%, 77.8%, 77% and 76.4% accuracies are achieved using Random Forest, Decision Tree, KNN and SVM classifiers respectively as shown in Table 3.

Table 3. Overall Success of the Classifiers (With Ground-truth Data)

	Accuracy (%)
SVM	76.4
KNN	77.0
Random Forest	78.6
Decision Tree	77.8

V. SYSTEM CONFIGURATION

Our system configuration are shown in Table 4.

Table 4. System Configuration

Processor	Intel(R) Core(TM) i7-4770K CPU @ 3.50GHz
Graphics Card	Gigabyte NVIDIA GeForce GTX 980 4 GB
RAM	16 GB
Webcam	Logitech HD Webcam C525
OS	Windows 10 Pro 64-bit
IDE	Qt Creator 4.2.1
Libraries	OpenCV (version 3.3.1), dlib (version 19.8)

VI. CONCLUSION

We have proposed a methodology to detect student engagement in e-learning environment based on Head Pose Estimation and EAR. We used UPNA Head Pose Database to produce our dataset and five human labelers annotated each image in the dataset with high inter-rater reliability (0.85 Fleis' kappa). We have illustrated that student engagement can be detected by machine learning models with 72.4% accuracy.

Our application is based on computer vision techniques and required a simple webcam. This method is more cost-effective, simple and applicable than sensor-based methods, which are also expensive and intrusive. However, sensor-based methods provide more accuracy. There is a natural trade-off between high accuracy and non-intrusiveness.

In future work, in order to increase student engagement accuracies, researchers might consider other contributors to the student engagement such as gaze, gestures and facial expressions as well as Head Pose Estimation and EAR. Additionally, utilizing more extensive dataset might boost classification accuracies.

REFERENCES

- [1] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, vol. 25, pp. 120-125, 2000. Available: <https://opencv.org>
- [2] D. E. King, "Dlib-ml: A Machine Learning Toolkit," Journal of Machine Learning Research 10, pp. 1755-1758, 2009. Available: <http://dlib.net>
- [3] Ö. Sümer, P. Goldberg, K. Stürmer, T. Seidel, P. Gerjets, U. Trautwein, and E. Kasneci, "Teachers' Perception in the Classroom," CVPR Workshops 2018, pp. 2315-2324.
- [4] H. Coates, "Engaging Students for Success: 2008 Australasian Survey of Student Engagement," Victoria, Australia, Australian Council for Educational Research, 2008.
- [5] G. Kuh, T. Cruce, R. Shoup, J. Kinzie, and R. Gonyea, "Unmasking the effects of student engagement on first year college grades and persistence," Journal of Higher Education, vol. 79 (5), pp. 540-563, 2008.
- [6] S. Günüş, "The relationships between student engagement and their academic achievement," International Journal on New Trends in Education and Their Implications (IJONTE), vol. 5, pp. 216-231, 2014.
- [7] M. J. Casuso-Holgado, A. I. Cuesta-Vargas, N. Moreno-Morales, M. T. Labajos-Manzanares, F. J. Barón-López, and M. Vega-Cuesta, "The association between academic engagement and achievement in health sciences students," BMC medical education, 2013.
- [8] V. Trowler, "Student Engagement Literature Review," York, Higher Education Academy, 2010.
- [9] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: potential of the concept, state of the evidence," in Review of Educational Research, vol. 74(1), pp. 59-109, 2004.
- [10] R. Vertegaal, R. Slagter, G. van der Veer, and A. Nijholt, "Eye Gaze Patterns in Conversations: There is more to conversational agents than meets the eyes," Proceedings of CHI 2001, Seattle, WA, vol. 3(1), pp. 301-308, 2001.
- [11] R. Stiefelhofen, and J. Zhu, "Head Orientation and Gaze Direction in Meetings," CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02), ACM, New York, NY, USA, pp. 858-859, 2002.
- [12] S. Aslan, Z. Cataltepe, I. Diner, O. Dunder, A. A. Esme, R. Ferens, G. Kamhi, E. Oktay, C. Soysal, and M. Yener, "Learner Engagement Measurement and Classification in 1:1 Learning," Proceedings - 2014

- 13th International Conference on Machine Learning and Applications, ICMLA 2014, pp. 545-552, 2014.
- [13] O. M. Nezami, L. Hamey, D. S. Richards, and M. Dras, "Engagement Recognition using Deep Learning and Facial Expression," eprint arXiv:1808.02324, 08/2018.
 - [14] N. Bosch, "Detecting student engagement: Human versus machine," In Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization, UMAP '16, pp. 317–320, New York, NY, USA, 2016.
 - [15] J. Whitehill, Z. Serpell, Y.C. Lin, A. Foster, and J. Movellan, "The faces of engagement: Automatic recognition of student engagement from facial expressions," IEEE Transactions on Affective Computing vol. 5, no. 1, pp. 86-98, 1 Jan.-March 2014.
 - [16] U. Ayvaz, and H. Gürüler, "Real-time detection of students' emotional states in the classroom," 25th Signal Processing and Communications Applications Conference (SIU), 2017.
 - [17] S. Asteriadis, K. Karpouzis, and S. Kollias, "Feature extraction and selection for inferring user engagement in an HCI environment," in Human-Computer Interaction, New Trends, Springer Berlin Heidelberg, pp. 22-29, 2009.
 - [18] M. Ariz, J. J. Bengoechea, A. Villanueva, and R. Cabeza, "A novel 2D/3D database with automatic face annotation for head tracking and pose estimation," Computer Vision and Image Understanding, vol. 148, pp. 201-210, ISSN 1077-3142, July 2016.
 - [19] K. L. Gwet, *Handbook of Inter-Rater Reliability*. 2nd ed., Gaithersburg: Advanced Analytics, LLC ISBN 978-0-9708062-2-2, 2010.
 - [20] N. Dalal, and B. Triggs, "Histograms of Oriented Gradients for Human Detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1, vol. 1, pp. 886-893, 2005.
 - [21] V. Kazemi, and J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees," 2014 IEEE Conference on Computer Vision and Pattern Recognition pp. 1867-1874, 2014.
 - [22] J. J. Moré, "The Levenberg-Marquardt algorithm: Implementation and theory," In: A. G. Watson (eds) Numerical Analysis. Lecture Notes in Mathematics, vol. 630. Springer, Berlin, Heidelberg, 1978.
 - [23] Y. Zheng, Y. Kuang, S. Sugimoto, K. Åström, and M. Okutomi, "Revisiting the PnP Problem: A Fast, General and Optimal Solution," Proceedings of the IEEE International Conference on Computer Vision, pp. 2344-2351, 2013.
 - [24] S. Mallick, "Head Pose Estimation using OpenCV and Dlib," Learn OpenCV, 26.09.2016. Available: <http://www.learnopencv.com/head-pose-estimation-using-opencv-and-dlib/>
 - [25] L. Sörgi, "Two-view geometry estimation using the Rodrigues rotation formula," In 2011 18th IEEE International Conference on Image Processing, pp. 1009-1012. IEEE, 2011.
 - [26] G. G. Slabaugh, "Computing Euler angles from a rotation matrix," 1999. Available: <http://www.gregslabaugh.net/publications/euler.pdf>
 - [27] T. Soukupová, and J. Cech, "Real-Time Eye Blink Detection using Facial Landmarks," 21st Computer Vision Winter Workshop, Rimske Toplice, Slovenia, February 3–5, IEEE. 2016.

Functional Brain Network Analysis Under Cognitive Task

S.DEMİR¹ and İ.TÜRKER²

¹ Karabuk University, Karabuk/Turkey, saitdemir@karabuk.edu.tr

² Karabuk University, Karabuk/Turkey, iturker@karabuk.edu.tr

Abstract – Functional Brain Networks (FBN), derived from electroencephalogram (EEG) signal records, provide a framework to analyze connectivity between brain regions by the means of complex network measures. To investigate evolution of FBN under a cognitive task, we performed an analytical study based on EEG signals recorded from 36 healthy volunteers of matched age, which are available through Physiobank platform. We generated adjacency matrices with edge weights indicating correlations between EEG signals from each electrode, using the phase coherence method. We applied gradually increasing thresholds to these edge weights to achieve unweighted networks. As a result, we uncovered that under resting conditions, FBNs of female individuals have slightly greater strength, network density and clustering, pairing with lower path length. On the other hand, all this scheme reverses under a cognitive task, putting forward FBNs of male individuals as more connected ones. Moreover, regardless from gender, FBNs of successful individuals in cognitive task are slightly more connected compared to unsuccessful ones.

Keywords – Functional brain networks, complex networks, EEG, signal processing, phase coherence.

I. INTRODUCTION

NETWORK science provides a useful framework for understanding systems composed of many interconnected components. Collective behavioral dynamics are better uncovered by evaluating such systems as networks consisting of numerous nodes interacting with each other. By the way, we can now understand how human jointly act in social media or how magnetism emerges from the collective behavior of millions of spins [1].

Human brain is also a complex system with distinct regions interacting with each other. To analyze the anatomical or functional connectome of brain, graph-theoretic approach attracts increasing attention [2].

To construct a functional brain network, signals recorded via electroencephalography (EEG), magneto cephalography (MEG), or magnetic resonance imaging (MRI) are often used [3, 4]. Based on these recordings, electrodes or sensors corresponding to specific anatomic regions serve as nodes of the network, where coherent behaviors between each node pairs indicate edge weights. As a result, networks are constructed by thresholding these weights and an unweighted and undirected scheme is provided [5]. In this scheme without loss of generality, connections having weights greater than a pre-determined threshold have weights 1 and low-weight

connections are set to 0 [6].

Analysis of such networked structure provides insight into neurological and psychiatric diseases like schizophrenia, Alzheimer's disease, dementia, each of which can be probed by changes in brain functional networks [7, 8]. Along with these developments in exploring brain functions, human brain remains as the most complex structure for today's science, composing of billions of neurons and synapses. The full-resolution mapping of brain, named as connectome, remains as a big challenge together with some recent advancements. However, low-level projections to connectome like functional networks are rather popular approaches [9].

In this context, we can discriminate brain connectivity studies into three mainstream fields. First is anatomical or structural connectivity, involving high resolution connectome composed of neurons, while second is functional connectivity with lower resolution, in which regions of brain are treated as nodes [10]. The third is the effective connectivity, referring to the inference between two or more neural systems, either at a synaptic or population level [11]. Recent studies outline that structural and functional connectivity are characterized by common features, while functional connectivity is also correlated with the structural one even when the brain activity evolves in the resting state [12, 13]. Beyond the resting state characteristics, effect of sequential cognitive tasks over functional connectivity remains as a warm topic of interest, involved by some studies with network science tools [12, 14].

Incorporating network science approach with signal processing, a portion of studies outline that brain is a hierarchical and complex organ with small-world, scale-free and modular characteristics. Aiming to explore, to what extent functional brain network architecture is correlated to cognitive processes, they uncovered that there exist correlation between short characteristic path length and higher IQ [15].

Our study aims to provide an insight into evolution of network metrics under a cognitive task with respect to resting state. EEG signals provided from Ref. [16] are processed to achieve network representations. Since success of this cognitive (specifically arithmetic) task together with genders of the volunteers are also captured during EEG recordings, we derived plots presenting evolution of network parameters under various threshold levels, with respect to gender and success of the volunteers. The rest of the paper is organized as: Section II: Data and Methods, Section III: Results, and Section IV: Discussion.

II. DATA AND METHODS

The dataset used in this study is provided by Zyma et al. and freely retrieved from Ref. [16], which is available under PhysioBank database [17]. Dataset consists of EEG recordings collected from 36 volunteers using a 23-channel system, under resting and arithmetic task conditions. The positions of the EEG electrodes are based on the International 10/20 system, visualized as in Fig. 1.

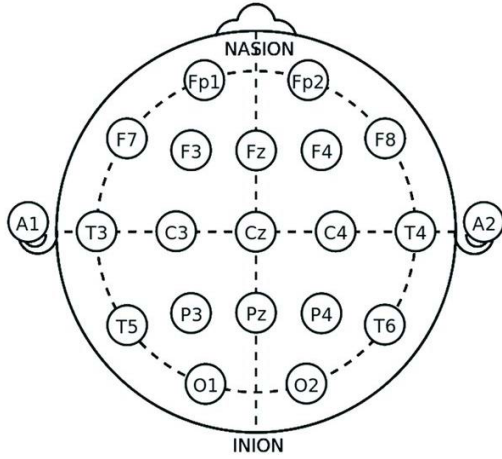


Figure 1: 10/20 EEG placement protocol [18].

During EEG recording, volunteers are subjected to a serial mental arithmetic task consisting of serial subtraction of two numbers. Each trial started with prompting to subtract 4-digit and 2-digit numbers (e.g., 1572 and 56). The number of operations completed for 4 minutes and correctness of the calculations resulting a mental arithmetic score was captured for each participant. According to this score, they were labelled as successful or not, subject to this mental process. Rest of the

details about EEG recording protocol can be accessed from Ref. [16].

We used beta band of EEG signals, involving 12-30 Hz interval of frequency spectrum, which is most involved with conscious focus, memory and problem solving activities of brain [19]. To achieve a projection from EEG signals to adjacency matrices of networks we used HERMES, a toolbox for MATLAB environment designed to study brain connectivity from neurophysiological data [10].

We employed coherence method, which depends on detecting how similar are the waveforms of each frequency after a various time-lag is applied to one of the signals, quantified by a cross-correlation function [15]. It measures the linear correlation between two variables $x(t)$ and $y(t)$ as a function of the frequency, f . It is the squared module of the coherency function (K), which is the ratio between the cross power spectral density, $S_{xy}(f)$, between $x(t)$ and $y(t)$, and their individual power spectral densities $S_{xx}(f)$ and $S_{yy}(f)$:

$$K_{xy}(f) = \frac{S_{xy}(f)}{\sqrt{S_{xx}(f)S_{yy}(f)}} \quad (1)$$

Once we achieved adjacency matrices holding cross-correlations of signals from channels, we performed network analysis using the Brain Connectivity Toolbox for MATLAB introduced by Rubinov and Sporns, containing tools for complex network measures and large-scale neuroanatomical connectivity datasets [20].

III. RESULTS

Adjacency matrices including cross-correlations between recordings from EEG nodes, form the basis for all results presented this section. Therefore, we start with presenting visualization of these matrices regarding genders, success status and rest/task status in Fig. 2.

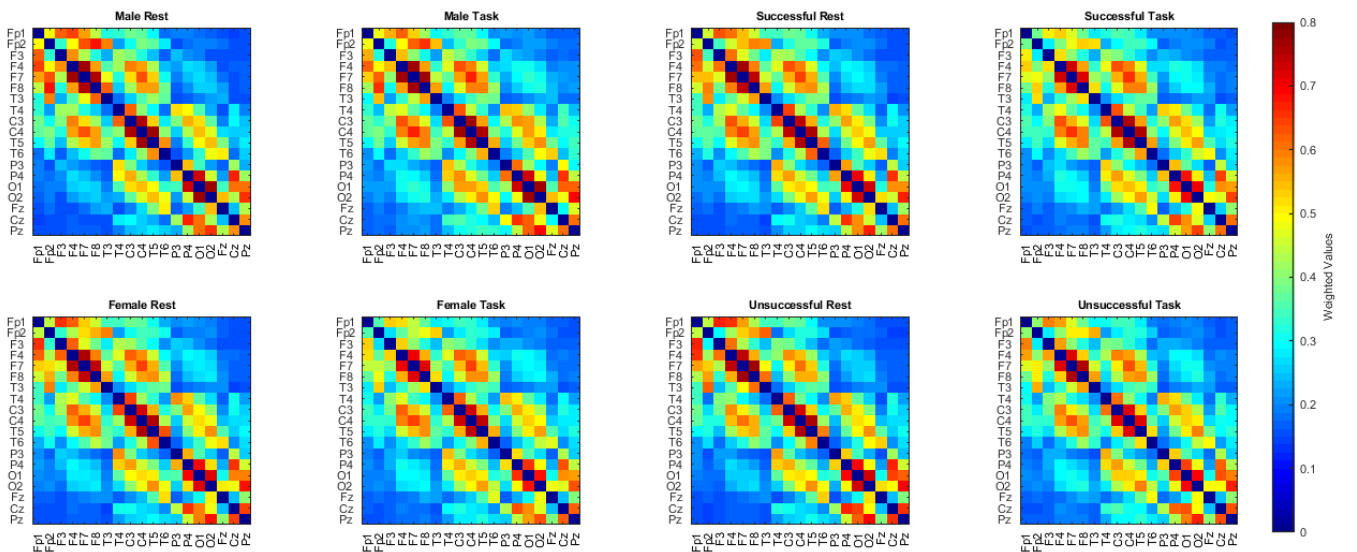


Figure 2: Adjacency matrices regarding male/female, rest/task and successful/unsuccessful status of the individuals

As seen in Fig. 2, the most dominant characteristics emerge as strong connections between spatially close nodes. This

property can be traced along the diagonal cells of matrices as lighter colors. Beside the diagonals, there also exists some strong clusters as between F4, F7, F8 and C3, C4, T5 modules. Discriminating the 8 figures seems a bit difficult since they have very similar cell values, whereas commenting on mean values (μ) is more convenient.

Matrices indicate that mean weight of brain network increases under cognitive task for male individuals compared to resting condition (from 0.3826 to 0.4051), while it decreases under cognitive task for female individuals (from 0.3931 to 0.3826). On the other hand, independent from gender, μ decreases under cognitive task for successful individuals (from 0.3939 to 0.3884), while it increases under cognitive task for unsuccessful individuals (from 0.3814 to 0.3846). we can conclude these results as brain functional network characteristics under cognitive task alters for gender, getting more connected for male and less connected for female individuals. Similar with female individuals, networks of successful individuals get less connected under cognitive task.

We also present network presentation of successful individuals under task condition with electrode locations, after a moderate weight threshold ($\tau=0.5$) is applied, in Fig. 3. Strong connections between frontal (F) lobe nodes are clear, while most of the nodes are somehow connected, while central Fz, Cz and Pz nodes are selectively connected to parietal (P) and occipital (O) lobe nodes.

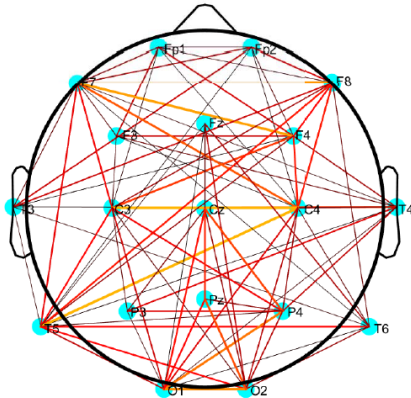


Figure 3: Network visualization for the network of average successful individuals under cognitive task condition. Edges with weight less than a threshold $\tau=0.5$ are omitted, while the other edges are colored with increasing darkness with weight.

To probe in network characteristics traced in adjacency matrices above, we performed analysis of corresponding networks, under various thresholds applied to edge weights. We start with Fig. 4, presenting evolution of network parameters for male and female individuals under resting condition. To briefly explain the network parameters in the coming figures; strength is defined as the sum of all link weights, averaged for all nodes, while path length is average shortest distances between node pairs. Clustering coefficient is the average neighboring rate of nodes' neighbors (excluding itself), while density is the rate of edges to the maximum possible edge count. The reader can be access the details of network parameters via Ref. [21].

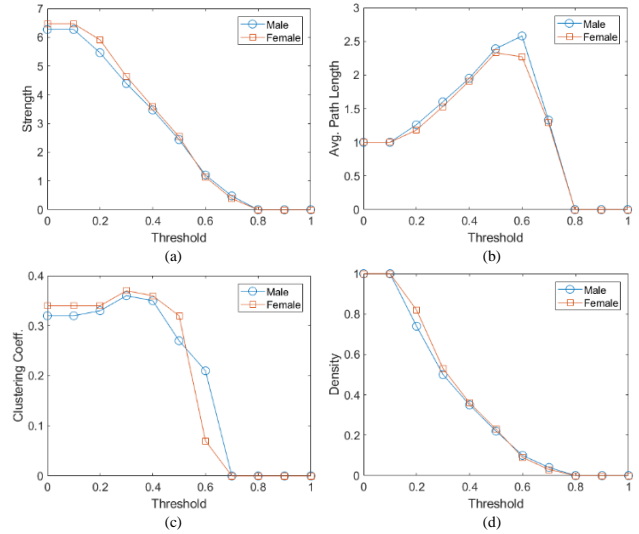


Figure 4: Evolution of network parameters under resting condition of male/female volunteers. (a) Strength, (b) Avg. Path Length, (c) Avg. Clustering Coeff., (d) Density.

Fig. 4 indicates that functional brain networks of female individuals have more strength compared to male peers under resting condition, with lower path length and higher clustering coefficient and density. This is an indicator that, **under resting conditions, female brain is more connected, or more "conscious" compared to male brain.**

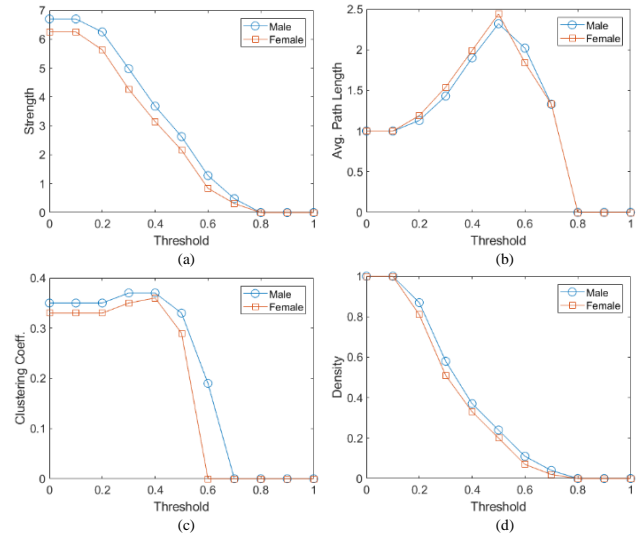


Figure 5: Evolution of network parameters under cognitive task condition of male/female volunteers. (a) Strength, (b) Avg. Path Length, (c) Avg. Clustering Coeff., (d) Density.

The same plots for cognitive task instead of resting condition are presented in Fig. 5. These figures outline that the outlook under resting condition is inversed for cognitive task, demonstrating higher strength, clustering and network density and lower average path length for male individuals. This shows that, **under cognitive task, male brain is more connected, or**

more “motivated” compared to female brain. This diversity in connectivity structure from rest to task conditions according to gender is a noteworthy output for these first two sets of network structure analysis.

The coming two sets of network evolution analysis involve in successful and unsuccessful distinction of volunteers under rest and task conditions. Fig. 6 presents the graph set for resting condition.

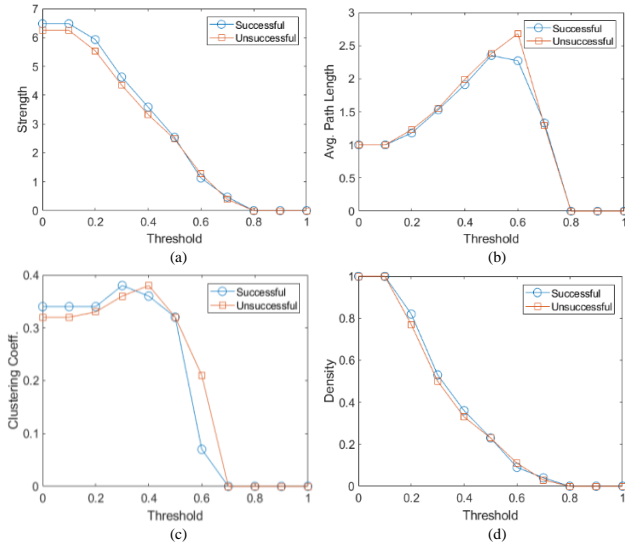


Figure 6: Evolution of network parameters under resting condition of successful/unsuccessful volunteers. (a) Strength, (b) Avg. Path Length, (c) Avg. Clustering Coeff., (d) Density.

Fig. 6 demonstrates that **successful individuals under resting condition have higher network strength and density with lower average path length, indicating a more-connected structure.** Clustering characteristics exhibits a two-phase characteristic, where successful individuals have more clustered brain networks for low-threshold status and less-clustered networks for high thresholds ($\tau > 0.35$). This indicates that high-weight connections of successful individuals exhibit a less-clustered structure and vice-versa for unsuccessful individuals.

The last set of plots presented in Fig. 7 involves with cognitive task condition, with distinction of successful and unsuccessful individuals. These set of plots exhibit similar properties with the resting state plots in Fig. 6, with minimal differences between successful and unsuccessful individuals. **This outcome indicates that arithmetically successful individuals differ from unsuccessful ones with their greater resting-state connectivity, while both success groups have very similar functional connectivity schemes under cognitive task.**

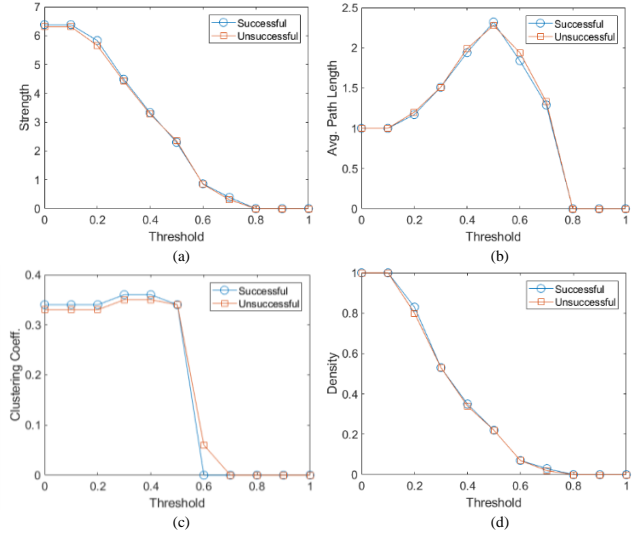


Figure 7: Evolution of network parameters under arithmetic task condition of successful/unsuccessful volunteers. (a) Strength, (b) Avg. Path Length, (c) Avg. Clustering Coeff., (d) Density.

IV. DISCUSSION

Evaluating functional brain networks with phase coherence method applied to EEG signals, we captured diverse outcomes in connectivity structure from rest to task conditions according to gender. Female individuals have greater connectivity compared to male peers under resting condition, whereas this outcome reverses under cognitive task, for which male individuals have greater connectivity.

On the other hand, successful individuals exhibit higher brain connectivity compared to unsuccessful peers under resting condition, while this diversity is preserved but reduced to a minimal level under a cognitive task. This is an indicator that arithmetically successful individuals differ by unsuccessful ones with their higher resting-state connectivity, since their connectivity schemes under cognitive task are very similar. We can conclude this situation that successful brains are the more active ones during resting state.

ACKNOWLEDGMENT

We thank to Zyma et al. for their excessive efforts for providing the EEG dataset and serving as publicly available at: <https://physionet.org/physiobank/database/eegmat/>.

REFERENCES

- [1] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, pp. 47-97, Jan 2002.
- [2] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, p. 186, 2009.
- [3] A. Joudaki, N. Salehi, M. Jalili, and M. G. Knyazeva, "EEG-based functional brain networks: does the network size matter?," *PloS one*, vol. 7, p. e35673, 2012.
- [4] M. Hassan, O. Dufor, I. Merlet, C. Berrou, and F. Wendling, "EEG source connectivity analysis: from dense array recordings to brain networks," *PloS one*, vol. 9, p. e105041, 2014.
- [5] O. Sporns, *Networks of the Brain*: MIT press, 2010.

- [6] K. Smith and J. Escudero, "The complex hierarchical topology of EEG functional connectivity," *Journal of neuroscience methods*, vol. 276, pp. 1-12, 2017.
- [7] F. Miraglia, F. Vecchio, and P. M. Rossini, "Searching for signs of aging and dementia in EEG through network analysis," *Behavioural brain research*, vol. 317, pp. 292-300, 2017.
- [8] A. Griffo, P. S. Baumann, J.-P. Thiran, and P. Hagmann, "Structural connectomics in brain diseases," *Neuroimage*, vol. 80, pp. 515-526, 2013.
- [9] F. Vecchio, F. Miraglia, and P. M. Rossini, "Connectome: Graph theory application in functional brain network architecture," *Clinical neurophysiology practice*, vol. 2, pp. 206-213, 2017.
- [10] G. Niso, R. Bruña, E. Pereda, R. Gutiérrez, R. Bajo, F. Maestú, *et al.*, "HERMES: towards an integrated toolbox to characterize functional and effective brain connectivity," *Neuroinformatics*, vol. 11, pp. 405-434, 2013.
- [11] K. J. Friston, "Functional and effective connectivity: a review," *Brain connectivity*, vol. 1, pp. 13-36, 2011.
- [12] S.-M. Cai, W. Chen, D.-B. Liu, M. Tang, and X. Chen, "Complex network analysis of brain functional connectivity under a multi-step cognitive task," *Physica A: Statistical Mechanics and its Applications*, vol. 466, pp. 663-671, 2017.
- [13] C. Honey, O. Sporns, L. Cammoun, X. Gigandet, J.-P. Thiran, R. Meuli, *et al.*, "Predicting human resting-state functional connectivity from structural connectivity," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 2035-2040, 2009.
- [14] Michael W. Cole, Danielle S. Bassett, Jonathan D. Power, Todd S. Braver, and Steven E. Petersen, "Intrinsic and Task-Evoked Network Architectures of the Human Brain," *Neuron*, vol. 83, pp. 238-251, 2014/07/02/ 2014.
- [15] E. C. van Straaten and C. J. Stam, "Structure out of chaos: functional brain network analysis with EEG, MEG, and functional MRI," *European Neuropsychopharmacology*, vol. 23, pp. 7-18, 2013.
- [16] I. Zyma, S. Tukaev, I. Seleznev, K. Kiyono, A. Popov, M. Chernykh, *et al.*, "Electroencephalograms during Mental Arithmetic Task Performance," *Data*, vol. 4, p. 14, 2019.
- [17] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, pp. e215-e220, 2000.
- [18] G. M. Rojas, C. Alvarez, C. E. Montoya, M. de la Iglesia-Vayá, J. E. Cisternas, and M. Gálvez, "Study of resting-state functional connectivity networks using EEG electrodes position as seed," *Frontiers in neuroscience*, vol. 12, 2018.
- [19] F. C. Galán and C. R. Beal, "EEG estimates of engagement and cognitive workload predict math problem solving outcomes," in *International Conference on User Modeling, Adaptation, and Personalization*, 2012, pp. 51-62.
- [20] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, pp. 1059-1069, 2010.
- [21] A. L. Barabási, *Network Science*. Cambridge: Cambridge University Press, 2016.

A Suggestion for Electronic Election System Based on Blockchain

B. ESEN¹, M. ÖZKURT² and İ. M. ORAK³

¹ Karabuk University, Karabuk/Turkey, esenburak09@gmail.com

² Karabuk University, Karabuk/Turkey, mutluozkurt@gmail.com

³ Karabuk University, Karabuk/Turkey, imorak@karabuk.edu.tr

Abstract - With the digital transformation era that we are in, the systems that we physically process are now transferred to electronic systems. In these days, Bitcoin [1] that is concept of virtual money which is one of the popular topics entered our lives. It is an end-to-end digital payment system over a decentralized network. This system works with the Blockchain algorithm. We can apply this technology to other areas in our digitalized life not only to money transactions. In this project, electronic election system is proposed and designed using the Blockchain technology and preserving privacy.

Keywords – blockchain, electronic election, privacy preserved, cryptography, consensus

I. INTRODUCTION

In this study it is intended to address the shortcomings of the current voting system and to reduce the slowness of the current system, to increase its reliability, to facilitate voting and to reduce election costs.

There will be no expenses such as printing of ballot papers, carriage of votes, seal, envelope, ink. People can use the vote wherever they want, even on the day of the election. In addition, the mobile application allows you to select the closest and least dense voting box via the map and vote quickly. When voting, it is possible to vote with a fingerprint instead of the present signature signing system.

The security of vote is provided by Blockchain's Consensus Algorithm. According to this algorithm, there is a chain in each voting center and this chain is added to the votes used in other centers. If those who try to manipulate the election process try to change the chain in any center, the chain of that center will be suspended because it will be inconsistent with other chains. The suspended chain is then updated again according to other centers where the chain is correct.

In 2011, 246 million 298 thousand 170 Turkish Liras were used for poll clerks, voting of ballots, carriage of votes and similar services [2]. According to election data made on 24 June 2018 Sunday, the number of registered voters is 59 million 367 thousand 469, voting voters 51 million 197 thousand 959, valid votes 50 million 68 thousand 627, invalid votes 1 million 129 thousand 332, participation rate %86.24[3]. According to these data, many voters cannot vote in elections due to road expenses, seasonal conditions, job intensities and other reasons.

Sometimes the votes are invalid due to incorrect voting. After the election is over, counting the votes takes a lot of time and sometimes the errors are counted. This reduces the efficiency of the election.

Protecting the security during voting is one of the most important issues. Transparency, reliability and confidentiality are important elements of the electronic voting system. With this project, a safe voting environment will be created in voting centers with fingerprints and ID cards. With the fingerprint and identification card, we can vote directly in the election centers without having to give the voters a separate password before the election. The most important characteristic is that the relationship between voter and ballot will not be disclosed. Instead of the ballots in the current system, the data will be stored and processed in a digitally distributed database blockchain. Thus, if there is any intervention in the votes in the database, it can be easily noticed due to the structure of the blockchain. In addition, all data in the system is transmitted over the encrypted network. When the election is over, the election results can be declared immediately.

II. RELATED WORK

Elections are the cornerstone of democracy in which the public expresses their opinion in the vote form. For the importance of society, voters must ensure that elections are transparent, reliable and confidential.

Some basic features to be considered when developing electronic voting system according to the system proposed by Rura L. [4] are as follows:

- Privacy: Keep the votes of the voters secret.
- Eligibility: Only the people who are able to vote can vote and the people who has voted cannot vote again.
- Receipt Freeness: Voters should not be able to show their votes to others.
- Convenience: Everyone should be able to vote easily and everyone should be eligible to vote.
- Verifiability: Real votes must be recorded.

Koç and his friends suggested an e-voting system by using Ethereum's Smart Contract system [5]. In another study "Block Chain Voting System" was developed by using Ethereum. In the system, voters can vote with any device that can connect to the internet and have a camera. In this system, the

Authentication Server is used confirm the voters through a database keeping the information of all citizens in the country. It also registers accounts of voters and candidates. Once the voter has been verified, the token is created through the Arbitration Server and the voting process is performed [6]. As a result, people who will vote for the operation of this system must have Ether in their Wallet and must register the Wallet address to the person who initiated the election. This process takes time. In addition, the Ethereum Network, the person using the vote and all the work is recorded as clear text and may cause problems after the election. It is unclear how the Ethereum Network will react under a system-wide load in a large nationwide election. There is no guarantee for scalability.

Blockchain system provides some benefits with its cryptographic foundations to transparent e-voting schemes. These can be considered in four layers as user interaction and front-end security, access control management, e-voting transaction management, and ledger synchronization [7].

World's first blockchain presidential election held on March 7, 2018 in Sierra Leone [8]. But the government of Sierra Leone said the government did not officially use a blockchain-based system [9]. According to this statement, voters gave their votes by conventional means and the government concluded by counting the votes with their own systems. Then Agora employees, a blockchain voting initiative of Swiss origin, handed the paper compasses manually to the blockchain database and explained the results of their election from their systems.

Another country using the blockchain system is the US. A blockchain-backed digital voting platform was used in local elections in West Virginia, USA on May 8 [10]. Members of the army and foreign voters from abroad due to compulsory services participated in elections by voting on their smartphones. In this election, which was a pilot application, voters downloaded a blockchain-based application to their smartphones and authenticated their identity through this application with facial recognition and fingerprint reading. Government officials have examined the election process and, according to the report they will prepare, will decide whether all West Virginia voters will use this system in the next elections [11].

Another country which used the blockchain system is Japan. On September 3, 2018, Tsukuba, a city on the island of Honshu in Japan, used Blockchain technology to allow citizens to choose a social contribution project in a fair way. According to the local newspaper, Tsukuba is the first city in Japan to implement the Blockchain voting system [12].

The new voting system is based on a 12-digit ID numbering system issued to all Japanese citizens to validate their voter identity before they gain the right to vote. As with most new technologies, there may be certain problems. Many voters have forgotten their passwords to vote and there are some uncertainties about whether the votes should be counted. Also, it is not known whether a person who reads a card is a real measure of whether the card is the real owner or not. However, once the system is in operation, the city government plans to expand the system to residents in remote areas, islands and

foreign countries [13].

HAVELSAN's "e-sandık" project completed in 2013. In order to use the machine, voters will show their IDs and receive randomly defined QR code slips from the ballot box committee. After holding the QR code on the machine, the voters will click on the 'start voting' tab and vote for the party they want. After the voting process is completed, the compass from the machine will be thrown into real boxes. With this system, the nationwide vote counting will result in 5-10 minutes. Also, the use of real ballot box in this system shows that there is no difference in the current system [14].

III. BACKGROUND

A. Blockchain

The blockchain, in its simplest definition, is the specialized state of a linked list structure. In the standard single-link list structure, each element of the list points to a pointer followed by the following element. In this way all elements from the starting element of the list to the last element of the list are connected to each other [15]. In the blockchain structure, each element (block) not only points to the next block also the hash value of that block as shown Figure 1. In other words, blockchain is a special linked list structure created with hash-pointers [16].

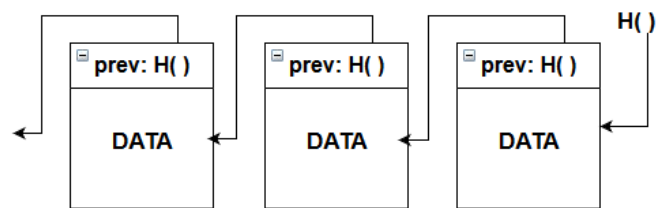


Figure 1: Blockchain Data Structure

The most important difference of the blockchain from the hash-pointer structure in relation to the classical connected list arises when any block in the list is requested to be changed. This type of change is easily noticeable. Because the newly added block hash value will be different from the value pointed to by the previous block. This feature is one of the important factors that make the blockchain a safe structure [17].

These are peer-to-peer networked computers that enable the realization of transactions and keep the entire Blockchain as shown Figure 2. These computers add blocks to the Blockchain by placing the data in the blocks and creating the hash code.

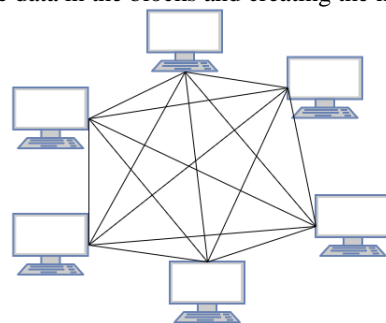


Figure 2: Node Structure

B. Advanced Encryption Standard

The privacy and security of votes are important elements of this system. For this reason, we need Advanced Encryption Standard system in order to ensure that the votes between the nodes are not visible and reach safely. The encryption algorithm defined by AES is a symmetric-keyed algorithm in which both the encryption and the keys used to decryption the encrypted text are related to each other. The encryption and decryption keys for AES are the same.

Both software and hardware performance of AES is high. The algorithm consists of an identical number of repetitive input open text, and identical conversion cycles (round) that convert the output to encrypted text. Each cycle consists of four steps, except the last round. These rounds are applied in reverse order to decrypt the encrypted text. The repeats of the rounds are 10, 12 and 14, respectively, for the 128-bit, 192-bit, and 256-bit key lengths [18].

C. RSA

RSA encryption is used to secure the symmetric key to the node safely. RSA is an asymmetric encryption algorithm. In RSA encryption, the user has 2 keys which are hidden and open. The party who wants to send a message will send the message that the other user wants to send with the public key by encrypting. The user who receives the encrypted message can access the open message value by decrypting the encrypted text with the private key. The user who wants to generate a private and public key will use the following algorithm.

- Select two major prime numbers close to each other. Let's call these two p and q
- Perform $n = p * q$
- We calculate the totient function of these numbers.
 $\varphi(n) = (p - 1) * (q - 1)$
- An integer is generated in the range of $1 < e < \varphi(n)$. This is the number of e and $\varphi(n)$ must be prime among them. The number e will be our public key.
- We find a number d with $d * e \equiv 1 \pmod{\varphi(n)}$. The value of d from here will be our private key.

As a result, our public key (e, n) consists of our secret key (d, n) . In addition, the p , q and $\varphi(n)$ values we have determined at the beginning should remain confidential. These values should remain confidential because we use them to find e and d values [19].

D. Digital Signatures

Another factor related to the security of the vote is to verify whether the votes from the nodes come from the actual nodes. That's why we use RSA digital signing. With RSA Digital Signing, the node from which the votes came can be identified. If the votes are changed during the communication or if a vote is sent from devices that are not included in the system, RSA Digital Signing will enable receiver to relise them easily. RSA Digital Signing is briefly explained with a following sample.

For example, Mutlu should confirm that the message he received was from Burak. Burak can do this with his private key. Burak produces a hash value of the message, raises it to the

power of d (modulo n) (as Burak does when decrypting a message), and attaches it as a "signature" to the message. When Mutlu receives the signed message, he uses the same hash algorithm in conjunction with Burak's public key. Mutlu raises the signature to the power of e (modulo n) (as he does when encrypting a message) and compares the resulting hash value with the message's actual hash value. If the two agree, he knows that the author of the message was in possession of Burak's private key, and that the message has not been tampered with since [20].

IV. PROPOSED SYSTEM

The system we propose is to be transparent, reliable and confidential shown on Figure 3. This system consists of five layers as shown on Figure 4.

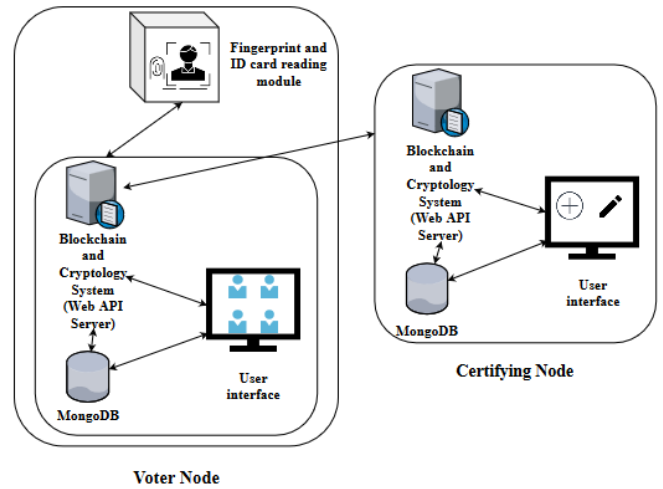


Figure 3: General Architecture and Relations

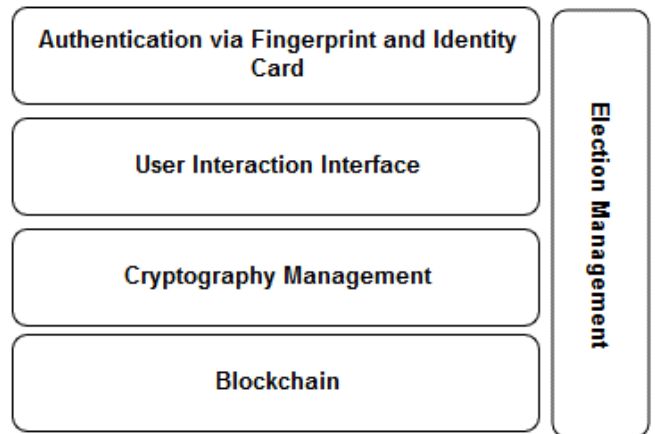


Figure 4: System Layers

A. Election Management

It is a system that runs on the node that controls the joining of the nodes and the voter information as shown flow-chart on Figure 5. A new node that want to join the system sends request with the poll clerk information. After checking the accuracy of this information, the system creates a digital signature. A

symmetric key (such as a certificate) and a previously created digital signature are sent to the node for use in encrypted communication. Thus, the relevant node is added to the system and ready to vote.

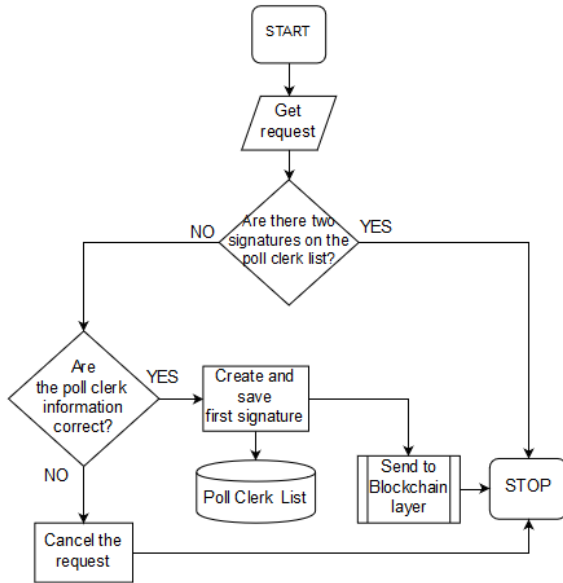


Figure 5: Poll Clerk List Management Flow-chart

This layer also keeps list of voters. It controls the voting process with two signatures as shown flow-chart on Figure 6. First step of the voting process is initiated by fingerprint and card reading. It checks the voter information from its own records. If the first signature is present, it creates a signature again and sends it to the user interacting layer. If the vote from the Blockchain Layer successfully dispatches the system with the first signature, the second signature is created and sent to the Blockchain Layer.

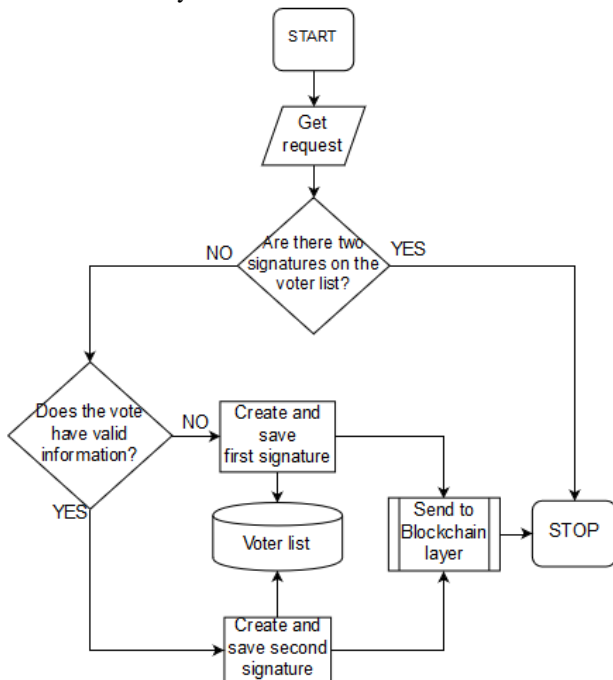


Figure 6: Voter List Management Flow-chart

B. Authentication via Fingerprint and Identity Card

This layer is modular and compatible with the country's chip card systems. It compares the information in the card and the information on the fingerprint sensor. It encrypts the person who gets his/her card read in the system and sends it to the Certifying Node who makes the election. This layer expects confirmation for the person whose information is sent if he/she is a voter and also whether he/she previously voted. In order to obtain statistical information, the voter sends information (such as province, district, neighborhood, ballot box, etc. which does not include any personal information) and the signature that the voter may vote in a coded manner to the system where the voter will vote. After this layer of the system, a return from the central system is received and it directs the voter to the voting interface.

C. User Interaction Interface

It is the interface in which the voters who have received the approval to vote can vote. If the voter does not receive the approval signature, he cannot access this layer. The vote used will be sent to the Blockchain Layer. The information from the Blockchain layer shows the informational message that indicates whether the vote has been successfully registered to the system.

D. Cryptography Management

It is the layer that encrypts and decrypts information of vote, voter, poll clerk by using symmetric cryptography as shown on Figure 7 and 8. In addition, this layer generates symmetric keys for the system to use in communication.

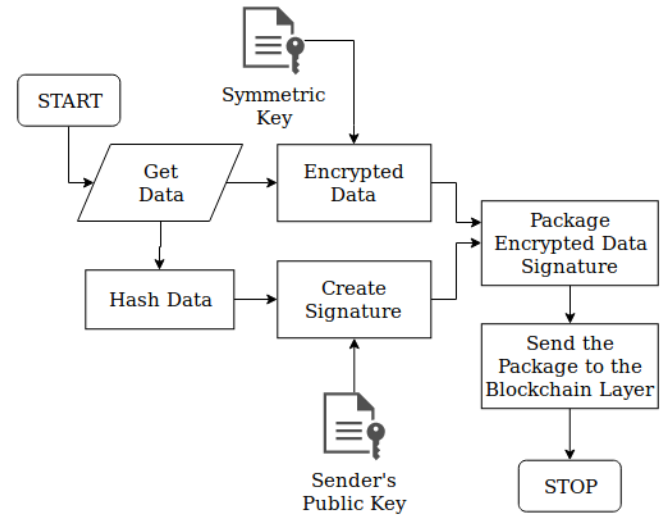


Figure 7: Encryption and Signing Flow-chart

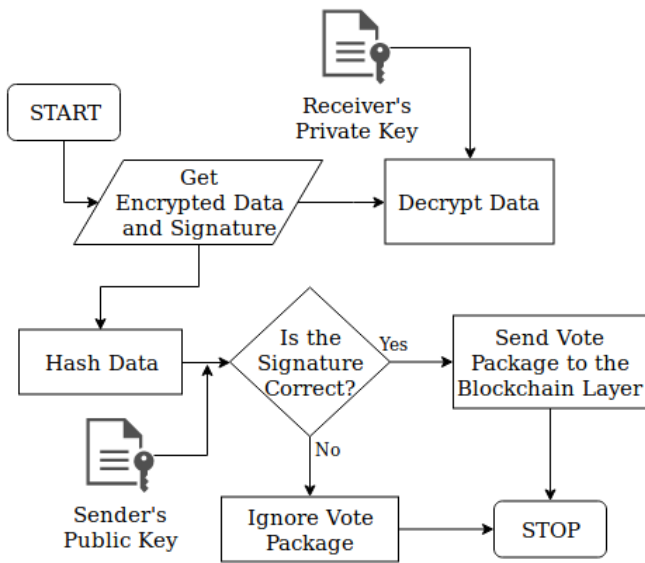


Figure 8: Decryption and Verifying Signature Flow-chart

E. Blockchain

It is the layer in which the votes are sent, processed, checked, recorded in the local database and the votes are counted as regional as shown flow-chart on Figure 9.

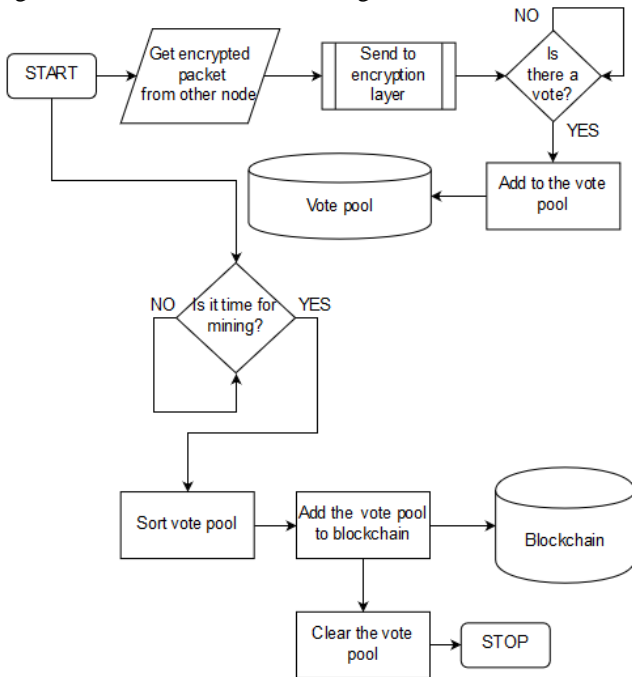


Figure 9: Blockchain Layer Flow-chart

The voter first assigns the polling package from the interface to the local buffer pool and then sends it to the encryption layer. It sends the encrypted data packet returned from the encryption layer to all nodes in the node list. If the vote has been submitted successfully, information is sent to the Election Management System with the first signature and the information is sent to the Voting interface. If the vote cannot be sent, the Election Management System cannot create the second signature. The

knowledge that the vote cannot be saved is also reported to the Voting interface.

There are two pools called "left" and "right" in the buffer pool section. In the process of adding votes to the buffer pool, the "left" or "right" goes to the pool according to when the ballot is created. For example, packages created in the "left" pool's time interval will be added to the "left" pool, even if they are in the "right" pool's time interval. If "Mining" is the time to implement, then the pool is "mining" if the current time zone is not the time zone of the pool and the required buffer time has elapsed. In this way, even if the packets are late, voting packages do not miss the "mining" process. If the package does not arrive within the buffer period and has not reached the other nodes, it will be considered that the package has not been sent. In this case, the Blockchain layer deletes the voting pack in the local blockchain and starts synchronization service. However, since the ballot box could not be sent, the second signature does not go to the system holding the voter list, indicating that the voting process is complete.

In the "Mining" process, the votes from the pool are sorted by the time of voting. The clock is always synchronized according to the clock time of the Certifying Node. There is also a service that synchronizes the time of the Certifying Node. The nodes sorted according to the time take the hash of the previous voting package and add it to their voting package, then take the hash of this voting package together with the voting package and create a separate voting block and the block is added to the chain. With the additional process, the vote meter in the tree structure is updated.

Blockchain, buffer pool, and node lists are kept in the local database in case of problems. Once the system is started, the blockchain is synchronized with other nodes. When the system restarts, the votes in the buffer pool are added to the system. If the signature does not comply with the rules (the system has taken a lot of time to restart), the voter can vote on other nodes or centers, as the signature indicating the completion of the vote does not go to the system holding the list of voters. All votes that do not comply with the rules in the buffer pool are deleted without being added to the system.

Voter Node has not authority to initiate and end the election. The signature of the polling clerk is required for this node to be included in the system. After obtaining the signature, the Voter Node contacts the Certifying Node. In this communication, it sends its node information, the information of the polling clerk. Then the Certifying Node asks for the other node information that is included in the system from the Node so that the Voter Node joins the system. This node accepts and answers all requests in the Blockchain Layer.

The Certifying Node can start and end the election. It does not respond to requests to add or delete votes. It can only create chains for synchronization. Keeps a list of all nodes in the voting system. In addition to the node information, this list also contains information about the polling clerk who started that node. According to the election rules, the vote does not respond to the vote result requests before the election ends, but when the election is over, both the election results and the blockchain are also accessible from outside the system.

V. DISCUSSION

In order to prove our design, we compared the characteristics of our system with others. As mentioned in the literature section, e-ballot box in the system of HAVELSAN counts the votes within itself. However, after the election is ended, the transfer of the results to the head office takes longer than our proposed system. Thanks to Blockchain and cryptology, the votes in our system can be counted as instant. Comparison of these two systems can be seen in Table 1.

Table 1: Performance of Vote Count

	HAVELSAN e-Sandık	Our proposed system
vote counting time	5-10 mins.	instant

When we compare our system with the blockchain based other systems, it can be seen that our system isolate voter info and his/her vote. This will provide more reliable and secure results.

VI. CONCLUSION

In line with our objectives, our suggested system will provide benefit to the country's economy by reducing the election costs. It will allow voters to vote in any place they want. Hence, the vote rate will be increased. On the other hand, no vote given in line with the rules will be considered as invalid. The relationship between voter and ballot is not disclosed. The election results will be declared instantly and statistically at any time. Since the system we propose is made up of layers, it can be easily aligned with different systems.

The reliability, transparency and security structure of our proposed system makes this electronic selection system as applicable in real life.

In future work this suggested system will be tested in a simulation environment and after that it will be proved by real time application in limited extend.

This work supported by 2209-A University Students Research Projects Support of TÜBİTAK.

REFERENCES

- [1] Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system." (2008)
- [2] <https://www.haberler.com/ysk-baskani-em-secimin-maliyeti-246-milyon-298-bin-2682442-haberi/>
- [3] <http://www.ysk.gov.tr/tr/24-haziran-2018-secimleri/77536>
- [4] Rura L., Issac B., and Haldar M. K. (2016) Implementation and evaluation of steganography based online voting, International Journal of Electronic Government Research.
- [5] Ali Kaan Koç, Emre Yavuz, Umut Can Çabuk, Gökhan Dalkılıç, "Towards Secure E-Voting Using Ethereum Blockchain", pp. 3-5, IEEE 2017.
- [6] Sagar Shah, Qaish Kanchwala, Huaqian Mi, "Block Chain Voting System", NORTHEASTERN UNIVERSITY, pp. 5-9.
- [7] Kashif Mehboob Khan, Junaid Arshad, Muhammad Mubashir Khan, "Secure Digital Voting System based on Blockchain Technology", January 2018
- [8] <http://uzmanpara.milliyet.com.tr/haber-detay/gundem/blok-oy-sisteminden-blok-zinciri-oy-sistemine/81353/>

- [9] <https://siberbulten.com/uluslararasi-iliskiler/sierra-leoneda-ilk-blok-zinciri-tabanli-secim-gerceklesti/>
- [10] <https://medium.com/dijitaldemokrasi/se%C3%A7im-g%C3%BCvenli%C4%B1-ve-blockchain-e32440bd1424>
- [11] <https://koinbulteni.com/abdnin-bati-virginia-eyaletindeki-vatandaslar-blockchain-uygulamasiyla-oy-kullanabilecekler-26954.html>
- [12] <https://kriptoparahaber.com/japonya-ilk-blockchain-oylama-sistemini-kullandi.html>
- [13] <https://webrazzi.com/2018/09/05/japonya-blockchain-oylama>
- [14] <https://digitalage.com.tr/e-secim-tartismalara-cozum-olur-mu/>
- [15] Cormen, T. H., Leiserson, J. E., Rivest, R. L., Stein, C. 2009, Introduction to Algorithms, 3rd Edition, MIT Press, A.B.D
- [16] Narayanan, A., Bonneau, J., Felten, E., Miller, A., Goldfeder, S. 2016 Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction, Princeton University, New Jersey.
- [17] Ersin ÜNSAL, Ömer KOCAOĞLU Avrupa Bilim ve Teknoloji Dergisi Sayı 13, S. 54-64, Ağustos 2018 "Blok Zinciri Teknolojisi: Kullanım Alanları, Açık Noktaları ve Gelecek Beklentileri"
- [18] Advanced Encryption Standard (AES),» FIPS PUB 197, 2001
- [19] Emre CERAN, Mehmet Sabır KİRAZ, Osmanbey UZUNKOL Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi Cilt 21, Sayı 2, 631-643, 2017 "RSA, Sifreleme Sistemlerinin Kleptografik Arka Kapıları için Güvenlik ve Karmaşıklık Analizi"
- [20] Rivest, R.; Shamir, A.; Adleman, L. (February 1978). "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems"

Assessing the Architectural Quality of Software Projects from an Organizational Perspective

F. CETIN, O. TAVILOGLU, C. YILMAZ, N. Z. UGUR and M. BAYAR

Siemens A.S., Istanbul/Turkey, fatihcetin@siemens.com
Siemens A.S., Istanbul/Turkey, onur.taviloglu@siemens.com
Siemens A.S., Istanbul /Turkey, caner.yilmaz@siemens.com
Siemens A.S., Istanbul/Turkey, nazim.ugur@siemens.com
Siemens A.S., Istanbul/Turkey, mujdat.bayar@siemens.com

Abstract – Software architecture has a big impact on creating a high-quality software product. In complex software organizations with multiple software products, it is an increasing need to validate architectural quality with quantitative metrics by an organizational point of view. Within customized quantitative metrics, the organization can create a common understanding of organizational software architecture quality, classify products based on maturity levels, inspect and adapt the improvement areas. This paper presents an empirical process definition for measuring software architecture quality in a multicultural, distributed organization.

Keywords – Software Architecture, Software Architecture Evaluation, Software Quality, Software Architecture Assessment

I. INTRODUCTION

Software architects are the primary responsible for the high-quality architectures that fulfill the business goals[1]. Overall architectural quality in a software organization is highly dependent on the architect's skills, the personality of an individual architect. Supporting architects with tools and methods have an undeniable impact on producing high-quality architectures. On the other hand, it may also be beneficial to use an organization-wide set of methods and tools that inspects architectures, exposes improvement areas. Using organization-wide tools and methods for assessing architectural quality help to create a common understanding and brings the ability to measure the quality from organizational perspective. In order to introduce a methodology for organization-wide architecture quality assessment we had twelve workshops with six contributors. Workshops focused to find the right answers to the following topics:

- Definition of the focus areas where the architectural quality can be assessed
- Definition of assessment levels where each level maps a focus area based on “W-Model of Software Architecture [2]” (see Section II for the details).
- Definition of done criteria for each level
- Definition of fundamental principles
- Create an assessment based on fundamental principles

Figure 1 depicts the overview of levels and transitions

between them. The levels indicate the maturity of the architecture and the process that is used to create architecture.

In this paper, we will discuss our approach assessing architectural quality by organizational perspective.

II. DEFINITION OF ASSESSMENT LEVELS

Assessment levels are the steps that get executed during the proposed architecture assessment. They are designed to be performed sequentially. It should be noted that the technologies used and the size of the projects are not related to the project assessment. The sequential transitions between the levels are depicted in Figure 1.

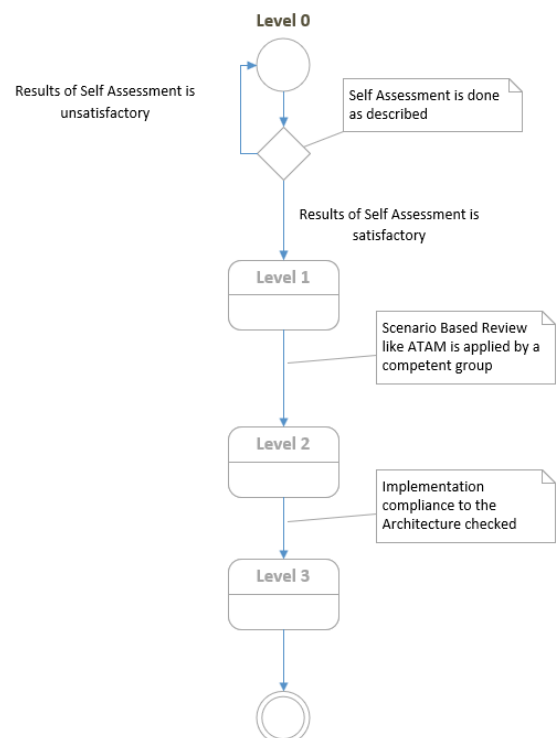


Figure 1: Transition of Levels

For the definition of the levels, the W-Model of Software Architecture constitutes a base for the study presented in this paper. Figure 2 shows how the Levels are mapped to the W-

Model of Software Architecture. The self-assessment done for Level 1 contains a questionnaire for the architect about the traceability (both horizontal and vertical) of each step of the W-Model while skipping the architecture/design review steps in the original model. The questionnaire especially focuses on the marginal legs of the shape "W" as shown in the Figure 2. In addition, Level 2 (Architecture Review) and Level 3 (Design Review) has already some correspondent representation in the W-Model. Note that the design review proposed in the W-Model of Software Architecture mostly focuses on the conformance of the detailed design on the architecture. However, in our work, as a difference from the W-Model's approach, the review for Level 3 mostly focuses on the conformance of the implementation with the architecture.

A. Level 0

This is the initial state for all projects in the organization. In order for a project at this level to pass to the next level, it must reach the level of qualification in the self-assessment tool.

B. Level 1

This level indicates project successes on self-assessment. Enough points on self-assessments ensure that adequate architectural documentation exists and scenario-based review methods [3] can be applied to the project.

C. Level 2

Scenario-based reviews (like ATAM) [4] are applied on Level 1 projects by the experts/other architects in the organization. Projects are shortly checked if they meet self-assessment criteria and the architecture review takes place. If there are not huge concerns about the architecture, the project is labeled as Level 2. Architecture mostly addresses Quality Attributes.

D. Level 3

This level is for checking if the code is compliant with the design. One of the following methods can be applied:

1. The code is tracked and controlled against the design with the tools. Tools, tools' configurations and the rules are examined and automatic tracking correctly validates code against architecture.
2. Expert review is (partially/fully) done. The developed code corresponds to the intended architecture.

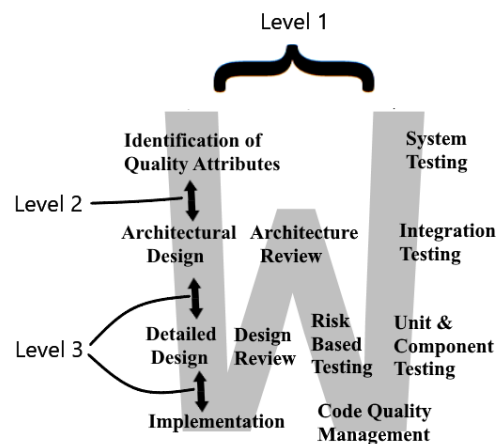


Figure 2: Mapping of Levels to W-Model of Software Architecture

III. FUNDAMENTAL PRINCIPLES

A. Architecturally Significant Requirements

Architecturally significant requirements compose the base for the design of the software architecture. Architecturally significant requirements are rarely explicit [5]. The architect must carefully expose architecturally significant requirements and makes them explicit.

Architecturally significant requirements should match to important requirements that the product needs to have, these requirements can be functional or non-functional. This is why these explications are essential to the assessment of projects.

1) Quality Attributes

Quality attributes are the list of features or characteristics of a software system by taking consideration of all impacting aspects. These aspects might not only customer requirements and software features but also hardware configurations as well. These attributes might be very easy or difficult to measure.

In this paper, we have preferred to use metrics of ISO standards [6] which are defined also quality characteristics. Since all quality attributes may match to the requirements then ASR's (Architecturally Significant Requirements) as well. Some of these quality attributes are below: Performance, compatibility, usability, reliability, security.

2) Business Goals

If the business requirements are aimed for high-level products to be developed, architectural vision and work are also indispensable. In this way, business goals should not contradict with architecturally significant requirements. These kinds of conflicts should be resolved immediately by workshops with the contribution of both sides.

3) Customer Requirements

It is highly recommended that customer requirements include some of the quality attributes so that the architect can design and create a software base on these important quality attributes.

This mission is not an easy task, in order to accomplish this, the architect and owner of the requirements must do hard work together. As a result of this hard work, some trade-offs might occur as a disadvantage of product. In this paper, we also aimed to measure by taking these trade-offs into account as well.

B. Architectural Documentation

Having concise, up-to-date, unambiguous and readable documentation is one of the key assets of the software architecture and also plays a substantial role in architecture evaluation. Architectural documentation constitutes a base common language for stakeholders of the architecture. The absence of proper documentation results in complexity, misunderstandings and more effort in the assessment process. Thus, a process that is proposed in this paper requires proper documentation for the Level 2 assessment. Use of organization-specific or publicly available templates such as arc42 [7] is possible for document creation.

C. Design Decisions

Design decisions for architecture is another important input for the evaluation of software architecture. Design decisions, the rationale behind them, addressing quality attributes (traceability of design decisions), risks, sensitivity points or trade-offs that they create shall be clearly defined and presented to the stakeholders of the architecture.

D. Testing

Software Testing is a crucial part of software development for assuring the quality and also measuring the maturity of software. The Quality Attributes addressed by the Architecture must be tested in the product. A proper testing strategy (e.g. Risk-Based Testing) [8] has to be used to efficiently verify the correctness of the implementation done for addressing QAs at an appropriate level of testing. It is also expected that testing levels are also in conformance with the development model. For instance, if W-Model is chosen for development, it shall be ensured that matching test step shall be planned for each development step of the model. Moreover, traceability and coverage of tests to the ASRs also shall be contained by the architecture definition of software.

E. Implementation

In order to address the ASRs in the product, the implementation should match the design. For the Level 3 assessments, the implementation is verified to confirm the design. Following options are all accepted to make the verification:

1. A model is built from the code and compared to the intended design [9]
2. An experience-based review is done that focuses on the code [10]

Also using an automated tool based on the SQALE [11] approach helps to see that the implementation really addresses the ASRs. However, as SQALE does not really compare the design with implementation it cannot be accepted as a methodology by itself for Level 3.

IV. SELF-ASSESSMENT

Self-assessment is the step that project members themselves qualify their project if it is ready for further assessments or not. In our workshops, we have constructed a self-assessment tool. The tool is composed of the important questions asked to the audience like a survey. It also evaluates the answers to the questions and outputs a total self-assessment grade. While answers to questions can be multiple selections, they can also be in other formats depending on the application.

A. General Questions

This section of the self-assessment tool focuses on requirements, testing, and documentation. This section has 15 questions with the weight distribution of requirement %50, test %35 and documentation is %15. Each question has a different coefficient. Questions and the proposed coefficients are:

- To what degree are your functional/nonfunctional requirements defined? (6)
- To what degree is your knowledge/understanding of your product's architecture? (1)
- What is your influence on architectural decisions? (1)
- To what degree are your Quality Attributes defined in your project? (4)
- Are Quality Attributes in your project ranked by priorities? (2)
- To what degree is your Quality Attribute Utility Tree defined? (2)
- To what degree are your Quality Attribute Scenarios defined? (2)
- To what degree is your Product Manager involved in defining your Quality Attributes? (1)
- Do you have architectural documentation? (2)
- How much does your architectural documentation fit in with the organization's architectural documentation template? (4)
- To what degree do the tests/test cases in your project address your Architecturally Significant Requirements? (3)
- To what degree are your Architecturally Significant Requirements covered in system test phase? (3)
- Which test levels (Unit test, Integration test, System test) do you have in your project? (6)
- To what degree do you have design concerns for Testability? (2)
- Is your test strategy defined? (1)

B. Quality Attribute Related Questions

This section inspects the quality attributes. The architect must assess each applicable quality attributes (Performance, Compatibility, Usability, Reliability, Security, Maintainability, Portability, Scalability) with the following question:

- What is the importance of {QA} requirements in your project? This question's answer will define general weight. If not applicable is selected rest of the questions will not be valid and only the next question will be asked.
- Do you foresee {QA} concerns in the future? This question is a conditional question. If the first question

is answered as not applicable this question will be asked. Otherwise, this question is not valid.

- Are {QA} Requirements defined in your project? (1)
- To what degree does your design address the {QA} Requirements? (4)
- Did trade-offs occur after architectural decisions? (1)
- To what degree are your Quality Attributes defined for {QA}? (2)
- To what degree are your Quality Attributes Scenarios' testable? (2)
- To what degree are your Quality Attribute Scenarios tests defined? (1)
- To what degree do your unit tests verify your {QA} requirements? (1)
- To what degree do your integration test verify your {QA} requirements? (3)
- To what degree do your system test verify your {QA} requirements? (3)

V. RESULTS

In our study, we defined maturity levels for architectures in an organization. We set an objective and quantitative criteria for Level 1. Setting such criteria for Level 2 is not necessary at this time, since we are using scenario-based review methods. For Level 3, the approach proposed in "Pragmatic Evaluation of Software Architectures" [9] will be a base for calculation of the metrics.

Self-assessment of projects will contribute the awareness of deficiencies from architecture perspective, following the levels will guide to projects to fulfill the proper architecture entirely.

VI. FUTURE WORK

Applying the process that is presented in this paper by collecting different measures and feedbacks will enable us to set quantitative criteria for other levels. There is also an opportunity to define another level for assessing to what degree the tests address quality attributes.

REFERENCES

- [1] L. Bass, P. Clements, R. Kazman, M. Klen, *Evaluating the Software Architecture Competence of Organizations*, Seventh Working IEEE/IFIP Conference on Software Architecture (WICSA 2008), Vancouver, BC, Canada, 2008.
- [2] O. Taviloglu, F. Cetin, *Tracking the Architectural Quality: "W-Model of Software Architecture"*, ICATCES 2018, Safranbolu, 2018
- [3] Mugurel T. Ionita, Dieter K. Hammer, Henk Obbink, *Scenario-based software architecture evaluation methods: an overview*, In Workshop on Methods and Techniques for Software Architecture Review and Assessment at the International Conference on Software Engineering, Orlando, FL, USA, 2002
- [4] L. Bass, P. Clements, and R. Kazman. *Software Architecture in Practice*. Addison Wesley, Reading, Mass., 1998.
- [5] P. Clements, L. Bass, "Relating Business Goals to Architecturally Significant Requirements for Software Systems", Technical Note, Software Engineering Institute, Carnegie Mellon University, 2010
- [6] ISO/IEC 25010 <https://iso25000.com/index.php/en/iso-25000-standards/iso-25010>
- [7] Arc42, <https://arc42.org>
- [8] F. Redmill, *Theory and practice of risk-based testing: Research articles. Software Testing, Verification and Reliability*, 15(1), 3–20. , 2005
- [9] Knodel, M. Naab, *Pragmatic Evaluation of Software Architectures*, Springer International Publishing, pages 83-95, 2016
- [10] P. Clements, R. Kazman, M. Klein, *Evaluating Software Architectures: Methods and Case Studies*. Addison-Wesley , 2002
- [11] K. Mordal-Manet, F. Balmas, S. Denier, S. Ducasse, H. Wertz, J. Laval, F. Bellingard, and P. Vaillergues. *The squal model - A practice-based industrial quality model*. In Proc. IEEE Int. Conf. on Software Maintenance (ICSM 2009), pages 531–534, 2009.

An Overview on Data-Driven Prognostic Methods of Li-Ion Batteries

E. TASKIN¹, I. BILIK² and N. TOPCUBASI³

¹ Canakkale Onsekiz Mart University, Canakkale/Turkey, erhantaskin@comu.edu.tr

² Turkcell Technology A.S., Istanbul/Turkey, ilker.bilik@turkcell.com.tr

³ Okan University, Istanbul/Turkey, nursen.topcubasi@okan.edu.tr

Abstract - The reliability of the electronic components is very important for the reliability of the whole system, so it is one of the key parameters for the products. The reliability of an electronic component can be defined as performing its task in the system over an expected time [1]. As a power source, batteries are one of the important components to be monitored in systems. This study provides an overview of the data-driven prognostic methods used to predict the future changes in the health status of lithium-ion batteries by examining the normal operating cycles.

Keywords – prognostic methods, machine learning, remaining useful life, battery health monitoring.

INTRODUCTION

Li-ion batteries, mobile phones, mobile computers, electric cars and electrical appliances are widely used both in daily life and in many critical systems such as aerospace, defense and space systems. End of life of the batteries can cause the system to be completely out of service. Therefore, monitoring the health status of the battery and collecting some values is very important for ensuring the continuity of the system. The change in values such as voltage, current and temperature gives information about the health of the battery and helps to predict the future status of the battery [2].

Prognostic methods estimate the future health status of batteries using current and historical health conditions. The predicted future health of a battery is expressed in the remaining useful life (RUL). Three parameters are used when estimating RUL: state of charge (SOC), state of health (SOH) and state of life (SOL). There are three different approaches to the useful life prognostics of the battery: model-based, data-driven, hybrid [3]. Model-based approaches are quite complex. It is difficult to obtain physical changes in advance and it is both costly and time consuming to capture with experiments. Data-driven approaches obtain RUL estimation by processing data with mainly machine learning and statistics-based approaches. Various data-driven methods are available for battery status prediction and prognostic. Hybrid approaches are a combination of model-based and data-driven approaches. Data-driven approaches try to derive models directly from collected data. These methods, which do not require knowledge of the mechanisms in the infrastructure of the system, have become more popular in recent years [4].

DATA-DRIVEN PROGNOSTICS

With the development of sensor technologies and data storage / processing systems, it has been possible to collect large amounts of data and thus the data-oriented approach has become a trend. This approach obtains important data from the usage data from the sensors and then creates a pattern of degradation for the RUL estimation. The concept of degradation is tried to be explained in relation to the change trend of the data. The historical status monitoring data is generated by processing the data related to the current health status of the system and the degradation of similar systems, with various approaches. Data-driven approaches are often divided into two main categories: machine learning and statistical-based approaches [5].

A. Machine Learning Approaches

Machine learning methods use the characteristic of the data to match the monitored data to a pattern of degradation and estimate the RUL. The main used methods are Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Relevance Vector Machines (RVM) [6].

Artificial Neural Networks (ANN): ANNs are one of the computational intelligence tools commonly used in the field of prognostic [7]. An ANN consists of a series of simple processing units called neurons that mimic the ability of the human brain to process and acquire information. Very suitable for system modeling. Due to compatibility to nonlinear functions, they are also suitable for modeling complex systems. NNs can learn and reorganize their internal structure to adapt to a changing environment. They are naturally data-oriented and can model the system without detailed physical knowledge of a system [8].

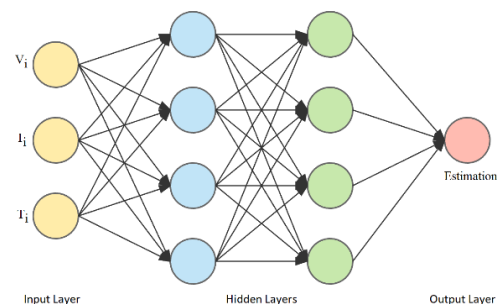


Figure 1: The structure of an Artificial Neural Network

An ANN consists of an input layer with input variables, one or more hidden layers capable of non-linear calculations, and an output layer representing the output of the system. An input layer that takes current, voltage and temperature values, hidden layers with activation function and output layer giving the estimation result can be shown as a battery prognostic sample model in Figure 1.

ANN methods have been used both for SOH [9], SOC [10] and RUL [11] estimations.

Support Vector Machines (SVM): Support vector machines (SVMs) [12] can be used to solve the problem when battery health status is considered as a classification problem. SVMs are a set of supervised learning methods for classification, regression and outliers. The purpose of the SVM is to find a hyperplane that best distinguishes the training samples according to the class label. This hyperplane will probably divide the area into two section for battery prognostics. One half will show that the battery is healthy, the other half is faulty or defective. The health status of the battery can be determined by which side the data will be obtained in the future.

SVMs can be used for linear or nonlinear data sets. A linear hyperplane cannot be drawn in a non-linear dataset. Therefore, the methods are used that called kernel tricks. The most commonly used kernel methods (Figure 2);

- Polynomial kernel
- Radial basis function (Gaussian) kernel
- Sigmoid kernel

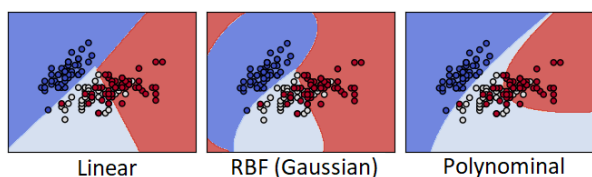


Figure 2: SVM Kernel Methods

SVMs have been frequently used for real-time prediction of remaining useful life [13].

Relevance Vector Machines (RVM): SVM is a well-known technique for classification and regression. There are some disadvantages such as lack of probabilistic outputs [9] which are more meaningful especially in health monitoring applications. RVM attempts to address these problems in a Bayesian framework. In addition to the probabilistic interpretation of its output, it uses a smaller number of kernel methods [14].

RVM supports uncertainty representation. RVM has been used often in studies where uncertainty is used for battery RUL estimation [15].

B. Statistical-Based Approaches

Statistical approaches estimate the beginning and progression of degradations according to the results of the examination on similar systems. These methods require a large amount of data set while they do not need information about the degradation mechanism [6]. Autoregressive (AR), autoregressive and moving average (ARMA), autoregressive

integrated moving average (ARIMA) models are widely used in modeling and estimating time series data.

Autoregressive (AR): The AR model is a time series model that uses the measured values in the previous steps as input to a regression equation to estimate the value in the next steps. It is one of the recommended methods for the analysis of this kind of data.

The AR time series prediction model can be used for real-time prognostications of Li-Ion batteries. An improved AR time series model is proposed in [16] for prediction of RUL. Another study on RUL prognostics is proposed an autoregressive model improved with particle swarm optimization (PSO) [17].

Autoregressive and Moving Average (ARMA): Autoregressive moving averages models are a time series estimation and prediction method and are applied in time series data observed at equal time intervals. When a time series data is given, the ARMA model is used to understand or even predict the future values of the series. The model consists of two parts: The autoregressive part (AR), the moving averages part (MA). In the time series analysis, the moving average model (MA model) is a common approach in the modeling of univariate time series. The moving average model indicates that the output variable is a linearly-defined term and is dependent on various historical values.

Autoregressive Integrated Moving Average (ARIMA): ARIMA is a popular and widely used statistical method for time series estimation. It is a model class that captures different standard temporary structures in time series data. There is an integral (I) section which is a model that uses the difference of raw observations between the autoregressive part (AR) and the moving averages (MA).

ARIMA has been used frequently in many studies for RUL estimation. Also, some studies compared Autoregressive Integrated Moving Average method with other techniques like Extended Kalman Filtering (EKF), Relevance Vector Machine, Particle Filter (PF) [18].

CONCLUSION

Data-driven methods are widely used for RUL estimation. These methods will continue to be developed with new approaches in the future. In this study, we have briefly touched on data-driven methods that include machine learning and statistical methods. In our next study, we will conduct a research on stochastic methods.

REFERENCES

- [1] K.C. Kapur, and M. Pecht. *Reliability Engineering*, John Wiley & Sons, 2014.
- [2] M. G. Pecht, and M. Kang, *Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things*, John Wiley and Sons, 2018.
- [3] C. Hubbard, J. Bavlsik, C. Hegde and C. Hu, "Data-Driven Prognostics of Lithium-Ion Rechargeable Battery using Bilinear Kernel Regression" in *Annual Conference of The Prognostics and Health Management Society*, 2016.
- [4] X. Si, W. Wang, C. Hua, M. Chen and D. Zhou, "A Wiener-process-based degradation model with a recursive filter algorithm for remaining

- useful life estimation”, *Elsevier: Mechanical Systems and Signal Processing*, Vol. 35, pp. 219-237, February 2013.
- [5] O. F. Eker, F. Camci and I. K. Jennions, “Major Challenges in Prognostics: Study on Benchmarking Prognostics Datasets”, *European Conference of Prognostics and Health Management Society*, 2012
 - [6] C. Su and H. J. Chen, “A review on prognostics approaches for remaining useful life of lithium-ion battery”, *IOP Conf. Series: Earth and Environmental Science* 93, 2017
 - [7] C. Chen, B. Zhang, G. Vachtsevanos and M. Orchard, “Machine condition prediction based on adaptive neuro-fuzzy and high-order particle filtering.”, *IEEE Transactions on Industrial Electronics* 58, pp. 4353–4364, 2011
 - [8] I.A. Basheer and M. Hajmeer, “Artificial neural networks: fundamentals, computing, design, and application” in *Journal of Microbiological Methods* 43, pp 3–31, 2000
 - [9] D. Yanga, Y. Wanga, R. Pana, R. Chenb and Z. Chen, “A neural network based state-of-health estimation of lithium-ion battery in electric vehicles”, *Elsevier: Energy Procedia*, Vol. 105, pp. 2059-2064, May 2017
 - [10] L. W. Kang, X. Zhao, and J. Ma, “A new neural network model for the state-of-charge estimation in the battery degradation process”, *Elsevier: Applied Energy*, Vol. 121, pp. 20-27, May 2014
 - [11] J. Wu, C. Zhang and Z. Chen, “An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks”, *Elsevier: Applied Energy*, Vol. 173, pp. 134-140, July 2016
 - [12] V. N. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.
 - [13] M. A. Patil, P. Tagade, K. S. Hariharan, S. M. Kolake, T. Song, T. Yeo and S. Doo, "A novel multistage Support Vector Machine based approach for Li ion battery remaining useful life estimation", *Elsevier: Applied Energy*, vol. 159, pp. 285-297, December 2015
 - [14] M. E. Tipping, “The relevance vector machine,” in *Advances in Neural Information Processing Systems*, vol. 12, Cambridge, MA: MIT Press, pp. 652–658, 2000
 - [15] D. Liu, J. Zhou, D. Pan, Y. Peng and X. Peng, “Lithium-ion battery remaining useful life estimation with an optimized Relevance Vector Machine algorithm with incremental learning”, *Elsevier: Measurement*, vol. 63, pp. 143-151, March 2015
 - [16] D. Liu, Y. Luo, Y. Peng, X. Peng and M. Pecht, “Lithium-ion Battery Remaining Useful Life Estimation Based on Nonlinear AR Model Combined with Degradation Feature” in *Annual Conference of Prognostics and Health Management Society*, 2012
 - [17] B. Long, W. Xian, L. Jiang and Z. Liu, “An improved autoregressive model by particle swarm optimization for prognostics of lithium-ion batteries”, *Elsevier: Microelectronics Reliability*, vol. 53, Issue 6, pp. 821-831, June 2013
 - [18] B. Saha and K. Goebel, “Comparison of Prognostic Algorithms for Estimating Remaining Useful Life of Batteries”, *Transactions of the Institute of Measurement and Control*, Vol: 31 issue: 3-4, pp. 293-308, 2009

Game Design for Rehabilitation of Children with Disabilities by Using Depth Sensor

Pembe Buse KAYALI¹, Caner OZCAN² and Rafet DURGUT³

¹ Karabuk University, Karabuk, Turkey, busekyl-95@hotmail.com

² Karabuk University, Karabuk, Turkey, canerozcan@karabuk.edu.tr

³ Karabuk University, Karabuk, Turkey, rafetdurgut@karabuk.edu.tr

Abstract - In this study, a game has been designed to provide rehabilitation for children with disabilities using Microsoft Kinect V2 which is a low-cost depth sensor. In order to keep the physical activity at the top, the exercise sets are intended to be carried out without loss of motivation. Kinect sensor is used to determine the joint points of the persons and transfer them to the characters within the game. The game is designed on the Unity3D, and the character in the game is an open source application called MakeHuman. The actions that people should perform are determined by the rule-based approach.

Keywords – Kinect, exercise, game, rehabilitation, motion recognition.

I. INTRODUCTION

One of the biggest problems in the implementation of the necessary exercise sets for people with disabilities or those who should perform physical therapy exercises is lack of motivation. For this reason, there is a big difference between expected and obtained performance from the exercise sets, which are realized as a result of the unwillingness and strain, especially in young children [1]. A gamification model is used to eliminate this problem [2]. Exercise sets that people need to perform are embedded in a game and are expected to perform them in a fun way. Thanks to this game design, the actions that were previously challenging and boring can be made more fun and enjoyable.

Many studies have been carried out to develop and evaluate motor skills of children with disabilities [3]. In these applications, the body joint points are determined to obtain action information from the user. This is usually done with RGB and RGB-D cameras. When the RGB cameras are used, the main problems such as the effects of the ambient light and the difficulties of background extraction are tried to be eliminated by using the depth sensor [4].

In particular, with the introduction of the Microsoft Kinect product in 2010, the use of depth sensors has increased significantly in human skeleton monitoring [5]. Kinect equipment can be used in homes and clinics because of its low cost. Although the Microsoft Kinect sensor is intended to be used in entertainment, it has applications in the field of military [6], engineering [7] and health [8]. In these studies, the position of the joint points on the 3D space in the human body is determined with the help of Kinect and presented as an

introduction to the developed game or application.

Particularly in the field of health, the Kinect sensor is often used to help with treatments [9]. In addition to orthopedic disorders such as paralysis, upper extremity problems and the recovery of lost motor skills, it is also actively used in the treatment of cognitive diseases such as Alzheimer [12] and Cerebral Palsy. In addition, Kinect hardware can be used in medical areas other than physical therapy as well as in the recovery stages after the heart attack [14]. Thanks to the developed application, joint positions are monitored for knee rehabilitation and contribute to the treatment process [15]. The findings showed that wearable sensors and Microsoft Kinect have acceptable accuracy. Methods for classifying the fall risk of individuals and focusing on the preventive measures under the highest risk are used to identify the most important management strategies to protect resources and prevent suffering [16]. Unsigned motion capture data is used to assess balance, core, and limb performance using time and velocity-dependent measurements [17]. It was determined that a depth sensor combined with the proposed frame could determine the performance subtleties between groups. In another study, it is used in physical therapy after hip surgery [18]. Therapists stated that the exercise games met more than motor rehabilitation criteria, and that they aimed to continue to use the game as part of their rehabilitation treatment.

In this study, a depth sensor supported game design has been realized in order to increase the physical activities of children with disabilities and to strengthen their motor skills. The characters in the game can be controlled by body joints. The angles of the joint points of the character are calculated and expected to reach the desired angle value transferred to the game using rules.

The application presented in this paper differs from other studies. In the designed game, the accuracy of the actions is done by angle determination. Thanks to real-time monitoring in the designed game, the character can access the actions of the player and check the accuracy of the angle. The columns and characters used in the design of the game enable the player to understand when the game starts and how the player moves.

II. MOTION DETECTION WITH MICROSOFT KINECT

In 2010, with the depth sensor, which was introduced by Microsoft and named as Kinect 360, motion recognition and rehabilitation work in the field of computerized vision gained speed [19]. The general appearance of the sensor and the equipment included are shown in Figure 1. The rays sent by the IR emitter on the sensor are detected by the sensor on the hardware and the depth map of the media is extracted.

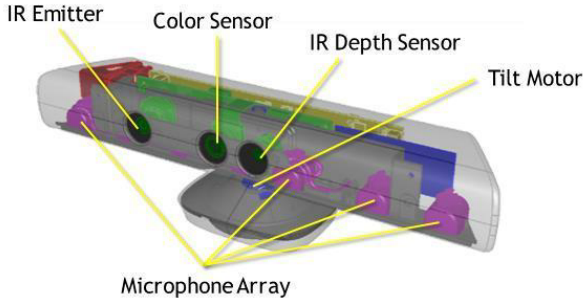


Figure 1. Internal structure of the Microsoft Kinect 360 depth sensor [14]

By using machine learning methods on the depth map, the positions of 20 joint points on the 3D space can be estimated [20]. The representation of 20 joint points is given in Figure 2. In Kinect v2 version, the number of these joint points has been increased to 25.

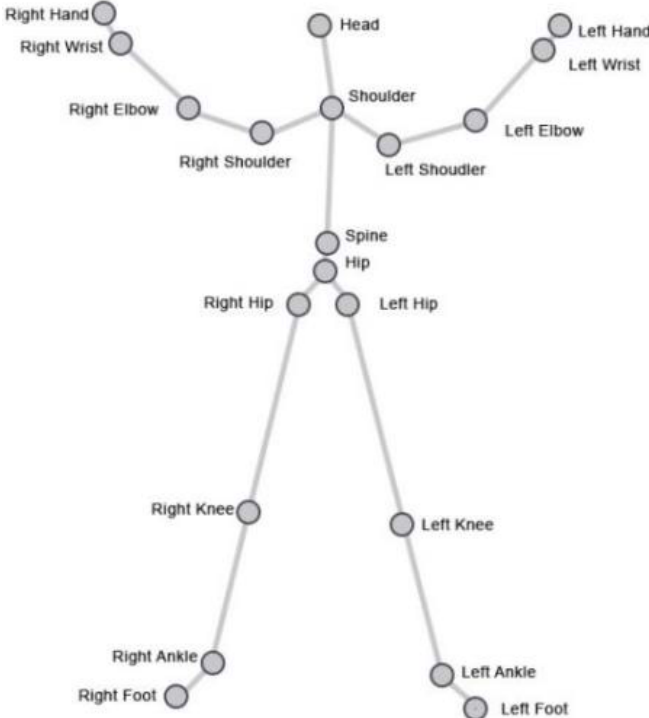


Figure 2. Kinect human body joints [14]

III. APPLICATION DESIGN

In this study, in order to support the physical rehabilitation of children with disabilities, 3D game design has been realized which can use the human body as a natural interface with the depth sensor. The main purpose of the application is to make the basic exercise sets that children need to practice by making them entertaining and increasing their physical activities. The exercise kits defined in the application for rehabilitation of disabled children are given in Figure 3.

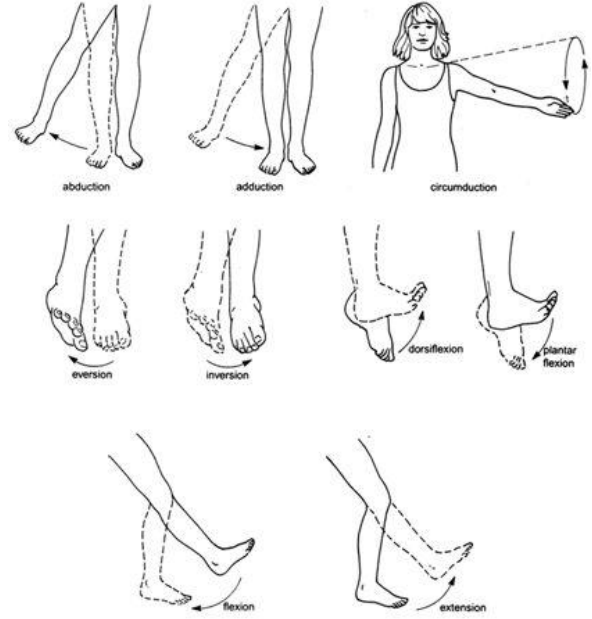


Figure 3: Exercise sets within the application

The general flow diagram of the application is given in Figure 4. When the application is first opened, the information of the exercise sets registered in the database is loaded. The user is then expected to be positioned opposite the sensor. When the person passes through the sensor, the exercise to be done to the user is determined and displayed in the game. The person is expected to perform the exercise within the specified time period. In order to reach the knowledge that exercise is done, joint angles are calculated by using cosine theorem with the Eq.1.

$$\cos\theta = \frac{x_1 \cdot y_1 + x_2 y_2 + \dots + x_n y_n}{\sqrt{x_1^2 + x_2^2 + \dots + x_n^2} + \sqrt{y_1^2 + y_2^2 + \dots + y_n^2}} \quad (1)$$

Joint angles are processed in the rule-based classifier to determine whether the correct exercise is performed. If the exercise is not performed within the specified time, the user will be shown the error message and the next exercise set will be displayed. The success message is displayed if it is performed correctly. After the person's score is increased, the next exercise set is started. Animations have been added to objects to make the game interesting. These actions contribute to the development of the mental monitoring mechanism by attracting user attention. The blocks and small characters in the application move up and down, indicating which action will be made in the game.

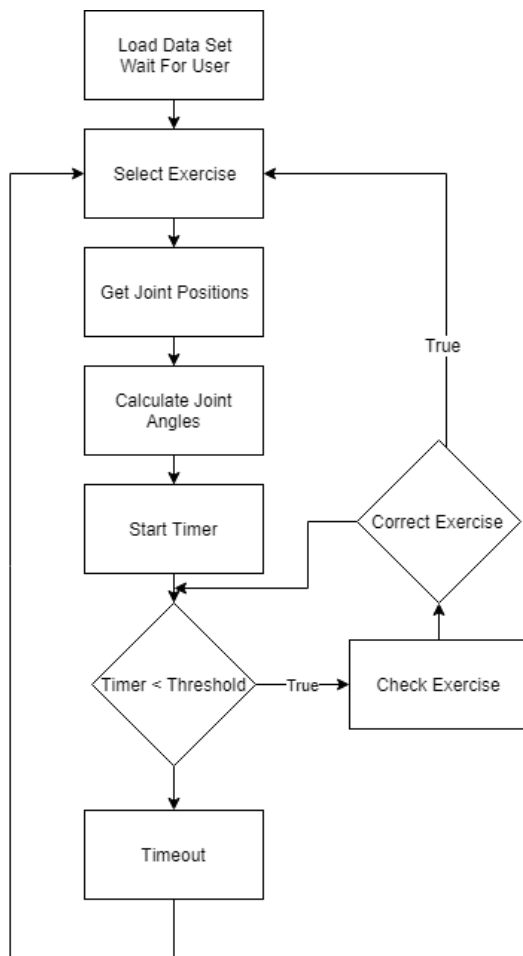


Figure 4: Flow diagram of the application

The theme of the game is thought of as a scene of columns and small characters. Columns and small characters in the theme of the game have a continuous action mechanism as long as the game continues. Thanks to this actions, the player is doing brain training by following the actions. It is designed to perform physical therapy actions correctly and to increase body coordination with the actions mentioned in the columns. The action of objects in the game is designed to create a more fun environment for the user. The visual interface of the application is prepared in 3D as shown in Figure 5.

IV. AVATAR DESIGN

The avatar design, which is used to visualize the action within the game, has been made with the open source application called MakeHuman. Many different human modeling can be done within the application. The MakeHuman approach uses common parameters such as height, weight, gender and muscularity. The development of MakeHuman is derived from a detailed technical and artistic study of the morphological features of the human body. It has been developed under different operating systems to make it available on all computers. It was developed with an architecture that includes OpenGL and Qt support using the Python programming language to work on all operating systems. [21].



Figure 5. Game Home Screen View

V.CONCLUSION

In this study, a game has been developed to contribute to the physical rehabilitation process of children with disabilities. Microsoft Kinect v2 depth sensor is used to use user bodies as a natural interface. Based on the information obtained from the sensor, it is determined whether the users do the exercises correctly. Users are congratulated as a result of correct actions and motivated for incorrect exercises. It is ensured that the individuals receive the highest efficiency from the rehabilitation process by keeping them highly motivated.

In our future works, a separate playground will be designed for each category of disabilities, not as general actions that people with disabilities have difficulty in making. In these areas, actions determined by experts will be used for each category of disabilities. These actions will be divided into small sub-designs as easy, medium and difficult and the game will be played in stages.

ACKNOWLEDGMENT

This study was supported by The Scientific and Technological Research Council of Turkey (TUBITAK) under 2209-A Research Projects Support Program for University Students (App. No: 1919B011802741). The authors thank TUBITAK for the financial and scientific support.

REFERENCES

1. Chang, Yao-Jen, Shu-Fang Chen, and Jun-Da Huang. "A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities." *Research in developmental disabilities* 32.6 (2011): 2566-2570.
2. Paavola, Jessica M., Kory E. Oliver, and Ksenia I. Ustinova. "Use of X-box Kinect gaming console for rehabilitation of an individual with traumatic brain injury: a case report." *Journal of Novel Physiotherapies* 3.01 (2013): 1-6.
3. Meleiro, Pedro, et al. "Natural user interfaces in the motor development of disabled children." *Procedia Technology* 13 (2014): 66-75.
4. Biswas, Kanad K., and Saurav Kumar Basu. "Gesture recognition using microsoft kinect®." *The 5th international conference on automation, robotics and applications*. IEEE, 2011.
5. Zhang, Zhengyou. "Microsoft kinect sensor and its effect." *IEEE multimedia* 19.2 (2012): 4-10.
6. Lun, Roanna, and Wenbing Zhao. "A survey of applications and human motion recognition with microsoft kinect." *International Journal of Pattern Recognition and Artificial Intelligence* 29.05 (2015): 1555008.
7. Jia, Weidi, et al. "3D image reconstruction and human body tracking using stereo vision and Kinect technology." *2012 IEEE International Conference on Electro/Information Technology*. IEEE, 2012.
8. Kamel Boulos, Maged N. "Xbox 360 Kinect exergames for health." *Games for Health: Research, Development, and Clinical Applications* 1.5 (2012): 326-330.
9. N. Silberman, D. Sontag, and R. Fergus. Instance segmentation of indoor scenes using a coverage loss. In *Computer Vision—ECCV 2014*, pages 616–631. Springer, 2014.
10. M. Abdur Rahman, A. M. Qamar, M. A. Ahmed, M. Ataur Rahman, and S. Basalamah. Multimedia interactive therapy environment for children having physical disabilities. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval, ICMR '13*, pages 313–314. ACM, New York, NY, USA, 2013.
11. Y.-J. Chang, S.-F. Chen, and J.-D. Huang. A kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in Developmental Disabilities*, 32(6):2566 – 2570, 2011. ISSN 0891-4222.
12. J. Chapinal Cervantes, F. L. G. Vela, and P. P. Rodríguez. Natural interaction techniques using kinect. In *Proceedings of the 13th International Conference on Interacción Persona-Ordenador, INTERACCION '12*, pages 14:1–14:2. ACM, New York, NY, USA, 2012. ISBN 978-1-4503-1314-8.
13. Luna-Oliva, Laura, et al. "Kinect Xbox 360 as a therapeutic modality for children with cerebral palsy in a school environment: a preliminary study." *NeuroRehabilitation* 33.4 (2013): 513-521.
14. LaBelle, K. (2011). Evaluation of Kinect joint tracking for clinical and in-home stroke rehabilitation tools. Undergraduate Thesis, University of Notre Dame.
15. Naeemabadi, M. R., Dinesen, B., Andersen, O. K., Najafi, S., & Hansen, J. (2018). Evaluating Accuracy and Usability of Microsoft Kinect Sensors and Wearable Sensor for Tele Knee Rehabilitation after Knee Operation. In *BIODEVICES* (pp. 128-135).
16. Using data from the Microsoft Kinect 2 to determine postural stability in healthy subjects: A feasibility trial *PLoS One*. 2017; 12(2): e0170890. Published online 2017 Feb 14. doi: 10.1371/journal.pone.0170890
17. Daniel Leightley and Moi Hoon Yap, Digital Analysis of Sit-to-Stand in Masters Athletes, Healthy Old People, and Young Adults Using a Depth Sensor *Healthcare*; 6(1): 21, 2018.
18. Ling, Y., Ter Meer, L. P., Yumak, Z., & Veltkamp, R. C. (2017). Usability test of exercise games designed for rehabilitation of elderly patients after hip replacement surgery: pilot study. *JMIR serious games*, 5(4).
19. Garduño-Ramón, M. A., L. A. Morales-Hernández, and R. A. Osorio-Rios. "Morphological filters applied to Kinect depth images for noise removal as pre-processing stage."
20. Shotton, Jamie, et al. "Real-time human pose recognition in parts from single depth images." (2011): 1297-1304.
21. M. Bastioni, S. Re, S. Misra. "Ideas and methods for modeling 3D human figures: the principal algorithms used by MakeHuman and their implementation in a new approach to parametric modeling". *Proceedings of the 1st Bangalore Annual Compute Conference, Compute 2008*.

Prioritizing Manual Black-box Tests Using Topic Modeling

O. C. SENEL¹, M. H. OZCANHAN¹ and M. ERSAHIN²

¹ Dokuz Eylul University, Izmir/Turkey, onur.senel@ceng.deu.edu.tr

¹ Dokuz Eylul University, Izmir/Turkey, hozcanhan@cs.deu.edu.tr

² Commencis Technology, Izmir/Turkey, mustafa.ersahin@commencis.com

Abstract - Software testing is commonly used for validating software changes but it is the most expensive phase of the software development life cycle (SDLC). Test case prioritization (TCP) aims reducing the cost by scheduling the running order of tests to increase the effectiveness of testing; so that most beneficial test cases are executed first and faults are detected in the early phases of testing. In this paper, we propose a new static TCP technique for manual black-box testing. We use a topic modeling (TM) algorithm to extract the functionalities of each test script. This approach allows differentiating and ranking test cases. Cases that test different parts of the system under test (SUT) get higher ranks. Our approach is compared with the manually prioritized test cases of test engineers in the case study of a commercial online banking project. The comparison shows that the average percentage of fault detection (APFD) rates of our approach is higher than the manual prioritization approach.

Keywords – Test case prioritization, topic modeling, black-box testing, software testing.

I. INTRODUCTION

NEW versions of application software are released frequently which always include some changes, in the original work. Changing some parts of software could cause new bugs. The testing process of the software development life cycle (SDLC) is used to assure the quality of the software. Every part of the software should be tested before the release of the new version. Automated or manual testing can be used to ensure the software runs correctly. Both of these techniques are time-consuming and costly. Due to the project's budget, there isn't enough time to run all test cases then fix bugs and re-run test cases again. That's why the detection of the failures in the early stage of the testing process is very critical for saving time. Test case prioritization (TCP) is a technique that sorts test cases for execution to reduce testing cost.

In our study, we focus on manual black-box testing. In the black-box testing, there are test case scripts written in natural language. These scripts contain required information (preconditions, steps to perform etc.) to perform a test case. Test engineers perform test cases one by one to test software. Since we are focusing on manual black-box testing, the only data are test case scripts, there are no test and source codes as in the white-box unit testing. Maximizing coverage and diversifying test cases are two common objectives of a TCP technique in the literature. Since we don't have any coverage information of

test cases, we choose a diversity-based TCP technique. The motivation behind diversifying test cases is the fact that similar test cases detect the same faults. A diverse set of test cases has the probability of detecting different faults, hence more number of faults. To diversify test cases, we use a text mining concept, called topic-modeling (TM). The idea of topic modeling-based TCP is that assign each test case to a topic. If you do not test a topic, you will not detect failures related to that topic. Furthermore, you should choose tests as much as diverse from each other to ensure test different parts of the software. We use Latent Dirichlet allocation (LDA) technique to generate our topic model [1]. We applied LDA on test scripts to extract the topics for each test case. After that, we have topic membership probability vectors for each test case. We use these vectors to calculate the distance between test cases also we choose Manhattan distance as our distance metric. To rank test cases, we calculate distances between all test case pairs and store these distances in a matrix. Then, we add a test case that has a maximum average distance to all others to prioritized tests list. This greedy approach prioritizes all test cases. We evaluate our approach on a commercial online banking project test cases and compare results to manually prioritized test cases by test engineers. The average percentage of fault detection (APFD) rates of our approach is higher than the manual prioritization approach. This higher score shows that faults are detected earlier when TCP technique is used for manual testing. As we mentioned before, it is very important and cost-effective to find faults earlier. According to this reality, TCP reduces testing costs of software significantly.

II. RELATED WORK

In general, there are five categories of TCP techniques based on the available data: white-box execution-based, black-box execution-based, grey-box model-based, white-box static and black-box static prioritization. In addition to these categories, there are two different maximization strategy can be used in TCP: maximizing coverage and diversifying test cases.

In execution-based approaches, the execution information of the test cases is required. Some of the previous studies prioritize test cases by maximizing source code coverage. They mostly use statement-level execution information in the source as the input prioritization. Wong et al. add source code change information between versions to the prioritization process to give high priority to test cases that likely related to changed parts of the source code

T11406092 Displaying User Phone Entry Screen (Macintosh; Intel MAC OS Sierra 10.12.6; Firefo...	
Preconditions	
C3115219	
Steps	
Step	Expected Result
1 When user presses on Accept Button on previous screen that Terms and Conditions Screen.	<p><i>If startSession() service call is successful;</i></p> <p>User Phone Entry Screen is displayed as below and also numeric keyboard is opened automatically.</p>

Figure 1: A sample test case script from our experiment.

[2]. Other studies use diversification to prioritize test cases. Simao et al. use a neural network to find the most dissimilar test cases [3]. Yoo et al. differentiate test cases' execution profiles by a clustering algorithm [4]. Execution-based based prioritization studies are generally in the white-box category. Sampath et al. present a black-box execution-based TCP technique [5]. They use user activity logs as the execution information in web applications. Accessing the source code is not required by using these logs.

Model-based TCP techniques use the specification models of the source code and test cases (e.g. UML state diagrams, paths on the state diagram). Since the source code of the system under test (SUT) is not required in this technique, all model-based TCP techniques are gray-box. In some projects, execution information may not be available therefore model-based approaches can be used in these projects. Hemmati et al. and Korel et al. proposed model-based TCP techniques using specification models of the source code [6, 7]. Korel et al. try to maximize the coverage of the model by computing the difference of the model between two versions [7]. Hemmati et al. use diversity-based algorithms to compute the similarity between test cases' paths in the state model then prioritize test cases by their paths similarity [6].

In a static prioritization approach, execution information and specification models are not required. Requires the source code of the SUT and test cases or test scripts. Zhang et al. propose a call graph-based technique as a white-box static prioritization [8]. This a coverage-based prioritization technique uses the static call graph of test cases. A greedy algorithm prioritizes test cases by maximizing the number of source code methods that are covered by the test case. Ledru et al. propose another technique that is string-based black-box static TCP technique [9]. They measure the distance between test cases to identify their similarity. They compare several distance metrics such as Euclidean, Manhattan, Levenshtein and Hamming and find that the Manhattan distance has the best average performance for fault detection. Ledru et al. also use a greedy algorithm to maximize diversity between test cases.

There are a few studies that only use test scripts as the input of the prioritization process. Hemmati et al. interested in the TCP problem in the context of manual

system-level black-box testing which is our focus in this paper also [10]. They use test scripts written in natural language (e.g. instructions in English) from the Mozilla Firefox projects. They implement several existing TCP techniques and adapt them to the domain of black-box system-level test prioritization. They use the APFD metric for comparing the effectiveness of these TCP techniques.

In this paper, we focus on black-box static prioritization in the context of manual testing. We implement a topic-based TCP technique. We use test scripts from a commercial online banking project which is being developed with Agile methodologies. We will compare results to manually prioritized test cases according to the APFD values in the results section.

III. THE PROPOSED TOPIC-BASED TCP

In using a topic-based TCP approach, our aim is to prioritize test cases effectively and help test engineers to save time when running the tests of a project. TM is a method to abstract textual data from any text document collection. Since it does not require any training data, TM can be applied to test cases of any project, quickly. In addition, it is fast enough to handle numerous documents.

The focus in our work is on test case scripts rather than test case source codes. Figure 1 shows the sample test case of our experiment, which is testing a user phone number entry in an online banking web application. The tests usually explain the test steps to perform in natural language (e.g. instructions in English). Some preprocessing to the text of test cases is the primary step. For example, removing special characters and common English stop words. Then we stem each word into its base form. Finally, we convert each word to lowercase. Our proposed method uses the preprocessed text as input data to apply the TM algorithm for extracting topics. We prefer one of the most commonly used model in natural language processing, namely the LDA model. And, we use Gensim package of Python programming language as our LDA implementation tool [11].

LDA needs the number of topic parameter, K. Choosing the optimal value of K is another research decision matter. Larger K values produce finer-grained topics and smaller K values produce coarser-grained topics. Previous studies in software engineering, commonly used K values ranging

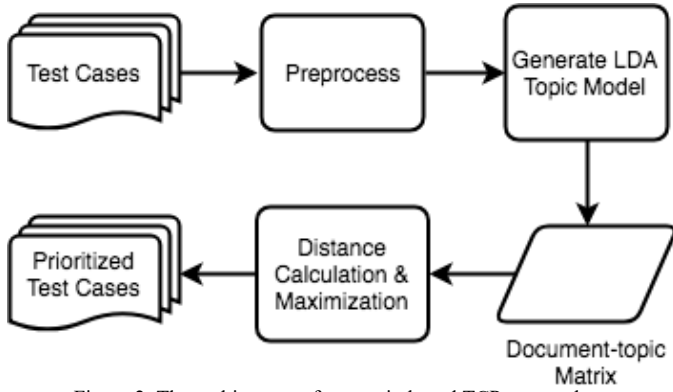


Figure 2: The architecture of our topic-based TCP approach.

from 5 to 500 [12, 13]. Thomas et al. use $K = N/2.5$ to produce medium-grained topics, where N is the number of documents in the SUT [14]. Similarly, we choose K as $N/2.5$ rounding to the nearest integer. We use the default values in Gensim package for other parameters of LDA.

After applying LDA, we have a document-topic matrix that represents the topic membership probability of each test case. We need to calculate distances between each test case using their topic membership probability values. Ledru et al. found that the Manhattan distance metric is optimal for string-based TCP, therefore, we use the same metric [9]. We implement a greedy maximization algorithm. This algorithm finds the most dissimilar test case from all test cases firstly and adds this test case to an initially empty list. Next, it finds the case that is most dissimilar to previously prioritized test cases and adds this test case to the same list. This algorithm continues until all test cases have been prioritized. Figure 2 summarizes the steps in our TCP implementation.

IV. EXPERIMENTAL RESULTS & DISCUSSION

In this section, we compare our topic-based approach to manual prioritization.

Test cases and their results are stored in TestRail test case management tool [15]. To collect the faults, we looked at the test execution results in TestRail. Each test case has been exercised by testers, and the results have been reported in TestRail. Test engineers create different test runs in TestRail for each alpha, beta and release candidate testing. These test runs contain manually prioritized test cases in order. We collect 10 different tests run using TestRail API. Table 1 shows the successful and failed test cases count of these runs.

TABLE 1: TEST CASE COUNT RESULTS OF TEST RUNS.

Test Run	Successful Test Cases	Failed Test Cases	Total	Percent Faults
1	152	21	173	12.14
2	138	13	151	8.61
3	100	6	106	5.66
4	45	2	47	4.26
5	36	2	38	5.26
6	34	1	35	2.86
7	31	2	33	6.06
8	12	16	28	57.14
9	17	1	18	5.56
10	12	1	13	7.69

TABLE 2: MEAN APFD VALUES OF THE EXPERIMENT.

Test Run	APFD	
	Manual	Topic-based
1	39.51	49.21
2	43.89	59.30
3	17.77	51.89
4	60.64	68.16
5	38.16	86.89
6	1.43	91.24
7	48.48	76.11
8	46.43	50.28
9	36.11	72.78
10	3.85	92.05

APFD metric is used to test the effectiveness of TCP techniques, which is originally introduced by Rothermel et al. in [16]. APFD captures the average of the percentage of faults detected by a prioritized test suite. APFD is given by

$$APFD = 100 * \left(1 - \frac{TF_1 + TF_2 + \dots + TF_m}{nm} + \frac{1}{2n} \right) \quad (1)$$

where n is the number of test cases, m is the number of faults and TF_i is the number of tests, which must be executed before fault i is detected. Larger APFD values indicate that increasing the effectiveness of TCP technique. In other words, more faults are detected with fewer test cases.

Mean APFD values of two different techniques are presented in Table 2.

The results show that our proposed TCP method performs better than manual prioritization on average by a margin of 36%. We got higher mean APFD values on all test runs in our case study. This means if we prioritize test cases with this method before the execution, we can catch faults with executing fewer test cases. The results of the sixth and tenth test run is quite striking. They have one faulty test and that test is the last executed one in the manual prioritization. Our TCP method puts it to the first order in the prioritized list.

TM is based on a machine learning algorithm and LDA uses a random seed to extract topic from a text collection. That's why LDA produces slightly different results in different runs. To solve this problem in our case study, we run our TCP method 30 times for every test run and take the average of the results.

V. CONCLUSION & FUTURE WORK

Delivering high quality softwares is an important goal for all software developers. To achieve their goal, too many test cases need to be executed on every new software release. Manual black-box testing is one of the most common software testing type, which has a high cost and takes too much time a. Few TCP studies focus on making manual black-box software testing more efficient. In our work, we focus on black-box static prioritization in the context of manual testing. The experimental results show that the efficiency of the software testing processes is increased significantly by using the proposed TCP method.

The proposed TCP method provides better fault detection rate on the testing process. The efficiency of the testing process is very critical because all possible faults should be detected with as few tests as possible within a limited testing budget.

In future work, we want to integrate information retrieval (IR) approaches to our TM approach. Currently, we don't use any source code change information between versions when prioritizing test cases. We plan to use IR approaches to capture the relation between source code changes and test cases. Using this relation, we can find the related test cases for the new version of the software automatically, then prioritize them using our topic-based TCP method.

REFERENCES

- [1] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," in *Advances in neural information processing systems*, 2002, pp. 601–608.
- [2] W. E. Wong, J. R. Horgan, S. London, and H. Agrawal, "A study of effective regression testing in practice," in *Proceedings The Eighth International Symposium on Software Reliability Engineering*, Albuquerque, NM, USA, 1997, pp. 264–274.
- [3] A. Silva Simao, R. De Mello, and L. Senger, "A Technique to Reduce the Test Case Suites for Regression Testing Based on a Self-Organizing Neural Network Architecture," in *30th Annual International Computer Software and Applications Conference (COMPSAC'06)*, Chicaco, IL, 2006, pp. 93–96.
- [4] S. Yoo, M. Harman, P. Tonella, and A. Susi, "Clustering test cases to achieve effective and scalable prioritisation incorporating expert knowledge," in *Proceedings of the eighteenth international symposium on Software testing and analysis - ISSTA '09*, Chicago, IL, USA, 2009, p. 201.
- [5] S. Sampath, R. C. Bryce, G. Viswanath, V. Kandimalla, and A. G. Koru, "Prioritizing User-Session-Based Test Cases for Web Applications Testing," in *2008 International Conference on Software Testing, Verification, and Validation*, Lillehammer, Norway, 2008, pp. 141–150.
- [6] H. Hemmati, A. Arcuri, and L. Briand, "Achieving scalable model-based testing through test case diversity," *ACM Trans. Softw. Eng. Methodol.*, vol. 22, no. 1, pp. 1–42, Feb. 2013.
- [7] B. Korel, G. Koutsogiannakis, and L. H. Tahat, "Model-based test prioritization heuristic methods and their evaluation," in *Proceedings of the 3rd international workshop on Advances in model-based testing - A-MOST '07*, London, United Kingdom, 2007, pp. 34–43.
- [8] L. Zhang, J. Zhou, D. Hao, L. Zhang, and H. Mei, "Prioritizing JUnit test cases in absence of coverage information," in *2009 IEEE International Conference on Software Maintenance*, Edmonton, AB, Canada, 2009, pp. 19–28.
- [9] Y. Ledru, A. Petrenko, S. Boroday, and N. Mandran, "Prioritizing test cases with string distances," *Autom. Softw. Eng.*, vol. 19, no. 1, pp. 65–95, Mar. 2012.
- [10] H. Hemmati, Z. Fang, M. V. Mäntylä, and B. Adams, "Prioritizing manual test cases in rapid release environments," *Softw. Test. Verification Reliab.*, vol. 27, no. 6, p. e1609, Sep. 2017.
- [11] R. Řehůřek and P. Sojka, "Gensim—statistical semantics in python," *Stat. Semant. Gensim Python LDA SVD*, 2011.
- [12] T. L. Griffiths, M. Steyvers, and J. B. Tenenbaum, "Topics in semantic representation," *Psychol. Rev.*, vol. 114, no. 2, p. 211, 2007.
- [13] S. W. Thomas, "Mining Software Repositories with Topic Models," *Tech Rep 2012-586 Sch. Comput. Queen's Univ.*, p. 43, 2012.
- [14] S. W. Thomas, H. Hemmati, A. E. Hassan, and D. Blostein, "Static test case prioritization using topic models," *Empir. Softw. Eng.*, vol. 19, no. 1, pp. 182–212, Feb. 2014.
- [15] "Test Case Management & Test Management Software Tool - TestRail." [Online]. Available: <https://www.gurock.com/testrail>. [Accessed: 28-Feb-2019].
- [16] G. Rothmel, R. H. Untch, Chengyun Chu, and M. J. Harrold, "Prioritizing test cases for regression testing," *IEEE Trans. Softw. Eng.*, vol. 27, no. 10, pp. 929–948, Oct. 2001.

Increasing Air Traffic Control Efficiency With Integrated Human Machine Interface

Busra TUZLUPINAR and Baha SEN

Abstract—The intensive use of air transportation vehicles increases the air traffic intensity day by day. Air traffic control is provided by air traffic control systems in a regular, fast and compatible manner. Flight sorting, flight plan preparation, flight arrival and departure commands are handled and supervised by the controller. The purpose of the controllers is to be quick, efficient and stable by ensuring that no mistakes are made. For this reason, large workload falls on the controllers. Our aim to make the air traffic control system faster and real in the electronic environment in order to reduce this large load on the controllers.

Index Terms—ATC, 3D Display, ADS-B, Asterix, HMI.

I. INTRODUCTION

Air Traffic Control (ATC) is to prevent aircraft from colliding with each other or any other obstacles on the The controller considers the aircraft movements and positions to manage air traffic in four dimensions along with the time dimension. It provides the aircraft with the necessary instructions in accordance with the multi-dimensional thinking. landing, departure or flight part; Air Traffic Service (ATS) type to provide fast and regular air traffic flow. The system is responsible for tracking, detecting, and descending a flight.

The controllers are responsible for managing all the situations between the take-off and landing of the aircraft, while maintaining the vertical and horizontal safety distance between them in all stages of their flights, other flights, obstacles and other vehicles at another point where the aircraft want to reach from one point to another. The controller considers the aircraft movements and positions to manage air traffic in four dimensions along with the time dimension. It provides the aircraft with the necessary instructions in accordance with the multi-dimensional thinking.

Air traffic control systems provide a 2 dimensional image. The height of the 2D image planes cannot provide the difference in height between each other. For this reason, the controllers read the information in the 2D image and interpret the height calculation so that they can think about the 3 dimensional image. When the traffic density is too high, it is difficult to separate the planes in the layers from each other, to view the collision situations, and to calculate the height difference between each other in the 2D image.

3D display support to see difference between other aircraft, how the plane passes over the mountain, easily decision making for controller any conflict will occur. Also when the controller follow the aircraft, they can see its departure,

landing and holding positions. 3D display greatly reduces workload of controllers by providing a more real image.

II. GENERATING 3D MAP

In this project, we decided to use of Unity for simulation of 3D view. Firstly, we need to generate 3D Turkey map for real display of air space. To obtain of this map, an elevation data set that's stored in ASCII (American Standard Code for Information Interchange) was converted to height map image. Height map image stores height values as pixel values. These height maps known as digital elevation model (DEM). By converting the elevation data in ASCII format to DEM format, the loss of image values between the pixels was prevented and a more realistic image was obtained.

For generating map in the Unity, some calculations was done. Because of Unity scene starting index is not same with elevation data set starting index. For each pixel, x and z axis values added to data set value. Then obtained value was divided into the highest value in the data set. These values were imported to a 3D terrain object and a map view was created.

III. PARSING ASTERIX FLIGHT DATA

ADS-B (Automatic Dependent Surveillance - Broadcast (ADS-B)) provides to tracking of air craft, gives an information about its position in real time. It collects air craft data via satellite navigation. ADS-B based on Communication, Navigation, and Surveillance (CNS) technologies all over the world. ADS-B provides pilots and air traffic controllers to control aircraft. It continuously broadcast air craft information to other air crafts and ground systems. Ground systems transmit the air craft's position to air traffic controller.

ASTERIX (All-purpose structured EUROCONTROL surveillance information exchange) is standard for exchange of information between ATM applications. To communicate with different users, the data classified to data categories. The propose of this categorization is easy identification of data. Because, each numerical data gives different information about the air craft. Data categories from 000 to 127 for civil and military applications, 127 to 240 for special military applications, 241 to 255 used for civil and military applications which are non-standard applications.

For this project, CAT021 Asterix real time data used. All air craft data are stored in astfin format. CAT 021 received tract data from ADS-B radar. The category consists of data blocks. A Data Block contains one or more record(s) of data. This category gives an information about air craft id, air craft position as longitude and latitude, its speed vector etc. To get particular air craft information, data file decoded in Java.

B. TUZLUPINAR is with Institute of Science, Computer Engineering, University of Ankara Yildirim Beyazit, 06010 Ankara, Turkey.

B. SEN is with Institute of Science, associate professor of Computer Engineering, University of Ankara Yildirim Beyazit, 06010 Ankara, Turkey.

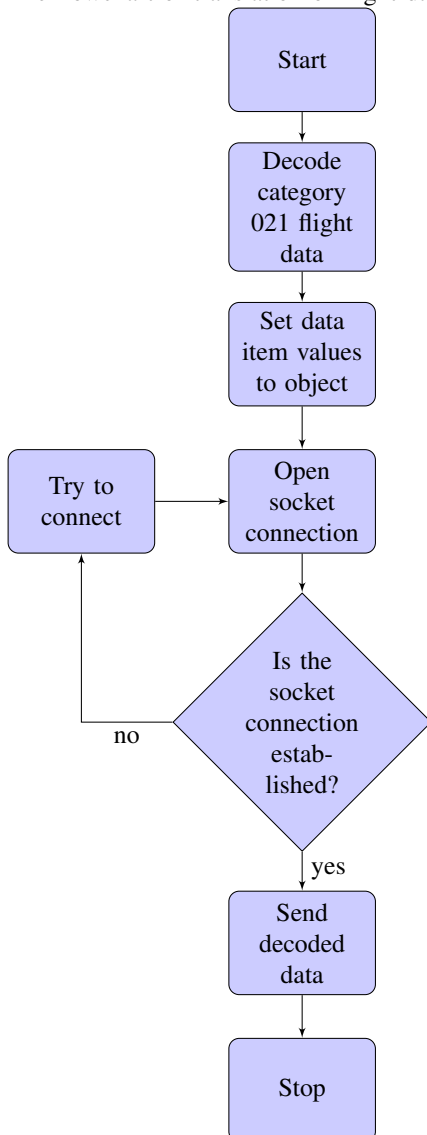
TABLE I
USED ASTERIX CATEGORY 021 DATA ITEMS

Category Number/Data Item	Description of Standard Data Items
Data Item I021/130	Position in WGS-84 Co-ordinates
Data Item I021/145	Flight Level
Data Item I021/146	Selected Altitude
Data Item I021/150	Air Speed
Data Item I021/161	Track Number

IV. TRANSLATION OF DECODED FLIGHT DATA

We decoded air craft information in Java language. But our simulation platform for 3D display is Unity platform and Unity uses C Sharp language. For this reason, we need to convert decoded data from Java to C Sharp. To converting issue, the WebSocket communication protocol was used to stream data with a single TCP connection. In C Sharp platform, one socket is opened and started to listen client to receive the data. After connection is establish from both Java and C Sharp sides, client started to sending data to server continuously with timer.

The flowchart of translation of flight data.



V. MOVING AIR CRAFT

3D air plane objects added to our scene which includes 3D elevation map. Received individual data imported to these 3D air plane object. As long as data received, air planes can move over the map. Adjusting of camera position, zooming in or out, we can see the distance between air craft or mountains. Also we can track one selected air craft by setting camera to air plane object.



Fig. 1. Example of a 2D display of air traffic control area

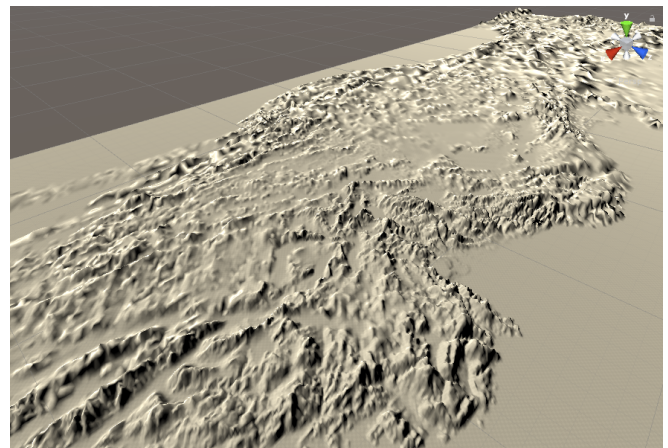


Fig. 2. Generated 3D map

VI. RELATED WORKS

The propose of this paper [1] presents benefits of effective and usable design in air traffic control tower. For this reason, they developed prototype of integrated decision support systems. These prototypes can be mention in three categories: A surveillance display, flight data manager and supervisor display. To enhance the tower systems, some features were added. Such as, giving different color for each small, large or heavy air crafts, to easy distinguish aircraft and surface, they are colored white and cyan colors. They also mentioned about, electronic-flight-data-display system. This system gives departing and landing information about for each aircraft. The system allows to controllers manage of flight data via touch screen. They improved prototype based on user requirements. This article highlights the importance of easy and usable systems and how they can be improved.

Researches conducted enhance our understanding of Air Traffic Management. The content of the research follows sequence from general to specific topics. The featured topics focus of recent researches are category of air traffic control, communication in air traffic control systems, human-system interaction and 3D - 2D displays for ATC that it is subset of air traffic control display literature.

Luigi Mazzucchelli and Antonio Monteleone mention about the combining of ground and on-board systems for ATC [2]. D4 technology supports management of big scale of real time data with visual reality systems. The system provides virtual display of air space with compatible monitors and glasses. In this context, radar tracks, flight plans, weather information and elevation data compute in distributed machines. To sending the information between distributed machine, they used TCP/IP protocol. Hence, they enable 3D representation of combined data to controllers in tower and approach positions.

The paper of [3] gives an overview about the 3D-in-2D displays for ATC that were implemented in the Augmented Reality. In this paper, they show the differences between 2D and 3D displays. Compare the advantage and drawback aspects of both display systems. In their project, they developed combination of 3D and 2D view of air traffic. With AR technology, they developed a simple ATC simulator. AR technology provides the controllers to explore the air space, quickly zoom in-out using their hand gestures. Thus, they aim to minimize the effort consumed by the controllers.

This article [4] describes the automated scenario production for the training of air traffic controllers in the air control system simulation project. Air traffic controllers need to give fast decision and correct due to their jobs. In the training of these persons, the situations that they may face from the take-off in flight are limited with the experience of their instructors. Therefore, they have developed a system that produces automatic scenario for every situation in the air traffic field. These scenarios also include conflicts that may occur in the take-off, landing and en route positions of aircraft.

VII. CONCLUSION

This paper presented our 3D work of ATC System. In that work, it was mentioned that infrastructure of the ATC System which are 3D elevation map, ADS-B radar, Asterix flight data, Asterix data item decoding and transformation of the data to 3D display. Our approach is improving ATC Systems issues on these infrastructures. For this purpose, we try to understand of the ATC system, then used 3D computing algorithms, techniques and methods. In our research, we collected flight data. Then, we made this data meaningful in the air traffic field. Consequently, we improved 3D air traffic control area to reduce the workload of the controllers by passing 3D display in air traffic control area.

REFERENCES

- [1] H.J.D. Reynolds, K. Lokhande, M. Kuffner, S. Yenson "Human-Systems Integration and Air Traffic Control" 2012.
- [2] L. Mazzucchelli, A. Monteleone "A 4D Immersive Virtual Reality System for Air Traffic Control".
- [3] B.L. William Wong, S. Rozzi, A. Boccalatte, S. Gaukrodger, P. Amaldi, B. Fields, M. Loomes, P. Martin "3D-in-2D Displays for ATC" 14 November, 2007.
- [4] M. Ulvi, M. Ozturk, O. Ince, G. Sinar, I. Tasdelen "Otomatik Hava Trafiklari Uretilmesi Yazilimi Calismasi".
- [5] EUROCONTROL "EUROCONTROL Specification for Surveillance Data Exchange ASTERIX Part 12 Category 21 ADS-B Target Reports" 07 July, 2014
- [6] EUROCONTROL "Model for Task and Job Descriptions of Air Traffic Controllers" 15 March, 1996.
- [7] Y. Cao, D. Sun "A Parallel Computing Framework for Air Traffic Flow Management" 01 July, 2011.
- [8] Y. Cao, D. Sun "A Parallel Computing Framework for Large-Scale Air Traffic Flow Optimization" December, 2012.
- [9] B. Sridhar, G.B. Chatterji "Computationally Efficient Conflict Detection Methods for Air Traffic Management" 1997.
- [10] R.J. Reisman, D.M. Brown "Design of Augmented Reality Tools for Air Traffic Control Towers" 25-27 September, 2006.

Comparison of Classification Algorithms in Terms of Model Performance Criteria in Data Mining Applications

E. KABULLAR¹ and F. ATASOY²

¹ Karabuk University, Karabuk/Turkey, elifkabullar@karabuk.edu.tr

² Karabuk University, Karabuk/Turkey, ferhatatasoy@karabuk.edu.tr

Abstract – Process of data mining is discovering anomalies, patterns and correlations from a large amount of data. Looking at today's data technologies is one of the most up-to-date machine learning methods. At the same time, data mining is the biggest solution for the effective processing of data that is parallel to the development of computer technologies. For this reason, it is important to use techniques that can process large data sets. The accuracy of these inferences; reveals the success of the created model on the data. Therefore, it is very important in terms of the success of the application in which data mining applications produce better results.

In this study, models were created by using various classification methods on bank marketing data set and performance comparisons (accuracy, precision, sensitivity, f-criteria) were made in terms of the data set model performance criteria. K-nearest neighborhood, Naïve Bayes, Support Vector Machine and Logistic Regression methods were used as classification method. As a result of the comparison, the most successful estimation method is logistic regression model. This algorithm is followed by SVM, K-NN and Naïve Bayes algorithm according to accuracy success rate, respectively.

Keywords – Classification, data mining, K-NN, logistic regression, model success criteria, Naive Bayes, SVM.

I. INTRODUCTION

NOWADAYS, the amount of data used together with rapidly developing computer systems is increasing rapidly.

Therefore, it becomes more difficult to examine the data. According to the studies, the amount of data increases twice each year. Data mining applications are rapidly evolving in recent years to analyze data and create meaningful information [1]. Data mining is derived from artificial intelligence, machine learning and statistics. It is also a discipline that can be used in different areas to discover unknown important information [2].

Data mining is widely used due to the need for decision-making in many areas such as marketing, education, finance, production, health, and customer relationship management [3]. According to functions, data mining models can be grouped as following:

1. Classification, Regression
2. Clustering,
3. Association Rules and Sequential Time Patterns [4].

Classification and regression models, estimating, clustering; patterns of association and consecutive-time patterns are descriptive models [5].

In this study, models were created by using various classification methods on Bank Marketing data set and performance comparisons (accuracy, precision, sensitivity, f-criteria) were made in terms of the data set model performance criteria. As classification methods; K-nearest neighborhood (K-NN), Logistic Regression (LR), Naïve Bayes (NB) and Support Vector Machine (SVM) methods were used.

II. RELATED WORKS

There are many different studies in the literature for comparison of data mining algorithms. In some of them existing algorithms were compared with existing or newly developed algorithms. In studies on the acceptability of the developed algorithm; the developed algorithm is evaluated by comparing the results with the results of the other studies with different known data sets.

When the data mining applications for the banking sector are examined, it is observed in the literature that credit card fraud detection, credit card spending, customer groups are formed, and credit demands are evaluated [5].

In 2013 Xiong et al. [6] developed a system based on personal bankruptcy by examining credit card information of bank customers with data mining. They combined conditional probability distribution model with model-based k-means algorithm to define bankruptcy features. Bad customers were identified with the bankruptcy features.

Banking is an important area and different studies on it were investigated in [7].

III. DATASET

The data set is related to the direct marketing campaigns of a Portuguese banking institution over the phone [1]. The general purpose of these phone calls is to sell a bank time deposit to the customer. The aim of this study is to predict whether customers will subscribe to this deposit based on this data set.

When we look at the results; we will try to estimate one of

the two classes. “Yes” or “No”, the customer will subscribe or not.

IV. CLASSIFICATION ALGORITHMS

A. K-Nearest Neighborhood

K-NN algorithm is one of the simplest and most widely used classification algorithms. The value of K is determined in the algorithm's operation. The meaning of K value refers to the number of elements to be looked at. When a value is received, the distance between the nearest K and the incoming value is calculated. Euclidean function is usually used in the calculation of distance. Euclidean function as an alternative to Manhattan, Minkowski and Hamming functions can be used. After the distance is calculated, distances are sorted, and value is assigned to the appropriate class depends on minimum value.

Let us assume that a new member has arrived according to the K-NN method as shown in Figure 1. Let us determine the 3 nearest neighbors to which this new member is closest. Since two of the 3 closest members are red round members, we can classify our new member in this way.

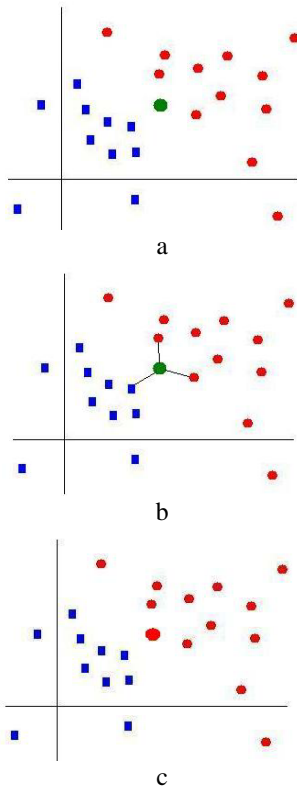


Figure 1: Classification with K-NN [2], a) New data point, b) calculation distances between new data and known points, c) inserting new data point in the nearest class.

B. Logistic Regression

Logistic Regression is a data analysis method to classify binary dependent variables (Yes / No, Male / Female, Fat / Weak, etc.) which has relations one or more independent variables. It is used to classify categorical or numerical data.

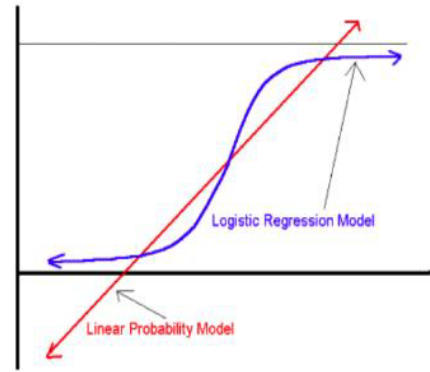


Figure 2: Comparison of linear versus logistic regression [3].

It is widely used in linear classification problems. Therefore, it is very similar to linear regression. Figure 2 shows the comparison graphic. Figure 3 represents how LR algorithm classifies entry visually.

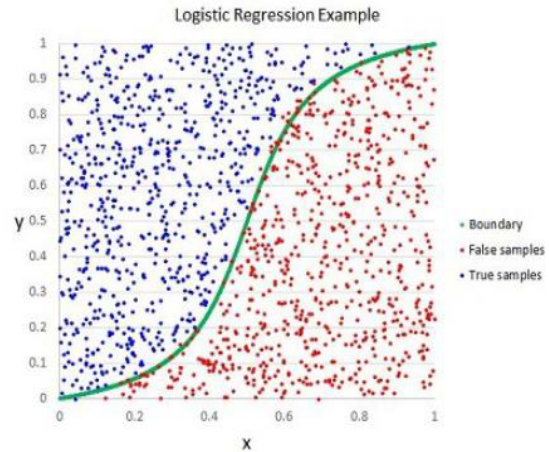


Figure 3: Classification with LR [3].

C. Support Vector Machine

SVM is a similar classification algorithm with LR. They both try to find the best line that separates the two classes. The algorithm allows the line to be drawn in two classes of the line to be drawn away from the most distant places. It is a non-parametric classifier. SVM can also classify linear and nonlinear data, but usually try to classify data as linear.[7] Figure 4 shows SVM algorithm classification visually.

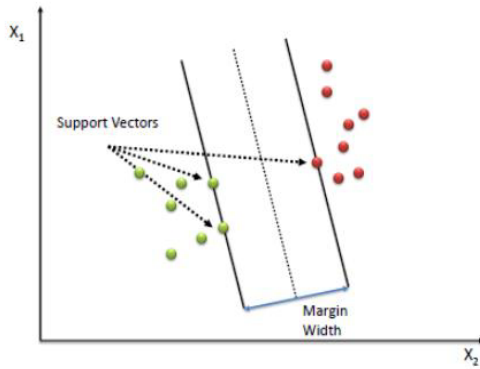


Figure 4: Classification with SVM [2].

D. Naïve Bayes

In the NB classification, the system learns certain rate of data that includes known outputs (class or category) for known inputs. The classes or categories of the data must be given the system. The data is processed by the probabilistic operations and when the new test data presented to the system is operated according to the previously obtained probability values and the category of the test data is tried to be determined. As the number of sample data in training increases, the success rate for classifying of test data increases.

V. MODEL SUCCESS CRITERIA

The basic methods used for evaluating model success are precision, error rate, sensitivity and F-criteria.

The number of samples assigned to the correct and wrong class state success of the model.

The confusion matrix provides the achievement information of the results obtained from test. In confusion matrix, rows show certain numbers of the samples in the test set, while columns show the estimation results of the model.

Table1: Confusion matrix [5].

Correct Class	Projected Class	Class 1	Class 0
	Class 1	a	b
	Class 0	c	d

a: TP (True Positive) b: FN (False Negative) [4]

c: FP (False Positive) d: TN (True Negative) [4]

A. Accuracy - Error rate

Accuracy rate is the most prominent and simple method for measuring model performance. The ratio of the number of accurately classified samples (TP + TN) to the total number of samples (TP + TN + FP + FN). The error rate is the value of this value. In other words, the ratio of the number of incorrectly classified samples (FP + FN) to the total number of samples (TP + TN + FP + FN) [4,5]. Equations are shown in (1) and (2).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$\text{Error rate} = \frac{FP + FN}{TP + FP + FN + TN} \quad (2)$$

B. Precision

The precision is the ratio of the number of True Positive samples estimated as class 1 to the number of all instances estimated as class 1. It is calculated by using (3).

$$\text{Precision} = \frac{TP}{TP + FN} \quad (3)$$

C. Sensitivity

Sensitivity given in (4), defines the ratio of correctly classified positive samples over the total number of positive samples:

$$\text{Sensitivity} = \frac{TP}{TP + FP} \quad (4)$$

D. F-Criteria

The precision and sensitivity criteria alone are insufficient to make a significant comparison. Combining both criteria together provides more accurate information. The f-criteria is defined for this purpose. The harmonic mean of sensitivity and precision is called F-criteria and it is calculated with using (5).

$$F - \text{criteria} = \frac{2 \times \text{Sensitivity} \times \text{Precision}}{\text{Sensitivity} + \text{Precision}} \quad (5)$$

VI. RESULTS AND DISCUSSIONS

The results of the models are given in Table 2 as following:

Table2: Classification results for bank data.

Algorithms	Accuracy	Precision	Sensitivity	F-criteria
K-NN	0.891	0.743	0.638	0.670
SVM	0.9071	0.816	0.907	0.712
LR	0.9073	0.818	0.667	0.711
NB	0.841	0.650	0.690	0.665

In Table 2, the classification results of the methods are given for Bank data. According to the results, the best results belong to the LR with 0,907 estimate accuracy. The accuracy performance order of other algorithms is SVM, K-NN, and NB respectively. The best result belongs to LR for precision with 0.818. The precision performance order of other algorithms is SVM, K-NN, and NB respectively. SVM has the best result for sensitivity and other algorithms is SVM, K-NN, and NB respectively.

Finally; according to f-criteria SVM is the best with 0.712 and the other algorithms follow it in order LR, K-NN, and NB respectively.

The best method for data mining depends on data set, feature extraction method, sampling, etc. With this study classification methods are implemented, and results are presented.

REFERENCES

- [1] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: A review," *Egypt. Informatics J.*, vol. 19, no. 3, pp. 179–189, 2018.
- [2] D. R. Carvalho and A. A. Freitas, "A genetic-algorithm for discovering small-disjunct rules in data mining," *Appl. Soft Comput. J.*, vol. 2, no. 2, pp. 75–88, 2002.
- [3] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, pp. 32–35, 1950.
- [4] C. Coşkun and A. Baykal, "Veri Madenciliğinde Sınıflandırma Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması," *Akad. Bilişim'11 - XIII. Akad. Bilişim Konf. Bildir.*, pp. 51–58, 2011.
- [5] J. C. Culberson, B. P. Feuston, V. Svetnik, C. Tong, A. Liaw, and R. P. Sheridan, "Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling," *Journal of Chemical Information and Computer Sciences*, vol. 43, no. 6, pp. 1947–1958, 2003.
- [6] T. Xiong, S. Wang, A. Mayers, and E. Monga, "Expert Systems with Applications Personal bankruptcy prediction by mining credit card data," *Expert Syst. Appl.*, vol. 40, no. 2, pp. 665–676, 2013.
- [7] Aslı ÇALIŞ, Sema KAYAPINAR, and Tahsin ÇETİNYOKUŞ, "Veri Madenciliğinde Karar Ağacı Algoritmaları ile Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama," *Endüstri Mühendisliği Derg.*, vol. 25, no. 21, pp. 2–19, 2014.
- [8] P. Moro ve R. Cortez and P. "Bank Marketing Data Set", June 2014. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>. [Accessed: 2 April 2019].
- [9] S. Sayad, "An Introduction to Data Science", [Online]. Available: https://www.saedsayad.com/support_vector_machine.htm. [Accessed: 2 April 2019].
- [10] M. Nandu, "Introduction to Logistic Regression", [Online]. Available: <https://blog.goodaudience.com/machine-learning-using-logistic-regression-in-python-with-code-ab3c7f5f3bed>. [Accessed: 2 April 2019].
- [11] E. ŞİRİN, "http://www.datascience.istanbul/", [Online]. Available: <http://www.datascience.istanbul/2017/07/07/hata-matrisi-confusion-matrix-python-uygulama/>. [Accessed: 2 April 2019].
- [12] K. Markham, "data school", 25 May 2014. [Online]. Available: <https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/>. [Accessed: 2 April 2019].
- [13] Ş. E. ŞEKER, "KNN (K nearest neighborhood, en yakın k komşu)", [Online]. Available: <http://bilgisayarkavramlari.sadievrenseker.com/2008/11/17/knn-k-nearest-neighborhood-en-yakin-k-komsu/>. [Accessed: 2 April 2019].

Privacy Scoring Over Professional OSNs: More Central Users are Under Higher Risk

Y. KILIC¹ and A. INAN²

¹ Adana Alparslan Turkes Science and Technology University, Adana/Turkey, ykilic@atu.edu.tr

² Adana Alparslan Turkes Science and Technology University, Adana/Turkey, ainan@atu.edu.tr

Abstract – Privacy scoring aims at measuring the privacy violation risk of a user over an online social network (OSN) based on attribute values shared in the user's OSN profile page. Existing work in the field work with possibly biased or emotional survey data. In this study, we work with real-world data from the professional LinkedIn OSN and empirically show that centrality of an OSN user is positively correlated with privacy scores. This finding contradicts existing survey-based work that suggest negative correlation with centrality.

Keywords - privacy scoring, online social networks (OSN), indirect privacy survey, network centrality, LinkedIn

I. INTRODUCTION

ACCORDING to recent statistics, the use of Online Social Network (OSN) services is ever increasing [1]. OSNs have turned into a vivid, digital reflection of our real life activities and social interactions [2]. Such popularity has created room for specialization in OSN services: Facebook and Google+ are mostly used for personal purposes such as acquiring new friends or connecting with old ones, Twitter is preferred for commenting on public events and news, Instagram is the focal point of sharing pictures and last but not the least, LinkedIn caters for professional goals such as exploring possible job opportunities and strengthening ties with past and present colleagues and peers.

Various data leakage incidents, one of the latest being the “Cambridge Analytica Scandal of Facebook” [3], have raised concerns on how the content posted on these platforms should be maintained towards protecting individual privacy. The right to privacy is recognized globally by the Universal Declaration of Human Rights. Protection in digital domains is guaranteed by additional layers of protection such as The European Union's General Data Protection Law (GDPR) and Turkey's Regulation of Personal Data Protection. However, these laws and regulations become void when users share consentingly on OSNs. Therefore, it is vital that OSN users be informed of the potential risks of sharing data over OSNs.

Notice that every user's risk is unique to herself and depends on the type of data shared, and the audience the data are shared with. Recent work on the subject [4-10] tries to quantify the overall risk on a personal basis and calls this a user's *privacy risk score*. The concept is adopted from credit risk scores of the

financial system. Higher privacy risk scores imply being more prone to a privacy violation.

In this study, we focus on the problem of privacy scoring over professional OSNs –specifically, the LinkedIn OSN, which is the most prominent, popular example of its kind. There are many reasons why professional OSNs deserve special attention:

- Most users of professional OSNs reveal their true identity. This is a basic necessity for benefiting from the OSN services. Fake, anonymous or pseudonymous accounts are very unlikely.
- Data posted on professional OSNs tend to be more truthful than other OSNs. No company would hire someone whose graduation dates on her LinkedIn profile disagrees with her diplomas.
- The privacy paradox [11] is at its peak in professional OSNs: as a user shares more, she gets more visible but also more prone to privacy violation. Visibility is key to benefiting from professional OSNs.

Existing work on privacy scoring over OSNs rely on survey data. Given a list of profile items such as mother's maiden name, phone number etc., the users are asked to specify how visibly they would share the item – with no one, only friends, friends of friends or every one. Our results are based directly on actual user profiles, collected automatically using a crawler program that visited the LinkedIn pages of 5,389 distinct users. As a result, we experiment with observed (versus alleged) behavior of OSN users. According to Braunstein et al., survey results might be biased [12]. We also reach much larger a sample size – Liu and Terzi surveyed only 153 users [5], Pensa and di Blasi surveyed only 101 users [13].

The main contributions of our study are as follows:

1. We study privacy scoring over professional OSNs for the first time. No previous study has focused on professional OSNs.
2. We rely on real world data and not user surveys.
3. We experiment on a much larger data set compared to existing work.
4. We show that over professional OSNs, privacy risk is positively correlated with network centrality. Existing work have claimed negative correlation on survey data.

The rest of the paper is organized as follows. We review related work in Sec. II. Material and methods are presented in Sec. III and the results of empirical experiments on real world OSN data are discussed in Sec. IV. We conclude in Sec. V.

II. RELATED WORK

The first study on privacy scoring of OSN users is due to Liu and Terzi [4, 5]. Their model is inspired by the Item Response Theory (IRT) [14] of psychology. IRT is applied mainly in educational testing. Every question q has a level of difficulty β_q and discrimination α_q . Primary goal of IRT is to model whether or not an examinee e will correctly answer a given question q based on the personal ability level θ_e of e , and parameters (β_q, α_q) of question q .

Liu and Terzi's privacy scoring approach draws the following analogy between educational testing and OSN usage: OSN users represent examinees and attributes that could be shared on user profiles represent questions. A user is assumed to pick the correct privacy setting for an item based on her ability level and the item's privacy sensitivity and discrimination power. When user decisions are binary (e.g., either share the profile item with everyone or no one – mapping to a true/false question), then the problem is said to be *dichotomous*. Otherwise, if there are more than 2 possible answers/privacy settings, the problem is said to be *polytomous*. A user's privacy score is simply set as the cumulative product of her privacy settings multiplied by the corresponding item's sensitivity [4].

This initial IRT solution by Liu and Terzi [4] has laid the foundation for various others [5-10]. In [5], Liu and Terzi incorporate in their model, the underlying OSNs structure. This is achieved by multiplying each item's visibility by the information propagation factor that represents the fraction of OSN users that are given access to the shared data through the chosen privacy setting. It is also shown in [5] that level of information propagation is negatively correlated with a user's privacy score. The faster any information shared by user u spreads on the OSN, the lower will be this user's privacy score. Srivastava and Geethakumari [6] introduce a measuring approach to privacy scoring for unstructured contents (post, reply, etc.) called *Privacy Leakage*. The approach extracts sensitive information (e.g., e-mail, birthday, age, etc.) from the non-structured contents. They then calculate the privacy risks according to the sensitivity values of the extracted information. Domingo-Ferrer [7] proposes a score called *Privacy Risk Functionality* which measures how much users share information compared to other users. It has been argued that the privacy risks of a user may vary depending on the sharing of other users.

Nepali et al. [8] name their score the *Privacy Index*. The score calculates how sensitive is the information shared by the users. Sramka argues that auxiliary information should be incorporated in user scores [9]. This involves locating the evaluated users through simple web searches or joining personal data obtained from several OSNs.

Petkos et al. [10] do not only regard the information declared by users for privacy measuring, but also consider the information that can be obtained through inference. They also provide an open source framework for privacy measurement.

It is important to emphasize once again that all of these studies are based on the initial work by Liu and Terzi [4]. Another common point is that work in [4-10] rely entirely on privacy

survey results, provided as an $n \times m$ response matrix for n items and m users. The survey asks each user to set the audience she would share her item with. Braunstein et al. criticizes this approach as survey results tend to be emotional or biased [12]. Our work separates from the approach of Liu and Terzi's [4] (and their followers [5-10]) because (i) we use real data instead of survey results, (ii) we work with professional OSNs, and (iii) we leverage structural network properties like Page Rank [16] and Eigen Vector [17] centrality of a user's node.

Very few studies utilize the OSN topology in measuring privacy risks of users [5, 13, 15]. Bioglio and Pensa [15] investigates how users' connections or neighbors in the network affect their privacy risk. Pensa and di Blasi shows in [13] that IRT based scoring mechanism in [4] can be expanded with the Page Rank score of the corresponding user. Supporting Liu and Terzi's finding in [5], they also show that privacy score of a user is negatively correlated with her Page Rank score.

III. MATERIAL AND METHODS

A. Materials

We developed a web crawler to collect OSN data from LinkedIn. Towards this end, we used the Python language and imported the Selenium library. Data collected between May 2018 and August 2018 were modelled and queried with the relational MySQL and the graph-based Neo4J database management systems.

The crawler visited all directly reachable OSN user profiles rooted at the first author's Linked account. In breadth-first traversal order, a total of 5,389 distinct profiles have been accessed. At each profile page, our program recorded every bit of visible data including education history, work experience, phone number, e-mail address and social relations such as connections. The complete list of attributes is given in Table 1.

Table 1: LinkedIn profile items from the dataset

Item No	Item Name
1	Birthday
2	Connections
3	Company Web Site
4	Personal Web Site
5	Phone Number
6	Also Viewed Users List
7	Education
8	Work Experiences
9	User Interests
10	E-mail
11	About
12	Address

Collected data has been transformed to a response matrix as follows. For each of the 12 profile items in Table 1, if the value of user u on an item i is visible from the crawler's (i.e., the first author's) viewpoint, we have set the corresponding item's response cell $R[i, j]$ to 1. Otherwise, we set the value to 0. This binary response matrix corresponds to the dichotomous problem setting in [4].

B. Methods

Liu and Terzi provide a clear, concise baseline model for privacy scoring called the Naïve model [4]. This model, together with their IRT solution is the basis of almost all studies on the problem of privacy scoring over OSNs, including this study. We first introduce this Naïve model, then the IRT model and finally the Page Rank solution of [13]. All notation used during this presentation is summarized in Table 2. For brevity, we exclude the polytomous problem setting.

Let PS^j denote the privacy score of user j . It is widely accepted in the literature that PS^j depends on

- Visibility of profile item i on user j 's profile: denoted with $V(i, j)$, visibility indicates how large a portion of the network is able to observe the value of i on the profile of j .
- Sensitivity of profile item i : denoted with β_i , sensitivity of an item represents how large a portion of the network considers i critical in protecting their privacy.
- Topology of the underlying OSN.

Table 2: Terms and their notation

Name	Description
PS^j	Privacy score of user j
β_i	Sensitivity of item i
$V(i, j)$	User j 's visibility on item i
β_i^{Naive}	Sensitivity of item i in the Naïve model
β_i^{IRT}	Sensitivity of item i in the IRT model
PS_{Naive}	Privacy score in of the Naïve model
PS_{IRT}	Privacy score of the IRT model
PR_C	Privacy score of the Page Rank model
E_C	Eigen Vector centrality of an OSN user

Proposed methods differ in how visibility or sensitivity of an item is to be measured and whether or not network topology should be involved. Yet, the basic formula for calculating privacy score PS^j of a user j rarely changes:

$$PS^j = \sum_i \beta_i \times V(i, j) \quad (1)$$

Eq. 1 reflects that making visible sensitive items will increase the privacy score of a user and makes great sense.

The Naïve model sets $V(i, j)$ based entirely on the row i and column j of the response matrix. The visibility formula is given in Eq. 2, where $P(i)$ is the probability of observing 1 on row i and $P(j)$ is the same on column j . Notice that the two are assumed to be independent.

$$V^{Naive}(i, j) = P(i) * P(j) \quad (2)$$

Eq. 3 explains how β_i is computed in the Naïve model. Sensitivity of an item is simply the proportion of users that have hidden away the item in their profiles. Here $|R_i|$ denotes the number of users sharing i on row i of the response matrix R .

$$\beta_i^{Naive} = \frac{N - |R_i|}{N} \quad (3)$$

The IRT model approximates the item characteristics curve that best represents the response matrix R to yield $(\alpha_i, \beta_i^{IRT})$ – discrimination and sensitivity of item i and θ_j – ability of user j . Then user j 's visibility on item i can be calculated according to Eq. 4.

$$V^{IRT}(i, j) = \frac{1}{1 + e^{-\alpha_i(\theta_j - \beta_i^{IRT})}} \quad (4)$$

In [5], Liu and Terzi adds to the IRT model a network topology measure inspired by information propagation theory. The term f_j represents the proportion of network users that can view the value of item i on user j 's profile. Eq. 4 is rewritten as Eq. 5.

$$V^{IRT-IC}(i, j) = V^{IRT}(i, j) \times f_j \quad (5)$$

Pensa and di Blasi suggest computing V^{IRT} according to Eq. 4, and then propagating privacy scores within each user's neighborhood according to the Page Rank algorithm by Brin and Page given in [16].

Among these methods, we have implemented the Naïve model, the IRT model and Page Rank model. We also implemented the Eigen Vector centrality measure of [17].

IV. EXPERIMENTS

We first compared the sensitivity values of the profile items listed in Table 1. The results are depicted in Figure 1. Notice that the item sensitivity values for β_i^{Naive} of Eq. (3) and β_i^{IRT} of Eq. (4) are quite distant. Recall that the sensitivity of an item determines the extent to which the item increases the privacy risk of the user sharing it publicly. For example, intuitively, the phone number is expected to be more sensitive than education. Both models respect this anticipation. However, notice also that in the Naïve model, the two items are only marginally different in item sensitivity. The IRT model's performance in sensitivity measurement is much better because this model captures each user's sharing attitudes into account.

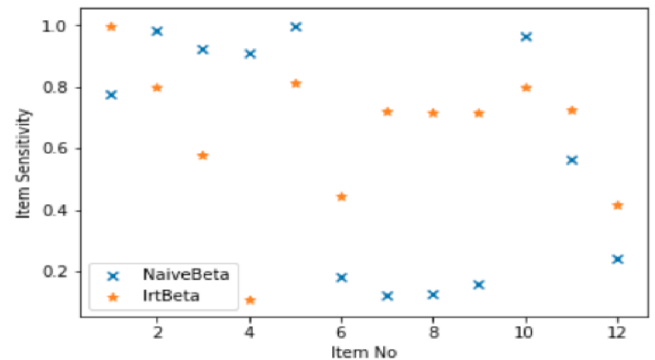


Figure 1: Item Sensitivity Per Item

Our next experiment involved goodness of fit tests. We relied on chi-square hypothesis testing to determine which of the models (e.g., PS_{Naive} and PS_{IRT}) better fits the underlying response matrix. We took the degrees of freedom to be $K-1$ for PS_{Naive} and $K-2$ for PS_{IRT} where K is the number of groups.

Under conditional independence assumption over items, we repeated the goodness of fit tests for each of the 12 items listed in Table 1. Table 3 summarized the results. Regardless of the number of groups (please refer to [4] for details of grouping and group invariance properties of IRT), more IRT hypothesis were accepted than Naïve hypothesis. This indicates that *PS_IRT* fits the response matrix better than *PS_Naive*.

Table 3: Goodness of Fit Test Results

Number of Groups(K)	<i>PS_Naive</i>	<i>PS_IRT</i>
3	5	7
4	5	7
8	5	7
10	2	10
12	3	9
14	3	9

The third experiment scenario focused on the relationship between network topology and privacy scores. Figure 2 investigates the correlation between Naïve and IRT models against the Page Rank algorithm in increasing damping factor [16]. We observe that regardless of the damping factor, both Naïve and IRT based privacy scores are positively correlated with Page Rank scores.

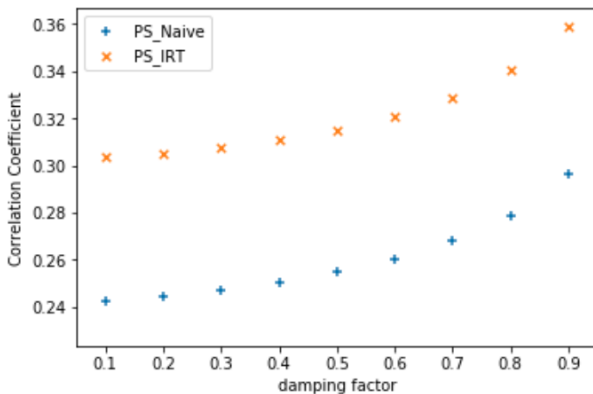


Figure 2. Page Rank Based Correlation Scores

Table 4 presents the most important results we obtained: the correlation between Eigen Vector Centrality (EVC) [17] and the Page Rank model is quite strongly positive. Notice that EVC has nothing to do with privacy scoring. Yet, higher EVC comes hand in hand with higher privacy scores, implying higher risk of privacy violation.

Table 4. Correlation Coefficient Between Centrality and Privacy Scores (Damping Factor = 85%)

Correlation Coefficient	<i>PS_Naive</i>	<i>PS_IRT</i>	<i>PR_C</i>	<i>E_C</i>
<i>PS_Naive</i>	1	0.312	0.286	0.306
<i>PS_IRT</i>	0.312	1	0.348	0.293
<i>PR_C</i>	0.286	0.348	1	0.506
<i>E_C</i>	0.306	0.293	0.506	1

To summarize, privacy attitudes of users appear to depend on their centrality in the OSN. It is possible that influential users are under higher risk because they are sharing aggressively. Unfortunately, based solely on correlation, drawing conclusions of causality between the centrality and privacy score of a user would not be correct.

V. CONCLUSION

Existing work on privacy scoring over OSNs worked with survey data that has been criticized for survey responses being emotional or biased. In this work, we collected real world OSN data from the professional LinkedIn OSN. Our empirical results over the profiles of 5,389 distinct users reveal that, (i) privacy score of a real-world OSN user can be modelled accurately using IRT, (ii) this score is correlated with the network topology, and (iii) unlike the claims of existing work, the correlation is positive. This means, more central users tend to have higher privacy scores while less central users tend to have lower privacy scores and vice versa.

Notice that our experiments only indicate correlation and not causality. In future work, we plan to more closely investigate whether there is any causality between these two dimensions as well as incorporating additional factors into the scoring mechanism such as granularity of the shares and the frequency of the shared data over the entire OSN.

REFERENCES

- [1] GLOBAL DIGITAL REPORT 2018. (n.d.). Retrieved January 25, 2019, from <https://digitalreport.wearesocial.com/>
- [2] Wani, M. A., Agarwal, N., Jabin, S., & Hussai, S. Z. (2018). Design and Implementation of iMacros-based Data Crawler for Behavioral Analysis of Facebook Users. arXiv preprint arXiv:1802.09566.
- [3] Confessore, N. (2018). Cambridge analytica and facebook: The scandal and the fallout so far. The New York Times, 4, 2018.
- [4] Liu, K., & Terzi, E. (2009). A Framework for Computing the Privacy Scores of Users in Online Social Networks. 2009 Ninth IEEE International Conference on Data Mining, 288-297.
- [5] Liu, K., & Terzi, E. (2010). A framework for computing the privacy scores of users in online social networks. ACM Transactions on Knowledge Discovery from Data (TKDD), 5(1), 6.
- [6] Srivastava, A., & Geethakumari, G. (2013, August). Measuring privacy leaks in online social networks. In 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 2095-2100). IEEE.
- [7] Domingo-Ferrer, J. (2010, October). Rational privacy disclosure in social networks. In International Conference on Modeling Decisions for Artificial Intelligence (pp. 255-265). Springer, Berlin, Heidelberg.
- [8] Nepali, R. K., & Wang, Y. (2013, July). Sonet: A social network model for privacy monitoring and ranking. In 2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops (pp. 162-166). IEEE.
- [9] Sramka, M. (2015). Evaluating Privacy Risks in Social Networks from the User's Perspective. In Advanced Research in Data Privacy (pp. 251-267). Springer, Cham.
- [10] Petkos, G., Papadopoulos, S., & Kompatsiaris, Y. (2015, August). PScore: a framework for enhancing privacy awareness in online social networks. In 2015 10th International Conference on Availability, Reliability and Security (pp. 592-600). IEEE.
- [11] Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. First Monday, 11(9).
- [12] Braunstein, A., Granka, L., & Staddon, J. (2011, July). Indirect content privacy surveys: measuring privacy without asking about it. In Proceedings of the Seventh Symposium on Usable Privacy and Security (p. 15). ACM.
- [13] Pensa, R. G., & Di Blasi, G. (2016, August). A centrality-based measure of user privacy in online social networks. In Proceedings of the 2016

- IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 1438-1439). IEEE Press.
- [14] Baker, F. B., & Kim, S. H. (2004). Item response theory: Parameter estimation techniques. CRC Press.
 - [15] Bioglio, L., & Pensa, R. G. (2017). Impact of neighbors on the privacy of individuals in online social networks. *Procedia Computer Science*, 108, 28-37.
 - [16] Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7), 107-117.
 - [17] Ruhnau, B. (2000). Eigenvector-centrality—a node-centrality? *Social networks*, 22(4), 357-365.

Recent Advances and Future Trends in Industrial Communication Networks: M2M, I-IoT and Interoperability Issues

H. SOY¹ and S. KOÇER¹

¹ Necmettin Erbakan University, Konya/Turkey, hakkisoy@erbakan.edu.tr

¹ Necmettin Erbakan University, Konya/Turkey, skocer@erbakan.edu.tr

Abstract – Over the last decade, the importance of industrial automation systems has been drastically increased to improve the productivity and to enhance the product quality in process control. Evolution of communication technologies has a strong influence on transformation of industrial automation systems. Thanks to the fourth industrial revolution, the traditional hierarchical architecture has to be replaced by distributed architecture in modern manufacturing operations. By this way, the factory floors and industrial processes will become more integrated and much smarter in order to maximize resource efficiency and profitability in near future. Machine-to-Machine (M2M) communication and Industrial Internet of Things (I-IoT) are key technologies to realizing the vision of smart factory. But, most of the field devices use vendor's proprietary protocols, and hence interoperability is often not possible entirely. The OPC UA (Object Linking and Embedding for Process Control) standard provides interoperability in industrial automation systems by creating open specifications that allows the real-time communication between field devices of different vendors. This study reviews the relevant and contemporary literature on the industrial communication networks that are shaping current automation applications, as well as the future trends in wireless networking technologies.

Keywords - M2M, I-IoT, Industry 4.0, OPC UA, TSN, industrial automation, smart manufacturing.

I. INTRODUCTION

TODAY, the manufacturing companies is toward the start of the new industrial revolution, commonly called as Industry 4.0 or Industrie 4.0, that offers many opportunities of digitalization of all stages of industrial production [1]. The main purpose of the Industry 4.0 is the emergence of digital manufacturing systems, also named as smart (connected) factory, which means networking of industrial automation systems to modernize the businesses of the companies by improving interoperability, efficiency and productivity [2, 3]. Industry 4.0 operates on several technologies, including the artificial intelligence, big data and cloud computing. But, the growth of the Machine-to-Machine (M2M) communication and Industrial Internet of Things (I-IoT) technologies are key to realizing the vision of Industry 4.0 [4-6]. Both M2M

communication and I-IoT have similar functionalities to enable the fully connected manufacturing systems [7].

M2M communication can be considered as an early form of I-IoT and allows automated data transmission between the machines and field devices (i.e., sensors, actuators and controllers) independently without human intervention. Traditionally, M2M technology focus on the industrial automation systems and telemetry applications for remote monitoring and control of single process [8]. M2M links are established both by wires as well as wireless interfaces. Due to the lower installation and maintenance costs and easier network scalability, wireless connectivity offers important cost savings compared to wired connections especially for hard-to-wire industrial applications. Wireless connectivity is often relying on cellular systems (3G, 4G LTE, LTE-A and LTE-M), Wi-Fi or another wireless link. To achieve M2M connection, the machines and field devices are equipped with embedded controller and RF transceiver module, which are specifically designed to application needs [9].

Internet of Things (IoT) is a global telecommunication infrastructure, which is used to connect the physical devices, vehicles, buildings and everyday objects to the Internet and/or other connected devices for remote monitoring and control purposes. Each physical object communicates with the existing Internet infrastructure via either wired or wireless connections. Initially, the IoT concept was implemented on commercial products. As Internet usage has grown rapidly, the IoT technology becomes an integral part of the industrial systems in growing number countries worldwide [10-12].

The I-IoT has a potential to drive the digital transformation in manufacturing by making possible to analyze data at every stage of production and extend the automated industrial processes to the cyber-physical systems (CPSs). The CPSs provide the basis for the I-IoT based applications and creates Industry 4.0-ready solutions for smart factories. In a CPS, the networked embedded devices monitor and control the physical process through closed feedback loops. Embedded devices usually have sensing and actuation functions. Thus, the computational resources and communication capabilities are integrated to interact with the physical process [13-15].

Manufacturing industry requires more intelligent and connected factory operations to provide rapid response to customer demands. Smart factories of the future need communication protocols that allow easier integration, configuration and maintenance of automation systems. There is no doubt that I-IoT and its subset M2M technologies will reshape the manufacturing ecosystem by integrating global digitalization trends within industrial plants. This study provides an overview for the current status of industrial communication networks and networking technologies. The interoperability among different communication protocols are also handled in this context.

II. INDUSTRIAL COMMUNICATION NETWORKS

The real-time monitoring of field devices and control of machines are challenging task in today's industrial processes, which consist of distributed controllers. Clearly, the efficient control of any industrial process depends on the reliable communication system. The real-time monitoring and control tasks can be achieved by exchanging data in several network architectures [16]. The most widely available industrial networks technologies can be classified into three main categories: traditional fieldbus networks, Ethernet based networks and wireless networks [17].

Initially, the industrial devices have been used serial protocols (RS-232, RS-422 and RS-485) to communicate one with another. These protocols have different characteristics, namely the number of devices that can communicate in the same time, bit rates for data transmission and the maximum cable length. The RS-232 protocol is the most common serial interface, in which only two devices are connected together at the same time. The main disadvantages of this standard are the limited cable length (20 meters) and possibility of noise occurrence. The RS-422 protocol uses balanced differential electrical signal (+5 and -5 V), as opposed to unbalanced signals referenced to ground (+5 and 0 V) with the RS-232, and permits more than one device to use the same bus. Differential data transmission increases the noise immunity and reduces the noise emissions in a harsh environment. Thus, the robust transmission over longer cable lengths (1200 meters) is possible securely and cost-effectively. The RS-422 protocol creates point-to-multipoint connections over twisted pair wire. Each device has a unique ID and master can broadcast data to several (up to 10) slaves [18, 19].

The RS-485 protocol is an improved version of the RS-422 protocol, because that can handle more devices (up to 32). The RS-485 allows more than one master device, but only one master device send data at the same time. The master devices can receive data back from each slave. The RS-485 protocol is a better fit for industrial applications and offers the multipoint communication which allows multiple devices to be connected over a single bus. It has been widely used for industrial automation applications. However, the field devices are usually connected with fieldbus from 1990 onwards [20].

The fieldbus is an isolated link that provides two-way connection between the devices in the field as shown in Figure 1. There are several fieldbus standards (PROFIBUS, INTERBUS, MODBUS, as well as CAN-based solutions such as Devicenet and CANopen) are built on the physical layer of the RS-485 protocol. Fieldbus offers several advantages, such as reduced cabling cost, simple configuration, reliability, flexibility and easy configuration when compared traditional point-to-point connections. These advantages put forward the fieldbus technology and made it more and more adoptable in industrial plants. The main drawback of the fieldbus networks is the lack of a unique standard among large number (on the order of about 100) of different and incompatible solutions of different manufacturers [20, 21].

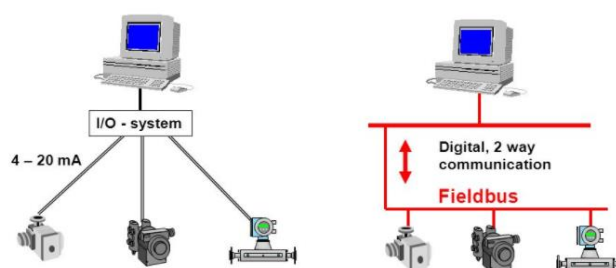


Figure 1: Traditional vs. fieldbus connections for industrial network.

At the beginning of the 2000s, the industrial Ethernet has gained popularity due to the advantages of higher speed, increased connection distance and the ability to connect more devices. Ethernet is the major wired local area network (LAN) technology that connects computers. The physical layer and data link layer properties of IEEE 802.3 standard is specified by Ethernet. Unfortunately, the Ethernet standard cannot provide the consistency requirements when timing is critical. The industrial Ethernet standard build on Ethernet in its original form. But the industrial Ethernet has a modified Media Access Control (MAC) layer to achieve very low latency and deterministic responses in a predictable timespan. There are many different industrial Ethernet protocols (EtherCAT, PROFINET, EtherNet/IP, Sercos-III) have been widely accepted in automation systems at the control level [22-24].

Today, while Fieldbus is a good candidate for the process automation, industrial Ethernet is also viewed as a potential solution for the vertical integration of automation devices, as it enabled a seamless data flow between the factory floor and upper layers. In fact, many companies are starting to use both fieldbus and industrial Ethernet solutions depending on the application. As shown in Figure 2, the sensors and actuators were connected to remote monitoring and control systems like Programmable Logic Controller (PLC), Distributed Control Systems (DCS), Supervisory Control and Data Acquisition (SCADA) using several fieldbus protocols. The PLC, DCS and SCADA systems became interconnected through

industrial Ethernet protocols. There is also a tight integration between SCADA with Manufacturing Execution System (MES) and Enterprise Resource Planning Systems (ERP) via Ethernet with TCP/IP protocols [25, 26].

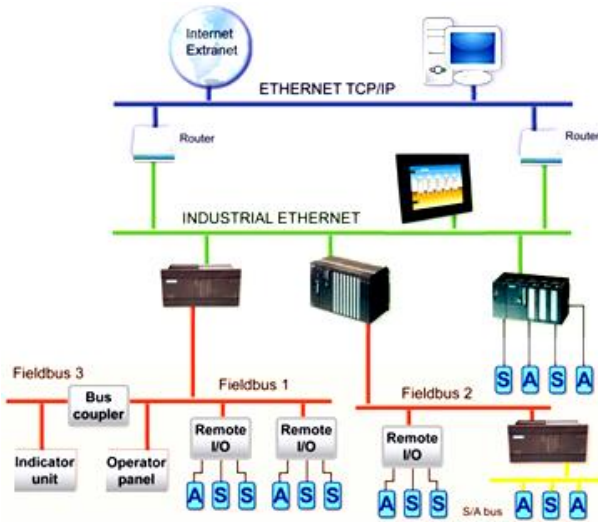


Figure 2: Fieldbus and industrial Ethernet for industrial automation.

In third-generation automation systems, the information flows increased across all levels of the plant in both vertical and horizontal lines. Therefore, the design of complex communication systems is not possible without a layered structure. In general, as shown in Figure 3, the network architecture of an industrial automation system consists of five zones in different levels of the hierarchy. This layered architecture is called as an automation pyramid. Each level has separate information characteristics such as package length, transmission speed and real-time communication requirements [27]. The automation pyramid is usually examined by partitioning into two distinct parts. The bottom levels of the pyramid cover the field devices, PLC, DCS and SCADA systems. The top levels refer to the MES, ERP systems for manufacturing management. The bottom levels of the pyramid focus on hardware functionalities. Upper levels use the data coming from lower levels to make intelligent decisions about the manufacturing operations [28].

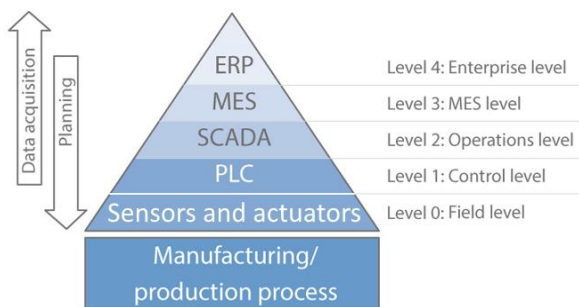


Figure 3: The automation pyramid and its functional structure.

Over the last decade, as machines and industrial control systems become increasingly more automated, ever larger amounts of information need to be processed. Industry 4.0 aims to bring together information from many distributed production facilities. The higher data volumes, nowadays called “big data”, and the integration of complex field devices demand better connectivity capabilities in industrial automation. This is especially true when taking into account the emerging smart factories, which are essentially needs vertical integration inside its boundaries. The M2M communication and I-IoT technologies can be used to break the constraints of hierarchical structure in automation pyramid by offering more flexible approach without closed barriers between adjacent layers [29].

Both M2M communication and I-IoT technologies allow networking of machines and field devices, but they slightly differ in the underlying platform to power their solutions. Figure 4 shows the network connection models for M2M communication and I-IoT applications. M2M communication aims to monitor the limited number of machines and does not necessarily require to use the Internet connection. Despite that I-IoT provides flexible network architectures via cloud computing technology, which enables the easy deployment, configuration and integration of new services into an automation system [30]. Indeed, I-IoT approach is an extended form of M2M that also includes Machine-to-Human (M2H) and Human-to-Machine (H2M) communications. Thanks to Internet Protocol version 6 (IPv6) with 128-bit address, it is possible to connect billions of things (machines and field devices) with I-IoT based applications [31].

The monitoring of industrial plants can be easily realized in a few second. Today, M2M communication applications are still used in modern industrial automation systems significantly when it comes to horizontal communication between networked machines, but I-IoT comes forward as a technological platform to make vertical communication when data needs to be transferred from field devices on the factory floor to the cloud and back again. M2M helps the realization of operational technologies (OT), while the I-IoT enables the smart factory concept for Industry 4.0 through information technologies (IT). However, the emergence of IT-OT convergence necessitates the interoperability among industrial devices from different vendors [32].

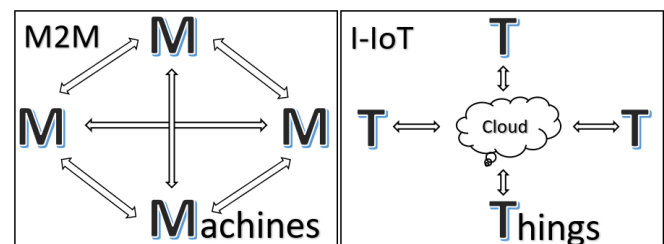


Figure 4: M2M and I-IoT based network models.

The I-IoT describes a smart manufacturing environment in which system components are continuously communicating with each other in order to support an automated production process. The integration of wireless technologies in fourth-generation automation systems is the new trend of industrial networking. Wireless networks offer reduced setup cost, faster installation and more application flexibility in industrial plants when compared with wired counterparts. Therefore, wireless networks are becoming more attractive by the day as they eliminate the problems inherent with wired networks. The traditional wireless technologies (ZigBee, Bluetooth, Wi-Fi) are standardized in the Institute of Electrical and Electronic Engineers (IEEE) 802 family and they have different range, data rate, power consumption specifications. These technologies are the basis of many home and office-based wireless network setups. But they are not designed to meet the industrial automation requirements such as low latency guarantee for real-time data transmissions and high service reliability [21, 33].

The employment of wireless networks enhances the industrial communication capabilities and also allows the manufacturers to build devices to increase overall efficiency and productivity. Today, the Wi-Fi and ZigBee technologies are still being used in non-critical applications in process control. In order to contribute the evolution of the industrial wireless networks, the well-known HART (Highway Addressable Remote Transducer) communication protocol is extended as wirelessHART. Besides, the ISA10.11a standard is redesigned, commercially referred to as ISA100 Wireless, to meet the operational needs of industrial applications. WirelessHART is a world-wide standard adopted by many organizations for process control. It supports mesh topology and channel hopping to improve the link reliability even in critical and harsh applications. The I-IoT based systems can be implemented with wirelessHART and ISA100 Wireless technologies. Figure 5 shows the I-IoT application with wirelessHART based network model [33, 34].

Low-latency guarantee is a critical requirement for real-time data transmission in wireless networks. For mission-critical applications in industrial automation scenarios, the cycle times for data transmission of time-sensitive process data are often less than 1 millisecond. IEEE 802.3 Ethernet-based communication protocols with TCP/IP are non-deterministic, and reaction time is often around 100 milliseconds. Industrial Ethernet protocols employ special preventive measures to avoid data collisions and offer lower latency and higher reliability with deterministic responses. Ether-CAT and Sercos-III protocols have response times on the order of 1 millisecond or faster [35, 36]. Recently, the Time Sensitive Networking (TSN) concept is also rising in popularity in industrial Ethernet based networks. TSN is an Ethernet extension defined by the Time-Sensitive Networking Task Group as part of IEEE 802.1 to satisfy real-time requirements in industrial communication [37, 38].

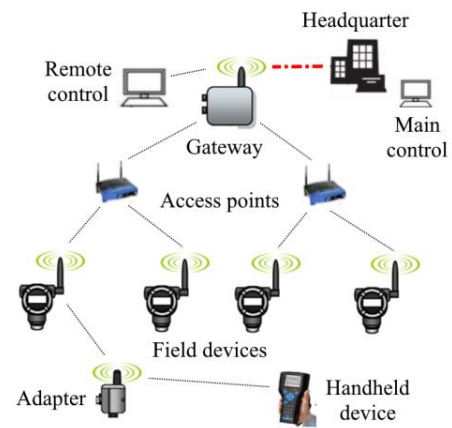


Figure 5: I-IoT application with wirelessHART based network.

III. OPC FOR INTEROPERABILITY

Automation systems operate by using continuously repeated the request and response sequences that are triggered by the controllers. The sensors gather data about the process send it to the PLC. The PLC forwards the data that are only needed from operator to the SCADA systems. The MES collects, organizes and records the process variables that can be monitored and visualized on SCADA screens. ERP system manages the process by planning future activities. To achieve efficient manufacturing process, one key issue is the realization of horizontal and vertical integration of heterogeneous industrial devices. In automation pyramid, most of the system components are communicated by using different protocols. At this point, interoperability between different communication protocols is a significant challenge to create a smart factory [39].

Since 1996, Open Platform Communications Foundation, also known as Object Linking and Embedding (OLE) for Process Control, is a de-facto organization whose standards provide an open connectivity platform over both wireless and wire-line networks for interoperability. OPC standard is a collection of specifications and describes, in plain words, a standardized communication framework for reliable data transmission between the layers of the automation pyramid. The field devices of different manufacturers and also the independent fieldbus deployments can be connected each other via OPC. Up to now, the OPC Foundation continues to set the standard specifications, with participation of industry vendors and software developers, that extending from OPC Classic to OPC UA [40, 41].

OPC Classic is a family of specifications (OPC DA: data access, OPC A&E: alarm and event data, OPC HDA: historical data access, collectively called as OPC Classic) to integrate the different automation products from multiple vendors. In OPC Classic, DA specification describes the access to current process data, A&E specification handles an interface for event-based information that specifically

includes acknowledgement about the process alarms and HDA organizes the functions to access archived data. OPC Classic specifications depend on the Component Object Model (COM) and Distributed COM (DCOM) technologies that are developed for Microsoft Windows operating systems. The COM technology enables the software components to interact with one another in a networked system. The OPC Classic architecture uses a client/server model for the information exchange and defines the COM objects with interface specification. The COM objects is used to transfer data between OPC servers and clients, as well as servers and servers. The OPC server encapsulates the access to the process data, provides an interface to the OPC objects and manages them. The objects can be on the same or on different machines. OPC client connects to the OPC server and access the data. The DCOM technology enables the client-server interaction through a network connection by allowing remote access [42].

OPC Unified Architecture (UA) standard, released in 2008, incorporates all successful and beneficial features of the OPC Classic specifications and improves them with XML and web services (HTTP, SOAP). OPC UA standard has a service-oriented architecture and provides a common protocol stack for the control needs of a connected factory. It allows operating system independent connectivity and provides communication interfaces among all component of the industrial automation system. The Microsoft's COM technology dependency and DCOM technology limitations are also removed by this way. OPC UA combines the functionality of OPC Classic specifications into single configurable address space and provides an information model for representing information data without being restricted to any specific communication protocols. The field devices from different vendors are modeled and integrated into the address space of OPC UA servers using the object-oriented techniques. Thus, the platform-independent solutions can be applied from field devices up to cloud-based ERP systems. As an example, a SCADA system establishes a connection with PLC by only using custom driver [40, 41].

The employment of I-IoT system model for automation applications is not so easy due to the used different communication frameworks. Because most of the industrial communication protocols are often incompatible with the communication protocols implemented on IoT platforms. Therefore, OPC-UA is increasingly seen as an indispensable part of I-IoT applications that require real-time interaction between devices with predictable latency [13, 43]. OPC UA standard is suitable for vertical and horizontal integration at all levels of automation pyramid from factory floor to the cloud as shown in Figure 6. In there, the gateway devices work as a bridge between the field devices and cloud-based servers. Messaging protocols (e.g. AMQP, MQTT, CoAP, JMS, etc.) are being adopted to I-IoT applications for this purpose [12, 44, 45].

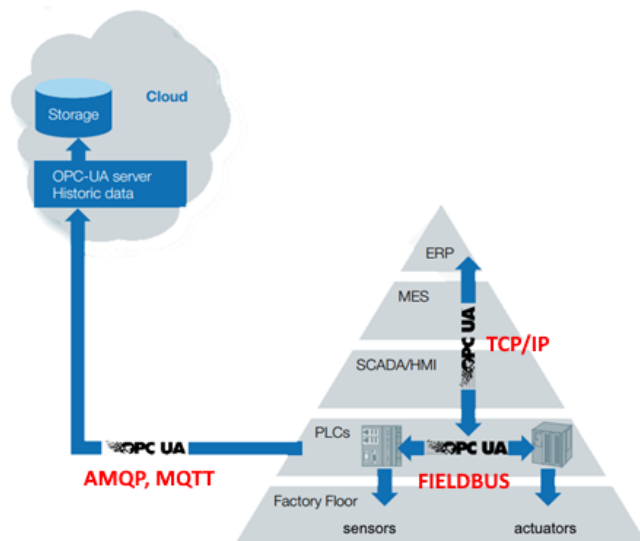


Figure 6: The interoperability in automation pyramid with OPC UA.

IV. CONCLUSION

This study has surveyed the existing communication networks in industrial automation systems. In addition, technical challenges and design objectives are described. Particularly, the fieldbus, industrial Ethernet and wireless technologies are discussed. Furthermore, the M2M and I-IoT based system implementations are presented in detail. As a consequence, the success of industrial automation systems depend on a reliable and efficient communication network that connects all the components of the factory to work together effectively. Today, many different communication protocols have been used in the industrial plants and the communication networks become more heterogeneous day by day. At this point, OPC UA represents a valuable contribution to realize horizontal and vertical integration between control levels of industrial automation systems. Recently, the wireless network technologies are evolving from consumer market to industrial applications. Today, the wireless network applications have been frequently used in industrial automation systems. On the other hand, the rise of the M2M and I-IoT technologies provides an opportunity for the remote monitoring and control of the whole industrial process. Especially, I-IoT based platforms can bring interoperability between OT and IT by communicating with most of the industrial devices. Looking to the future, there is no doubt that the industrial automation systems will continue to evolve to deliver time-sensitive communication links with better reliability. This means that the latency and packet dropout problems can be considered as two major technical challenges to significantly degrade the control performance of industrial plants, which need to be solved. Our next studies will aim to find out how the networked control systems meet today's demanding industrial challenges.

REFERENCES

- [1] Ustundag, A., and Cevikcan, E., *Industry 4.0: Managing The Digital Transformation*, Springer International Publishing Switzerland, 2018.
- [2] Zhong, R.Y., Xu, X., Klotz, E., and Newman, S.T., "Intelligent Manufacturing in the Context of Industry 4.0: A Review", *Engineering Journal*, vol. 3, no. 5, pp. 616-630, 2017.
- [3] Deloitte LLP, Connected factory: Digital solutions on the manufacturing floor, 2017. Available: <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/Consumer/IndustrialProducts/deloitte-uk-cip-connected-factory-experience.pdf>
- [4] Slama, D., Puhlmann, F., Morrish, J., and Bhatnagar, R.M., *Enterprise IoT: Strategies and Best Practices for Connected Products and Services*, O'Reilly Media, 2015.
- [5] Roblek, V., Mesko, M., and Krapež, A., "A Complex View of Industry 4.0", *SAGE Open Journal*, vol. 6, no. 2, pp. 1-11, 2016.
- [6] Zhou, H., *The Internet of Things in the Cloud: A Middleware Perspective*, CRC Press, 2013.
- [7] Balamuralidhar, P., Misra, P., and Pal, A., "Software Platforms for Internet of Things and M2M", *Journal of the Indian Institute of Science*, vol. 93, no. 3, 2013.
- [8] Verma, P.K., Verma, R., Prakash, A., Agrawal, A., Naik, K., Tripathi, R., Alsabaan, M., Khalifa, T., Abdelkader, T., and Abogharaf, A., "Machine-to-Machine (M2M) Communications: A Survey", *Journal of Network and Computer Applications*, vol. 66, pp. 83-105, 2016.
- [9] Aeris Communications, 2G, 3G, 4G... OMG! What G is Right for M2M?, 2013. Available: https://connectedworld.com/wp-content/uploads/2014/07/Whitepaper_Aeris_2G3G4GOMG.pdf
- [10] Baras, K., Brito, L., *Introduction to the Internet of Things*, Hassan, Q.F., Khan, A.R., Madani, S.A. (Eds.), *Internet of Things: Challenges, Advances, and Applications*, CRC Press, 2018.
- [11] Tsiatsis, V., Karnouskos, S., Höller, J., Boyle, D., and Mulligan, C., *Internet of Things Technologies and Applications for a New Age of Intelligence*, Academic Press, 2019.
- [12] Schoder, D., *Introduction to the Internet of Things*, Hassan, Q.F. (Ed.), *Internet of Things A to Z*, John Wiley & Sons, Inc., 2018.
- [13] Jeschke, S., *Industrial Internet of Things and Cyber Manufacturing Systems*, Jeschke, S., Brecher, C., Song, H., Rawat, D.B. (Eds.), *Industrial Internet of Things: Cybermanufacturing Systems*, Springer International Publishing Switzerland, 2017.
- [14] Wang, L., Wang, X.V., *Cloud-Based Cyber-Physical Systems in Manufacturing*, Springer International Publishing AG, 2018.
- [15] Liu, Y., Peng, Y., Wang, B., Yao, S., and Liu, Z., "Review on Cyber-physical Systems", *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 1, pp. 27-40, 2017.
- [16] Bergweiler, S., "Intelligent Manufacturing based on Self-Monitoring Cyber-Physical Systems", *UBICOMM 2015 : The Ninth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, 2015.
- [17] Jämsä-Jounela, S.-L., "Future Trends in Process Automation", *Symp. Cost Oriented Automation*, La Habana, Cuba, February 2007.
- [18] Park, J., Mackay, S., Wright, E., *Practical Data Communications for Instrumentation and Control*, Newnes, 2003.
- [19] National Instruments, RS-232, RS-422, RS-485 Serial Communication General Concepts, Available: <http://www.ni.com/white-paper/11390/en/>
- [20] B+B Smartworx, A Practical Guide to Using RS-422 and RS-485 Serial Interfaces, Available: <http://www.bb-elec.com/Learning-Center/All-White-Papers/Serial/RS-422-and-RS-485-Applications-eBook/RS422-RS485-Application-Guide-Ebook.pdf>
- [21] Sen, S.K., *Fieldbus and Networking in Process Automation*, CRC Press, 2014.
- [22] Lounsbury, R., *Industrial Ethernet on the Plant Floor - A Planning and Installation Guide*, ISA, 2008.
- [23] Zurawski, R., *Industrial Communication Technology Handbook*, CRC Press, 2015.
- [24] Aristova, N.I., "Ethernet in industrial automation: Overcoming obstacles", *Automation and Remote Control*, vol. 77, no. 5, pp. 881-894, 2016.
- [25] Softing, Products and solutions for digital communication in process automation, 2018. Available: https://industrial.softing.com/fileadmin/softing/files/pdf/en/ia/info/PA-Broschuere_EN_12Seiten_Layout_052018_online.pdf
- [26] Thomesse, J.-P., "Fieldbus Technology in Industrial Automation", *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1073 – 1101, 2005.
- [27] Calvo, I., Pérez, F., Agiriano, I.E., and Albéniz, O.G., Designing High Performance Factory Automation Applications on Top of DDS, *International Journal of Advanced Robotic Systems*, vol. 10, no. 4, 2013.
- [28] Mendes, C., Osaki, R., and Costa, C., Internet of Things in Automated Production, *European Journal of Engineering Research and Science (EJERS)*, vol. 2, no. 10, pp. 13-16, 2017.
- [29] Wollschlaeger, M., Sauter, T., and Jasperneite, J., "The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0", *IEEE Industrial Electronics Magazine*, vol. 11, no. 1, pp. 17-27, 2017.
- [30] Sethi, P., and Sarangi, S.R., "Internet of Things: Architectures, Protocols, and Applications", *Journal of Electrical and Computer Engineering*, vol. 2017, Article ID 9324035, 25 pages, 2017.
- [31] IPv6 for IoT, https://iot6.eu/ipv6_for_iot
- [32] Wang, S., Wan, J., Li, D., and Zhang, C., "Implementing Smart Factory of Industrie 4.0: An Outlook", *International Journal of Distributed Sensor Networks*, vol. 2016, Article ID 3159805, 10 pages, 2016.
- [33] Kim, D.-S., and Tran-Dang, H., *Industrial Sensors and Controls in Communication Networks: From Wired Technologies to Cloud Computing and the Internet of Things*, Springer, 2019.
- [34] Chung, T.D., Ibrahim, R., Asirvadam, V.S., Saad, N., and Hassan, S.M., *WirelessHART: Filter Design for Industrial Wireless Networked Control Systems*, CRC Press, 2018.
- [35] Hammer, A., "Ethernet Trends in the Industrial/OT Environment How machine builders and machine users are improving productivity and more", Belden (White Paper), 2018.
- [36] Lin, Z., and Pearson, S., "An inside look at industrial Ethernet communication protocols", Texas Instruments, 2018.
- [37] Danielis, P., Skodzik, J., Altmann, V., Schweissguth, E.B., Golasowski, F., Timmermann, D., and Schacht, J., "Survey on Real-Time Communication via Ethernet in Industrial Automation Environments", *IEEE Emerging Technology and Factory Automation (ETFA)*, 2014.
- [38] Varis, P., and Leyrer, T., "Time-sensitive networking for industrial automation", Texas Instruments, 2018.
- [39] Smoothing Out Interoperability Issues in Smart Factories, Available: <https://www.moxa.com/Event/integrated-solutions/smart-factory/industry-4.0/protocol-interoperability.htm>
- [40] Mahnke, W., Leitner, S.H., Damm, M., *OPC Unified Architecture*, Springer-Verlag Berlin Heidelberg, 2009.
- [41] OPC Foundation, "OPC Unified Architecture Interoperability for Industrie 4.0 and the Internet of Things", Available: <https://opcfoundation.org/wp-content/uploads/2016/05/OPC-UA-Interoperability-For-Industrie4-and-IoT-EN-v5.pdf>
- [42] Tan, V.V., and Yi, M.J., "Development of an OPC Client-Server Framework for Monitoring and Control Systems", *Journal of Information Processing Systems*, vol.7, no.2, pp. 321-340, 2011.
- [43] Cavalieri, S., Di Stefano, D., Salafia, M.G., and Scroppo, M.S., "A Web-based Platform for OPC UA integration in IIoT environment", *22nd IEEE International Conference on Emerging Technologies and Factory Automation*, 2017.
- [44] Patil, K., Lahudkar, S.L., "A Survey of MAC Layer Issues and Application layer Protocols for Machine-to-Machine Communications", *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 2, pp. 742-747, 2016.
- [45] Smart Industry Forum, "OPC UA TSN: A Small Step for Mankind, But a Giant Leap for Industry!", Available: <https://smartindustryforum.org/opc-ua-tsn-a-small-step-for-mankind-but-a-giant-leap-for-industry/>

Raspberry Pi Based Personalized Encoding ID System Inspired by the Quick Response Code Software and Hardware Design

B.Y. ÇOLDAŞ¹ and M. PALANDÖKEN¹

¹ Izmir Katip Celebi University, Izmir/Turkey, burakyasarcoldas@gmail.com

¹ Izmir Katip Celebi University, Izmir/Turkey, merih.palandoken@ikc.edu.tr

Abstract - In this paper, the encoding algorithm developed by MATLAB-based Random Number Generator has been applied to the encoding / decoding ID system inspired by the QR (Quick Response) code. However, this implementation has been conceptual manner differently from the traditional QR code implementation which are frequently used. On the other hand, in order to detect the QR code from a specific label and decode the embedded information, necessary image processing algorithms and QR code analysis algorithms have been developed using MATLAB. These algorithms have been implemented in Raspberry Pi to create an encoding / decoding ID system inspired by the QR code standard as prototype with the help of MATLAB Coder.

Keywords – Encoding, MATLAB, QR-code, Image Processing, Raspberry Pi, MATLAB Coder

I. INTRODUCTION

Barcode technology is one of the most effective parts of automatic identification technology. Decoded data can be obtained by analysis of a barcode. According to the coding type the barcode is divided into two classes: one-dimensional barcode and two-dimensional barcode [1]. QR (Quick Response) Code is two-dimensional barcode, enhanced by Denso Wave-Japan in 1994 [2]. Two-dimensional barcode such as QR code is nowadays being widely used in a variety of work. QR code is a method of encoding more information than traditional one-dimensional barcodes. And most important issue, it contains information that can be easily decoded at high speed [3]. With the development of technology, one of the basic of transferring information using some electronics devices is QR codes [4]. QR codes can be used as a method of sharing and storing information suitable for various applications, including URL sharing, mobile application, social networks, online payment, ticket services [5]. Today growing exchange of information around the world and the use of telecommunication networks such as the Internet, the availability of digital documents has become very essential by reason of it may be destroyed or attacked intentionally or unintentionally [6]. QR codes, the data that may be encoded and outline the obvious threat of a malevolent QR code. Most software developers do not consider coded

information as an unsafe entry. Since QR codes are a popular way of encoding information, there is a strong possibility that these codes can be used to launch a common attack [7].

In this paper, an algorithm has been implemented with MATLAB Random Number Generator in order to encoding the TR ID number of the person in the Inspired by the QR Code Standard. Later developed image processing algorithms and decode algorithms have been transferred to raspberry pi with the help of MATLAB Coder to detect QR code from a label and decode the embedded information.

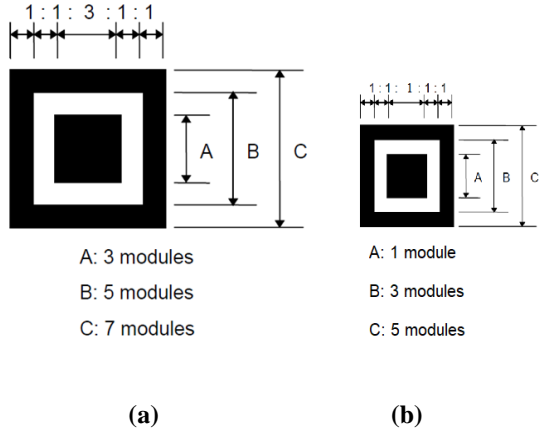
The remainder of this paper is structured with the following sub-sections: Section A presents a encoding ID system design inspired by the QR code standard. In section B, image processing and pattern recognition are being systemized with decoding ID system design inspired by QR Code standard and message extraction is explained.

II. DESIGNED SYSTEM

A. Encoding ID System Design Inspired by the QR Code Standard

In this section, firstly the implemented finder patterns and alignment pattern have been used as in the QR standards forms. Then created software with Matlab Random Number Generator is used to embed the TR ID number in the QR code.

Three finder patterns are located at exactly three position of the symbol. These are upper left, upper right and lower left of the symbol. Each finder pattern can be viewed as three interwoven squares and is constructed of a black (dark) 7*7 modules, white (light) 5*5 modules and finally a black (dark) 3*3 modules. The wideness ratio of alternative black and white module widths in each finder pattern is approximately 1:1:3:1:1 from any direction [8]. This important characteristic help us to find the locations of the finder patterns. Also Alignment patterns are similar to finding patterns. Outside the ratio of 1: 1: 1: 1: 1: 1. The alignment patterns are placed in fixed positions defined in the designed QR code. Alignment pattern can be viewed as three interwoven squares and is constructed of a black (dark) 5*5 modules, white (light) 3*3 modules and finally a black (dark) module [9].



The following figure shows the layout of these two modules in the designed QR code.

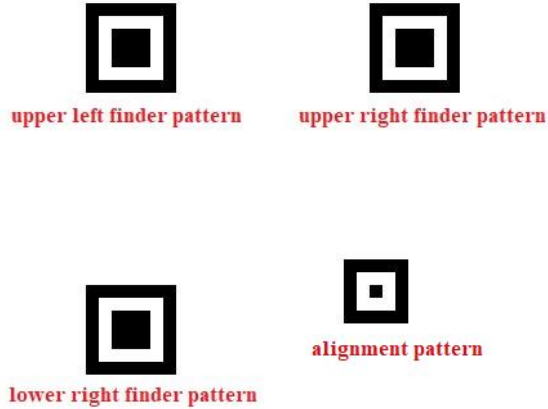


Figure 2: Finder Patterns, and Alignment Pattern

In the software which has been implemented in the form of interface to be created with Matlab Gui, the user has to enter 11-digit TR ID number. These values are placed randomly to be known only by the program. Coordinate values operate like a similar structure for the analysis process. The Matlab Random Number Generator instruction *randperm* has been used to calculate the without replacement random index permutations of the values and this operation works with different permutation values each time the program is run. For the 11-digit TR ID number, each digit is 4-bits on a binary basis, a total of 44 bit data information is randomly embedded in a specific region of the QR code. Then The 6-bit coordinate values, including the horizontal and vertical axis information of the region where the data is embedded, would be embedded in any other region of the QR code so that the reader can decode it. In the figure below, the identification number 2 0 8 3 9 7 0 6 3 8 6 is written with the software developed in the QR code, and the coordinate values of these numbers with the binary base are placed in the QR code.

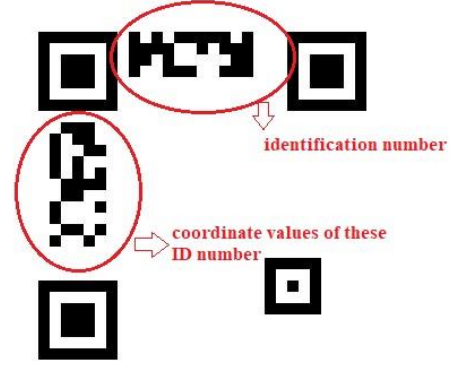


Figure 3: Embedding of TR ID and Its Coordinate Values

Then the other Matlab Random Number Generator instruction *randi* has been filled randomly with black and white modules in the QR code.

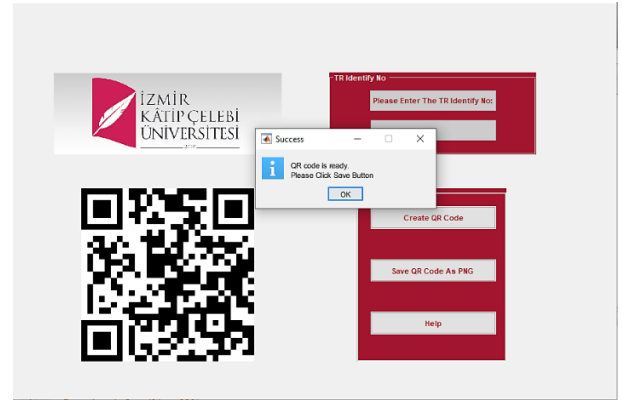


Figure 4: The Interface to be Created With Matlab Gui

The following figure has been created successfully with the generated algorithm and saved. It is also important to create a different QR code even if the same values are entered with the TR ID number in Figure 3 and Figure 4.



Figure 5: Successfully Generated QR Code

B. Image Processing and Pattern Recognition

Image processing is a technique to apply specific operations on an image, so as to get an enhanced image or to extract some useful information from image. Image processing is a kind of signal processing in which input is an image and output may be image or characteristics/features associated with that image [10]. The basic component of an image is the pixel. When mention to an image, an MxN pixel matrix should be come to mind. The figure below shows the processes performed in this section step by step.

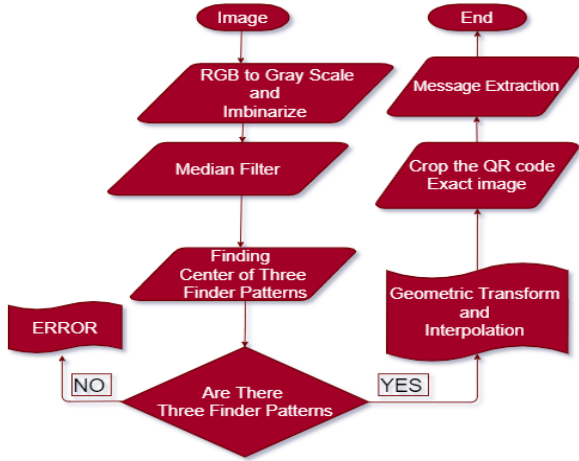


Figure 6: Image Processing and Pattern Recognition

The techniques and software applications applied in this project start with transforming the RGB (Red-Green-Blue) image into a grayscale image. The grayscale image using with some thresholding value then convert to binary image because the QR-code use only black and white module which is binary 0 and 1 respectively, the grayscale image has been converted to binary image. Some noise occurs during shooting on the original picture and also during conversions. The Median Filter used for binary image to remove noise in an image is a nonlinear digital filtering technique.

After these preparation operations, some data are needed for geometric transformation and interpolation. Geometric transformations change the spatial relationships between pixels in an image. Image interpolation occurs when we resize or distort your image from one pixel grid to another. Interpolation works by using known data to estimate values at unknown points [11]. For this purpose it must scan each row and column of the image and save how many pixels in a row and column are in the same module as black or white. As mentioned before, the intersection point of rows and columns with a ratio of 1: 1: 3: 1: 1 is a candidate to be the finder pattern. Another condition is that these candidates start with the black module and finish with the black module. After all the candidates have been found, the procedure is to choose three correct ones. K-means algorithm has been used for this process from known clustering algorithms. The candidates of finder patterns has been divided into 3 parts and exact finder patterns have been found with correct order. It is also known to the coordinates of the center points of the three finder

patterns. These coordinates have been used respectively in geometric transformation, interpolation and cropping processes.

Geometric transformation, a mapping technique that displaces the image points. Transformations can be examined under 2 main titles. Global or local in nature. Global transformations are generally defined by a single equation which is applied to the whole image. Local transformations are applied to any part of image and they are harder to express concisely. Most common global transformations are perspective, polynomial and affine transformations. The remaining transactions are progressed through the affine transformation. Having $p = (x, y)^T$ as old and $p' = (x', y')^T$ the new coordinates a pixel, a 2D affine transformation can be written as:

$$p' = A \times p + t$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (1)$$

The vector t is the translation component of the above equation. The matrix A controls scaling, rotation and shear effects:

$$A = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} : \text{scale and rotation} \quad (2)$$

$$A = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix} : \text{vertical shear} \quad (3)$$

$$A = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} : \text{horizontal shear} \quad (4)$$

$$A = \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} : \text{altered aspect ratio} \quad (5)$$

The affine transformation can be represented by a single matrix multiplication in homogeneous coordinates:

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (6)$$

Where equation 1 and 6 are related by ($a_{11} = a$, $a_{12} = b$, $a_{21} = e$, $t_x = c$, $t_y = f$) Introducing a scaling parameter W , the transformation matrix A can be modified to handle perspective corrections:

$$\begin{bmatrix} X' \\ Y' \\ W \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (7)$$

The image coordinates can be obtained by

$$x' = \frac{X'}{W} \quad (8)$$

$$y' = \frac{Y'}{W} \quad (9)$$

There are two different ways to do mapping that can be used for geometric transformations: forward mapping and backward mapping. Forward mapping is done by scanning input image, finding new coordinates and assigning the pixel value to it. Backward mapping is done by scanning output image, finding old coordinates and assigning the pixel value from it. Neither backward nor forward mapping guarantees a one to one mapping between the samples of input coordinates and output coordinates. Forward mapping produces some spaces in the target image since some of the output coordinates are never utilized. Hence backward mapping is preferred since it guarantees that the target image is completely filled.

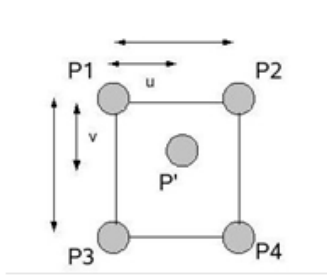


Figure 7 : Interpolation Exemplar

The problem to be faced in backward mapping is to find the value of the input image coordinates, Figure 7 The most widely used interpolation techniques are nearest neighbor interpolation and bilinear interpolation. The remaining transactions are progressed through the only bilinear interpolation. P' gets its value with the help from the plane on the 4 neighboring pixels and intensity value of P' can be found with the following bilinear equation.

$$P' = A_1 \cdot (1 - v) + A_2 \cdot v \quad (10)$$

$$A_1 = P_1 \cdot (1 - u) + P_2 \cdot u \quad (11)$$

$$A_2 = P_3 \cdot (1 - u) + P_4 \cdot u \quad (12)$$



Figure 8 : Process of Taking the First Image

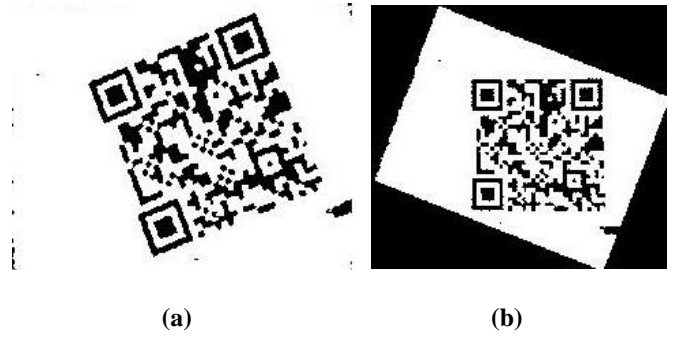


Figure 9 (a) : First Image Taken with Camera Via Raspberry Pi .

Figure 9 (b) : The Output of Geometric Transformation and Bilinear Interpolation



Figure 10 : All Designed System and Image Processing Algorithms Print Out the Applied Image.

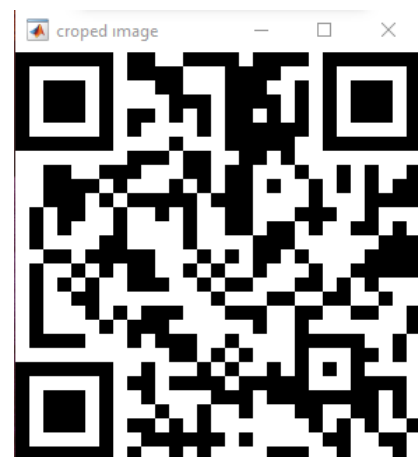


Figure 11 : Cropped the QR Code from Exact Image

The cropped QR code obtained to the algorithms applied to the image input is analyzed by an algorithm which is the opposite of the QR code generation process. The QR code reader decodes the QR code using the 6-bit coordinate values containing the horizontal and vertical axis information of the region where the data is embedded, and then finds out where the randomly placed data is stored, and the 11 digits are correctly sorted, illustrative the data that the user entered.

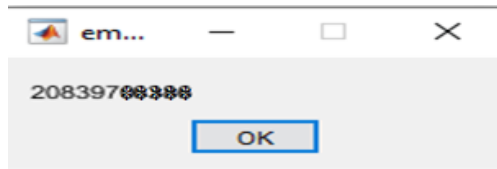


Figure 12 : Getting the Embedded TR ID Number in the QR Code by the Decoder

MATLAB Coder™ generates C and C++ code from MATLAB® code for a variation of hardware platforms, from desktop systems to embedded hardware. MATLAB Coder™ supports most of the MATLAB language and a wide range of toolboxes. It enables us to integrate the generated code as source code, static libraries or dynamic libraries. Also the generated code is readable and portable [12]. In order for the reading process to be carried out in an easy way, Raspberry Pi kit and USB camera are enough to create hardware. The algorithm is successfully embedded in the Raspberry Pi kit. This hardware provides convenience to read the QR code wherever possible.



Figure 13 : MATLAB Coder Success Built Report

III. CONCLUSION

The proposed encoding algorithm can be deduced to be inspired by the QR code in a form of a two-dimensional barcode with unique features where TR ID number can be placed with an algorithm similar to the encryption methodology. A Raspberry Pi based reader hardware has been implemented in which the image data transfer in conjunction with the image processing and decoding algorithms has been successfully accomplished. As a result of the real time hardware tests, the system has been operated at the desired level with the expected output decoded personal information. The designed system can be easily adapted for many systems and personal information to be unintentionally heard can be embedded in this designed structure and only the relevant person can be given access to this information with the personal key.

ACKNOWLEDGMENT

This work has been supported by the Scientific and Technological Research Council of Turkey (TUBITAK), Grand no: 1139B411801772

REFERENCES

- [1] Jeng-An Lin and Chiou-Shann Fuh, "2D Barcode Image Decoding," *Hindawi Publishing Corporation Mathematical Problems in Engineering* Volume 2013, Article ID 848276, 10 pages.
- [2] History of QR code. Available: <https://www.qrcode.com/en/> (Retrieved: 11/03/2019).
- [3] Phaisam Sutheebanjard and Wichian Premchaiswadi, "QR-Code Generator," *Eighth International Conference on ICT and Knowledge Engineering* 2010.
- [4] Tiago Nuno Pereira Mota, "Spectral Multiplexing of QR Codes," Thesis to obtain the Master of Science Degree in Electronics and Computer Engineering, November 2015.
- [5] Julian Brackins and Mengyu Qiao, *A Secure QR Code Scheme*. Department of Mathematics and Computer Science South Dakota School of Mines and Technology 501 E St Joseph St, Rapid City, SD 57701.
- [6] Afsaneh Arabzadeh and Alireza Naghsh, "Detection, Reconstruction, and Repairing the Distortion in Quran Pages Based on Watermarking" *Majlesi Journal of Electrical Engineering* Vol. 12, No. 3, September 2018.
- [7] Vishrut Sharma "A Study of Malicious QR Codes" *International Journal of Computational Intelligence and Information Security*, May 2012 Vol.3, No.5 ISSN: 1837-7823.
- [8] QR Code Tutorial. Available: <https://www.thonky.com/qr-code-tutorial/module-placement-matrix> (Retrieved: 14/03/2019).
- [9] QR Code Tutorial. Available: <https://www.thonky.com/qr-code-tutorial/module-placement-matrix> (Retrieved: 14/03/2019).
- [10] Gonzalez, R.C. and Woods, R.E. (2008). *Digital Image Processing*. Pearson 3rd edition.
- [11] Şükrü Ozan, "A Case Study on Logging Visual Activities Chess Game" Izmir Institute of Technology in Partial Fulfillment of the Requirements for the Degree of Master of Science in Electrical and Electronics Engineering November 2005 İzmir.
- [12] MATLAB Coder Generate C and C++ code from MATLAB code. Available: <https://www.mathworks.com/products/matlab-coder.html> (Retrieved: 17/03/2019).

The usage statistics of new HTML5 semantic elements in the ClueWeb12 dataset

G. ÇIPLAK¹, A. AYDIN² and A. ARSLAN³

¹ Eskişehir Technical University, Eskişehir/Turkey, gciplak@eskisehir.edu.tr

² Eskişehir Technical University, Eskişehir/Turkey, ahmetaydin@eskisehir.edu.tr

³ Eskişehir Technical University, Eskişehir/Turkey, aarslan2@eskisehir.edu.tr

Abstract - The ClueWeb12 dataset is the newest and the largest TREC test collection used in information retrieval research. The dataset consists of 733,019,372 English Web pages, collected between February and May 2012. The HTML5 Living Standard was published in the same year and it has been continuously updated since then. The most interesting new HTML5 elements are the semantic elements, which made easier to identify the correct Web page content for commercial search engines such as Google, Yandex, and Bing. This paper reports the usage statistics of new HTML5 semantic elements in the ClueWeb12 dataset. Although HTML5 was not widely adopted in 2012, we have found that 4% of the Web pages in the ClueWeb12 dataset have at least one of the new HTML5 semantic elements. We report also the most frequently used abbreviations and acronyms along with their expanded forms. We propose a couple of interesting ways to leverage new semantic elements in Web information retrieval.

Keywords - HTML5, Semantic elements, ClueWeb12, Web retrieval.

I. INTRODUCTION

THE first HTML5 public draft was released in 2008. In 2012, HTML5 Living Standard was published by the Web Hypertext Application Technology Working Group¹. HTML5 introduced new semantic elements with the motivation of allowing commercial search engines to easily identify the correct Web page content. This paper reports the usage statistics of HTML5 semantic elements in the ClueWeb12 dataset [1], which is crawled between February and May 2012. Although HTML5 was not widely adopted in 2012, we have found that at least one of the HTML5 semantic elements are used in 4 percent of Web pages of the ClueWeb12 dataset. The most frequently used semantic elements are identified as footer, header, nav, time, and article.

The remainder of this paper is organized as follows. We introduce HTML5's new semantic elements in Section 2. We discuss idiosyncratic challenges of Web retrieval in Section 3, which also includes a brief summary of existing related work. We present the usage statistics of semantic elements in the ClueWeb12 dataset in Section 4. We propose a couple of interesting ways to leverage semantics elements in Section 5. Finally, the conclusion and future work are given in Section 6.

¹ <https://whatwg.org>

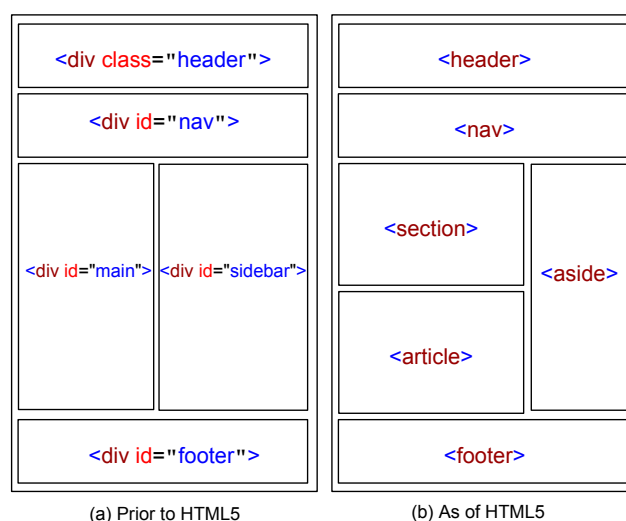


Figure 1: HTML5 offers new semantic elements to define different parts of a Web page.

II. HTML5 SEMANTIC ELEMENTS

With HTML4, developers used their own id/class names to style elements: header, top, bottom, footer, menu, main, container, content, article, sidebar, navigation, etc. This old-fashioned coding style is shown in Figure 1 (a). Such convention gives no standard clue about the actual content of a Web page. To see why, consider that a vertical search engine for news retrieval has been built using a crawler. Assume that for a news website, text of the article resides inside of `<div id="articleBody" itemprop="articleBody"> ... </div>` elements. On the other hand, another news website might publish its article inside of `<div class="newsbody"> ... </div>` elements. The developers of such crawler should both identify and update HTML markers for each news website separately. Indeed, such mechanism is labor and time intensive and may be possible for a vertical search engine that crawls a limited number of news websites. However, this approach is infeasible for generic search engines (Google, Yandex, Bing) to identify the correct Web page content.

As illustrated in Figure 1 (b), the HTML5 semantic elements can be used to specify the main content of a HTML page by the webmaster. In the figure, `<section>` or `<article>` elements include main content. On the other hand, `<nav>`, `<header>`, `<aside>` or `<footer>` elements contain template/boilerplate content (i.e. common for each page of a website and may not be very useful for full-text search).

A simple content extraction strategy (based on the semantic elements) would simply return what is inside of the `<section>` or `<article>` elements. Or alternatively, it would just remove `<nav>`, `<header>`, `<aside>` or `<footer>` elements from the whole HTML and return what is left. As demonstrated by this simple content extraction strategy, creating a Web page using HTML5 semantic elements produces more formatted page layout making content extraction applications' task easier.

The HTML5 semantic elements will contribute not only to search engines but also to the semantic Web that allows data to be shared and reused across applications, enterprises, and communities. Table 1 lists new semantic elements² and their descriptions in HTML5.

Table 1: HTML5 semantic elements and their descriptions.

<code><article></code>	Defines an article
<code><aside></code>	Defines content aside from the page content
<code><details></code>	Defines additional details that the user can view or hide
<code><figcaption></code>	Defines a caption for a <code><figure></code> element
<code><figure></code>	Specifies self-contained content, like illustrations, diagrams, photos, code listings, etc.
<code><footer></code>	Defines a footer for a document or section
<code><header></code>	Specifies a header for a document or section
<code><main></code>	Specifies the main content of a document
<code><mark></code>	Defines marked/highlighted text
<code><nav></code>	Defines navigation links
<code><section></code>	Defines a section in a document
<code><summary></code>	Defines a visible heading for a <code><details></code> element
<code><time></code>	Defines a date/time

III. WEB INFORMATION RETRIEVAL

Early Information Retrieval (IR) datasets were constructed from newspaper articles (e.g. Financial Times) or abstracts of academic journal articles (e.g. Communications of the ACM). These documents are plain text files with a very limited markup. The newest datasets such as ClueWeb09, and ClueWeb12 are crawled from the Web, therefore the documents are actually Web pages. Different from news articles and research abstracts, Web pages are written in HTML, which stands for Hyper Text Markup Language.

This idiosyncratic property of Web retrieval brings another challenge to the IR research: Identify the useful textual content of a given HTML page. It is worthwhile to note that without this step, searching the query word *'html'* would return all the documents in the collection since all Web pages would contain `<html> . . . </html>` at the bare minimum. More recently, Roy *et al.* [2] experimented with an approach of indexing Web

pages as-is (including the HTML tags) under the assumption that HTML tags are unlikely to occur in a user query. The authors report that (i) HTML tags should be indeed cleaned prior to indexing, (ii) cleaning generally improves the retrieval effectiveness for the most of the retrieval models [2].

The simplest solution to this cleaning problem would be just stripping HTML tags, which is the prevalent method among IR researches. There are many different open source solutions for the task of stripping HTML tags. However, even this simple method can be very tricky to implement that leads these solutions to produce different results. Some HTML mistakes (e.g. unclosed elements) or the use of JavaScript in HTML code could be the reason for obtaining different results.

Just stripping HTML tags is far away from optimal since this approach would keep boilerplate content such as advertisements. A sophisticated approach³ was proposed in 2010, which uses machine learning techniques to detect main content of a Web page [3]. Since content extraction is an important preprocessing step in Web data mining, a number of approaches have been proposed in the literature [3-10].

Song *et al.* [4] have proposed a hybrid method that combines textual and visual information with respect to text density and main content location for content extraction of a Web page. Insa *et al.* [5] have used words/leafs ratio in the DOM tree for discovering hierarchical relations between the Web page nodes.

Uzun *et al.* [6] have proposed a method that learns a pattern from DOM-based features to determine decision tree rules which are used to extract information from content blocks. Peters and Lecocq [7] integrated the semantics of several specific HTML tags for content extraction. Also, other similar works [8, 9] have implemented a machine learning method with complementary techniques such as grouping and filtering. Banners, advertisements, navigations, etc. could be considered as noise for content extraction applications. Detecting and removing these noises are also a method to obtain the main content of a Web page. Since several Web content extraction algorithms have been published in the literature, the researchers have carried out a meta-analysis [10].

We would like to emphasize that, none of the existing content extraction studies mention or utilize HTML5 semantic elements except [7,10]. In the next section, we present some statistical information about Web pages of the ClueWeb12 dataset.

IV. EXPERIMENTAL RESULTS

The ClueWeb12 dataset includes about 733 million Web pages crawled in 2012. This paper is the very first study that examines the ClueWeb12 dataset in terms of HTML versions of Web pages and the usage statistics of the HTML5's new semantic elements. Table 2 reports HTML versions of the Web pages in the ClueWeb12-B13 dataset, which is a subset of the full dataset. As can be observed from Table 2, the most frequent HTML version is XHTML, which covers 42.2% of the dataset. XHTML is followed by the HTML5 with 22.6% coverage. While detecting HTML versions, 21.44% of pages could not be determined because of the misuse of document version identifier.

² https://www.w3schools.com/html/html5_semantic_elements.asp

³ <https://boilerpipe-web.appspot.com>

Table 2: HTML version statistics for the ClueWeb12-B13 dataset.

HTML Version	Frequency	Percentage
XHTML	22,091,113	42.2
HTML5	11,299,308	22.6
HTML4	7,122,570	13.6
HTML3	442,112	<1
Not Detected	11,219,584	21.44
Others	148,736	<1
Total Documents	52,323,423	

Table 3 shows the usage percentage (i.e. document frequency divided by the total number of documents) of the 13 new semantic elements introduced in HTML5. The numbers reported in the table are obtained using the full dataset.

Table 3: The usage percentage of the 13 new semantic elements.

Element	Percentage	Element	Percentage
footer	2.682	figure	0.381
header	2.484	summary	0.273
nav	2.215	details	0.270
time	1.930	figcaption	0.105
article	1.624	mark	0.003
section	1.553	main	0.001
aside	1.520		

A semantic element is the one that clearly describes its content to both the browser and the webmaster. The semantic elements listed in Table 3 are specific for defining different parts of a Web page. However, there exists other semantic elements intended for different purposes.

For example, the `<acronym>` and `<abbr>` elements are used to embed expanded form of an acronym or abbreviation in an HTML page. The `<acronym>` element is supported until HTML4 but it has been replaced by `<abbr>` since HTML5. The expanded form is not rendered by default. When the user moves the mouse over the acronym, the expanded form is displayed as a tool tip text. This semi-hidden information can be utilized by search engines: a set of acronyms and abbreviations extracted from Web pages can be used to construct a database of acronyms and their expanded forms. The Google already supports such functionality. Figure 2 shows the rich snippet returned by the Google in response to the query “doi acronym”. Some popular expanded forms of DOI are listed in the response.

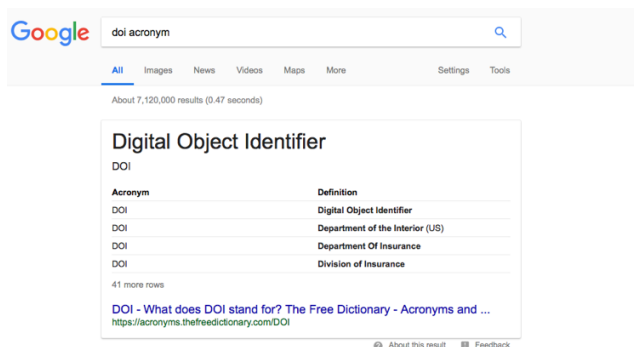


Figure 2: Example of an acronym search supported by Google.

The `<dfn>` element indicates a definition of a term: “`<dfn>FOMO</dfn>` is a pervasive apprehension that others might be having rewarding experiences from which one is absent”.

The `<cite>` tag defines the title of a work: “`<cite> The Rise of the Network Society </cite>` by Manuel Castells. Published in 1996”. These tags clearly define their content to both the users and search engines.

The usage statistics (document frequencies) of four semantic elements in the ClueWeb12-B13 dataset are as follows: `<abbr>` (4.33%), `<cite>` (1.92%), `<acronym>` (0.44%), and `<dfn>` (0.12%). Tables 6 and 7 list the most frequently used acronyms/abbreviations and their expanded forms extracted from the ClueWeb12-B13 dataset.

Table 4: The most frequent usage of the `<acronym>` tag.

Expanded form	Acronym	Frequency
us dollar	usd	33,177
uniform resource identifier	uri	25,434
really simple syndication	rss	18,175
copyright	©	16,532
Hypertext markup language	html	12,767
portable document format	pdf	10,874
perform a search by keyword, date, and/or name...	search	7,946
frequently asked questions	faq	5,019
natural resources conservation service	nrcs	4,850
cascading style sheets	css	3,596

Table 5: The most frequent usage of the `<abbr>` tag.

Expanded form	Abbreviation	Frequency
really simple syndication	rss	1,523,032
Hypertext markup language	html	287,699
points	pts	129,150
uniform resource identifier	uri	57,071
average	avg.	33,806
cascading style sheets	css	15,501
dollar	\$	12,584
universal resource locator	url	12,530
frequently asked questions	faq	5,577
digital object identifier	doi	3,581

The proper usage of semantic elements during the development of a Web page makes its content more interpretable by machines or robots. For instance, “DOI” attached with its expanded form using the `<abbr>` element would provide more information to robots. Using semantic elements to markup content could be helpful for search engines to identify relevant parts of Web pages.

In this perspective, `<dfn>` tag introduced before HTML5 can also be considered as a semantic tag because it includes a definition of a term or concept. Unlike the `<abbr>` tag, definition is rendered by the browser. This is an important difference because if an information is not rendered by default, human users cannot see it in the browser. Thus, this information is mostly intended for Web crawlers or robots.

V. OTHER INTERESTING APPLICATIONS

In this section we propose two interesting ways of leveraging the semantic elements to enhance Web IR: (i) definition lookup and (ii) text-based image retrieval. These two functionalities are already supported by the Google search engine.

A database of keywords and their definitions extracted from Web pages (the `<dfn>` tag) can be created. Specific queries looking for a definition of a term can be served using this database. For example, when a query such as “*paralegal definition*” is issued to the Google search engine, as can be seen in Figure 3, a rich snippet that shows two definitions taken from different sources is displayed to the user.

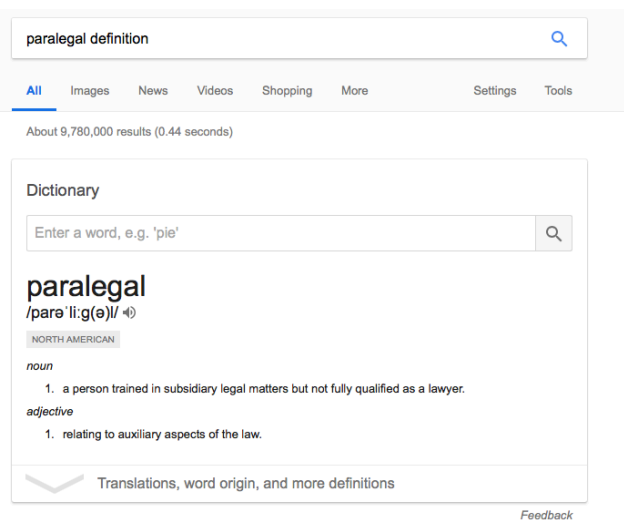


Figure 3: Example of a definition lookup supported by Google.

Similarly, `<figcaption>` and `<figure>` tags can be used for text-based image retrieval. A crawler can be written to extract figure captions and the image URLs from Web pages. A separate index created from figure captions would allow search queries against the `<figcaption>` field and matching images could be displayed to the user. Figure 4 shows Google’s image search results in response to the query “*black fox*”. The response shows several images of black colored foxes.

Images for black fox

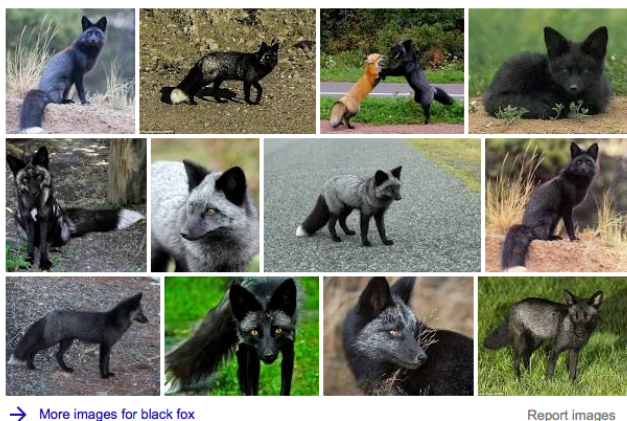


Figure 4: Example of Google’s image search.

VI. CONCLUSIONS

Content extraction is an important task for Web IR because removing navigational elements, templates and advertisements from Web pages plays significant role while retrieving relevant documents for a given query from large amounts of Web pages. HTML5 introduces new semantic elements, which can be used to make Web pages more understandable for search engines.

This paper presents the usage statistics of the HTML5’s new semantic elements in the ClueWeb12 dataset. We also demonstrate three practical applications of the semantic elements applied to Web IR: acronym finder, text-based image retrieval, and definition lookup. As we demonstrate in this paper, these applications are already supported by commercial search engines.

The HTML5 semantic elements will contribute not only to search engines but also to the semantic Web that allows data to be shared and reused across applications, enterprises, and communities.

We actually implemented a hybrid content extraction method that takes into account the HTML5 semantic elements and empirically evaluated our method using the ClueWeb12 dataset and its associated query sets. However, our approach did not improve the retrieval effectiveness when compared to a simple HTML tag stripping strategy. This is probably due to the fact that only 4% of the ClueWeb12 dataset contains the semantic elements of HTML5. When a newer benchmark dataset for Web retrieval is released, we will test our method on it.

REFERENCES

- [1] Callan, J. (2012). “The Lemur project and its ClueWeb12 dataset,” *Invited talk at the SIGIR 2012 Workshop on Open Source Information Retrieval*. Available: <http://opensearchlab.otago.ac.nz/SIGIR12-OSIR-callan.pdf>
- [2] Roy, D., Mitra, M., & Ganguly, D. (2018). “To clean or not to clean: Document preprocessing and reproducibility,” *Journal of Data and Information Quality (JDIQ)*, 10(4), 18:1-18:25.
- [3] Kohlschütter, C., Fankhauser, P., & Nejdl, W. (2010). “Boilerplate detection using shallow text features,” in *Proceedings of the third ACM international conference on Web search and data mining* (pp. 441-450). [Online]. Available: <http://doi.acm.org/10.1145/1718487.1718542>
- [4] Song, D., Sun, F., & Liao, L. (2015). “A hybrid approach for content extraction with text density and visual importance of DOM nodes,” *Knowledge and Information Systems*, 42(1), 75-96.
- [5] Insa Cabrera, D., Silva Galiana, J. F., & Tamarit, S. (2013). “Using the words/leafs ratio in the DOM tree for content extraction,” *Journal of Logic and Algebraic Programming*, 82(8), 311-325.
- [6] Uzun, E., Agun, H. V., & Yerlikaya, T. (2013). “A hybrid approach for extracting informative content from web pages,” *Information Processing & Management*, 49(4), 928-944.
- [7] Peters, M. E., & Lecocq, D. (2013). “Content extraction using diverse feature sets,” in *Proceedings of the 22nd International Conference on World Wide Web* (pp. 89-90).
- [8] Wu, S., Liu, J., & Fan, J. (2015). “Automatic web content extraction by combination of learning and grouping,” in *Proceedings of the 24th international conference on World Wide Web* (pp. 1264-1274). International World Wide Web Conferences Steering Committee.
- [9] Oza, A. K., & Mishra, S. (2013). “Elimination of noisy information from web pages,” *International Journal of Recent Technology and Engineering*, 2(1), 115-117.
- [10] Weninger, T., Palacios, R., Crescenzi, V., Gottron, T., & Meriardo, P. (2016). “Web content extraction: a meta analysis of its past and thoughts on its future,” *ACM SIGKDD Explorations Newsletter*, 17(2), 17-23.

Internet of Things based Remote Monitoring System Design for Industrial Plants

H. SOY¹ and S. KOÇER¹

¹ Necmettin Erbakan University, Konya/Turkey, hakkisoy@erbakan.edu.tr

¹ Necmettin Erbakan University, Konya/Turkey, skocer@erbakan.edu.tr

Abstract – In this study, the design of an Internet of Things (IoT) based industrial remote monitoring system is discussed. The proposed system model consists of a gateway and several sensor nodes that can detect the physical quantities in the plant floor. A gateway wirelessly collects the data from sensor nodes, aggregates the collected data and transmit it to the cloud server via the internet connection. The hardware prototypes of sensor nodes and gateway are designed by using STM32F103C8T6 microcontroller, nRF24L01 wireless communication module, ESP8266 Wi-Fi module and BME280 environmental sensor. The proposed system helps to remotely monitor the changes in temperature, humidity and pressure values. Beyond that, the critical situations can be avoided and preventive measures are successfully implemented.

Keywords - Remote monitoring, IoT, STM32F103C8T6 nRF24L01, ESP8266, BME280.

I. INTRODUCTION

IN recent years, the use of wireless communication technologies, especially in home and office environments, have attracted the attention of researchers with the idea to also use them in industrial plants and manufacturing systems [1]. Industrial environments often have very difficult operating conditions due to the chemicals, vibrations and moving parts. Employing wireless communication offers significant advantages as it does not require cable infrastructure and thus provides significant flexibility while reducing installation and maintenance costs in remote monitoring and control [2]. As a result of ongoing researches, there are many applications have been introduced for the use of wireless connectivity in industrial automation systems. In these applications, wireless network technologies, such as Wi-Fi and ZigBee, are often used to establish an infrastructure that supports required services [3-5].

Industrial control systems can be realized either as open or closed control loops by using sensor and actuator pairs. All of the system components have a connection with each other for real-time operation. In a typical industrial process, the most important problem related to communication is to ensure a reliable, predictable and secure data transfer between field devices (sensors, actuators and controllers). In conventional system setup, the field devices are usually connected to each

other by an isolated connection, called as FIELDBUS [6]. In this context, various standards have been introduced by the different companies, such as PROFIBUS, INTERBUS, MODBUS, CANBUS [7]. But, the digital transformation with Industry 4.0 involves more intelligent and interconnected factory operations to quickly respond the customer demands by industrial plants [8]. As machines and industrial control systems become more automated, it has become increasingly difficult to process a large amount of data collected from many points on industrial plants and also transfer them to the required control units. These high data volumes, today called as “big data”, make it necessary to achieve fast and reliable communication systems in industrial applications [9, 10].

The use of Internet of Things (IoT) technology, which has become increasingly widespread in many areas in recent years, can be considered to meet the increased needs in industrial applications [11]. The IoT concept refers to the global scale telecommunication infrastructure which is used to connect objects (machines, buildings and even devices used in daily life) to the internet for remote monitoring and control purposes. The objects with this feature include a self-contained embedded control system to communicate through existing internet infrastructure via wired or wireless connections [12, 13]. Initially, IoT technology was used in commercial products and it has been become an integral part of industrial systems in many countries with the widespread use of the internet [14].

The Industrial Internet of Things (IIoT) paradigm offers an opportunity to analyze the data at every stage of modern manufacturing system and it gives possibility to transform the industrial processes into the Cyber-Physical System (CPS) [15, 16]. IIoT can detect the failures, react to the unexpected changes and trigger maintenance. The digital transformation in industrial plants starts with IIoT based applications. Besides, a wireless sensor network (WSN) has a key role to play in providing of the IoT vision. WSNs collect data about the environmental parameters and provides a distributed sensing function with more accurate data. In fact, a WSN can be a well-suited tool for data acquisition in IIoT application. Recently, there are lots of WSN integrated remote monitoring system are adopted to industrial plants. A WSN can be completely independent from the Internet, be able to exchange information with Internet hosts (gateway), or share a compatible network layer protocol (TCP/IP) [17-19].

In this study, we have designed an IoT based remote monitoring system that can be used for industrial plants. The hardware and firmware design of system components were prepared by using low cost and locally available components. The communication protocol that will provide functionality to the system components is investigated in this context.

II. SYSTEM MODEL

The proposed system model is built on a WSN that allows remote monitoring of the industrial plant and the system framework contains three parts: sensor nodes, gateway and server as shown in Figure 1. The sensor nodes are deployed in different places to remotely monitor the factory floor. It is assumed that each sensor node contains a battery pack and the battery recharge operation is impractical on the factory floor. The number of sensor nodes depends on the physical size of the process. Although it is practically possible to transfer physical parameters measured by each sensor node to a remote monitoring center via a direct internet connection, such an application would make the system model considerably more complex. Besides, the direct internet connection to server with machine type communication also increases the cost of the SNs. In many applications, the sensor nodes should be low cost and energy efficient since their limited energy resource determines their lifetime.

In proposed system, the sensor nodes sense the monitored process with their sensors. After the data capture operation, the collected sensor data is delivered from the sensor nodes to a gateway with single-hop transmission. A gateway that acts as a common sink and collects data from the sensor nodes. Then, it eliminates the redundancy and prevents unnecessary data transfer by applying data aggregation techniques on the sensor data. This approach guarantee that only meaningful data will be forwarded to cloud server, instead of wasting the valuable network bandwidth. The gathered data are unloaded to server of remote monitoring center by a gateway via wired or wireless internet connection. The cloud server receives and stores the collected data, performs data analytics and creates visual illustrations for easier data interpretation.

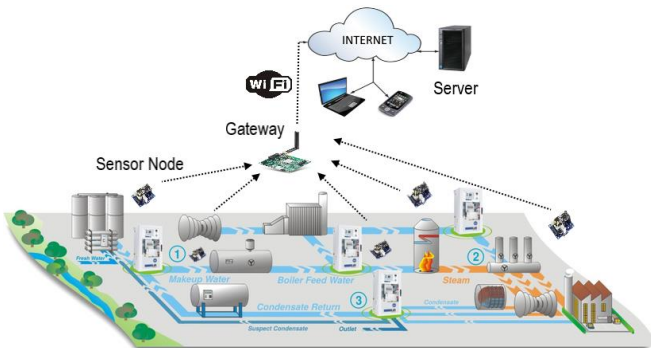


Figure 1: Proposed system model for remote monitoring.

III. HARDWARE DESIGN

The hardware design is an important part of the proposed system. The performance and stability of remote monitoring operations are directly depending on well-designed hardware components. In our study, we use STM32F103C8T6 as the microcontroller chip with 72MHz frequency, 64KB flash program memory, 20KB data memory, 3 general timers, 1 high-grade timer, 2 SPI communication interfaces, 2 I²C communication interfaces, 2 modules of 10 channel 12 bit synchronous ADC, 1 module of 8 channel pulse width modulation and 37 digital I/O ports [20]. To simplify the hardware design, we use STM32F103C8T6 development boards for both sensor node and gateway prototypes. These development boards are shown in Figure 2, respectively.

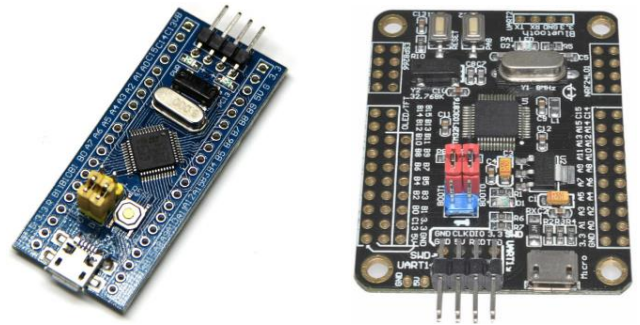


Figure 2: Development boards for hardware prototypes.

Sensors play an important role in all kind of industrial plant. In this study, we assume that the production process is kept under temperature control by observing humidity and pressure measurements. According that each sensor node is equipped with a Bosch BME280 environmental sensor to measure the temperature, humidity and pressure values in the surroundings of the process [21]. The sensor nodes communicate with a gateway through the nRF24L01 transceiver modules operating on the 2.4 GHz ISM band [22]. The MISO and MOSI pins nRF24L01 module are connected to PA6 and PA7 pins of STM32F103C8T6 via SPI interface. Figure 3 shows the prototype design of sensor node.

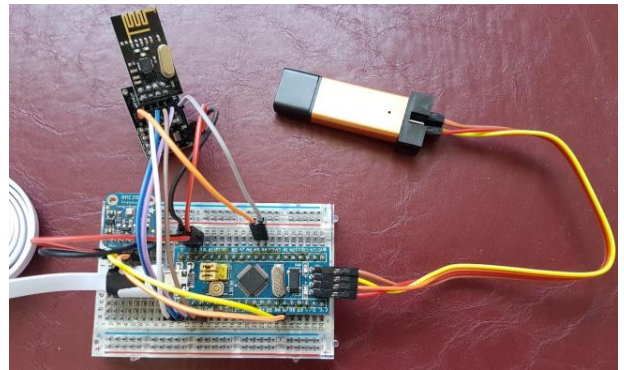


Figure 3: Sensor node hardware prototype on development board.

A gateway bridges the communication gap between the sensor nodes and cloud server. We assume that the Wi-Fi connection is available around the industrial plant. A gateway uploads aggregated data to server via Espressif ESP8266 Wi-Fi module [23]. The receive (Rx) and transmit (Tx) pins of the ESP8266 module are connected to PA2 and PA3 pins of STM32F103C8T6 respectively. Figure 4 shows the prototype design of gateway.

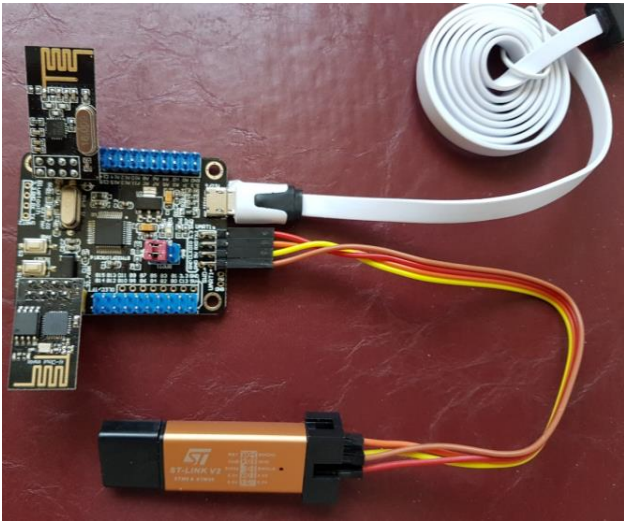


Figure 4: Gateway hardware prototype on development board.

IV. COMMUNICATION PROTOCOL

The remote monitoring of an industrial plant depends on the reliable communication protocol that enables real-time data transmission between the system components. In WSN applications, the most important criterion in the design of the communication protocol for battery operated sensor nodes is to extend their working life as far as possible. Therefore, the communication of sensor nodes is usually organized by using TDMA (Time Division Multiple Access) based channel access approach. In TDMA the available bandwidth is divided into a number of time-slots. Each time-slot can then be assigned to a certain sensor node. The TDMA based protocols needs a highly accurate and precise clock to work reliably, especially across back-to-back links. But time synchronization among randomly located sensor nodes is very difficult to achieve in harsh industrial environments.

In our study, we prefer to use a contention-based communication without common time schedule for sensor nodes. By this way, we aim to improve the energy efficiency by reducing control signal overhead for time synchronization. Our communication protocol is based on well-known CSMA/CA mechanism. In proposed communication protocol, the sensor nodes listen the channel before sending their data packets to detect other ongoing transmissions. If the channel is free, the sensor nodes send their packets to gateway. Otherwise, they delay (or back off) transmissions for a random

amount of time and sense the medium again. Sensor nodes employ a random exponential back-off algorithm that backs-off for a random interval of 250-1000 milliseconds before retransmission. After successful packet transmission, the sensor nodes switches to the sleep mode and minimizes their power consumption.

The payload and packet format are the other important factors. The nRF24L01 transceivers use an Enhanced ShockBurst packet structure, which is broken down into 5 different fields. The Enhanced ShockBurst packets consist of Preamble, Address, Payload, Cyclic Redundancy Check (CRC) and newly introduced Packet Control Field (PCF) fields for more robust communications. The PCF allows for variable length payloads with a payload length specifier (6 bit) and the payload field can vary from 1 to 32 bytes. The PCF also provides each sent packet with a packet ID (2 bit), which allows the receiving device to determine whether a message is new or whether it has been retransmitted. Besides, each packet transmission can request an acknowledgement (1 bit) to be sent when it is received by another device. The packet checksum can be optionally enabled 0 to 2 bytes CRC field. The nRF24L01 stores every preloaded packet into transmit First-In-First-Out (FIFO) buffer and send it on the listening pipe address. When packet arrives, the receiver recognizes the preamble, matches the address and performs the CRC check. If all checks are successful, the receiver extracts the packet payload and transfer it to the microcontroller [24, 25].

In our packet setup, we use 3-byte address and configure the transceiver for 6 bytes of payload, which is 2 bytes for each sensor measurement without CRC. Each sensor node has unique address and transmitted packets contain a source address, destination address, packet length and sensor data. The Enhanced ShockBurst packet structure and our payload field detail are shown in Figure 5. In our experimental study, the necessary firmware was prepared in the Keil MDK-ARM v5.24 development environment and it loaded into the microcontroller with the help of the ST-Link V2 USB programmer. We have revised and used the nRF24L01 module HAL library (available on GitHub) [26] and ESP8266 module HAL library (available on GitHub) [27] for STM32F103C8T6 microcontroller operation.

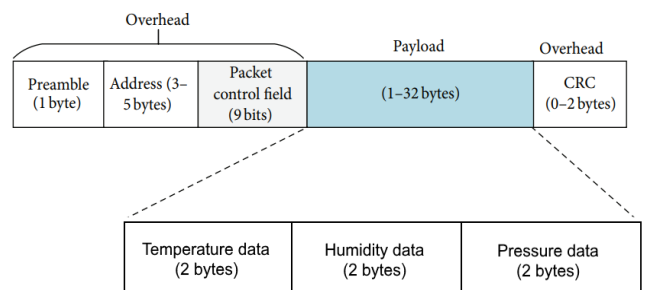


Figure 5: Enhanced ShockBurst packet structure with payload.

V. CONCLUSION

This study has aimed to construct an application model of the IoT based remote monitoring system for industrial plants. The remote monitoring system was implemented using low-cost STM32F103C8T6 boards, nRF24L01 transceivers and ESP8266 Wi-Fi module to experimentally validate the proposed model. A WSN were established to provide a distributed sensing function on the factory floor. All of the sensor nodes were connected to a gateway through nRF24L01 modules. A gateway was linked to the cloud server via ESP8266 Wi-Fi module. In this way, the collected data from the factory floor was sent to the cloud platform. We hope that the insights gained in this study will inspire further discussions and researches to make it easy industrial IoT applications. In the future, we would like to add the missing features of proposed system.

REFERENCES

- [1] Egea-Lopez, E., Martinez-Sala, A., Vales-Alonso, J., Garcia-Haro, J., and Malgosa-Sanahuja, J., "Wireless communications deployment in industry: a review of issues, options and technologies", *Computers in Industry*, vol. 56, no. 1, pp. 29-53, 2005.
- [2] Ikram W., and Thornhill, N.F., "Wireless communication in process automation: A survey of opportunities, requirements, concerns and challenges", *UKACC International Conference on Control*, pp. 1-6, Coventry, U.K., 2010.
- [3] Moyne J.R., and Tilbury, D.M., "The Emergence of Industrial Control Networks for Manufacturing Control, Diagnostics, and Safety Data", *Proceedings of the IEEE*, vol. 95, no. 1, pp. 29-47, 2007.
- [4] Raut, A.R., and Malik, L.G., "ZigBee in Building Industrial Control and Automation", *International Journal of Wireless Communications and Networking*, vol. 3, no. 2, pp. 149-154, 2011.
- [5] Paavola, M., "Wireless Technologies in Process Automation - Review and an Application Example", University of Oulu, 2007.
- [6] B+B Smartworx, *A Practical Guide to Using RS-422 and RS-485 Serial Interfaces*, Available: <http://www.bb-elec.com/Learning-Center/All-WhitePapers/Serial/RS-422-and-RS-485-Applications-eBook/RS422-RS485-Application-Guide-Ebook.pdf>
- [7] Sen, S.K., *Fieldbus and Networking in Process Automation*, CRC Press, 2014.
- [8] Ustundag, A., and Cevikcan, E., *Industry 4.0: Managing The Digital Transformation*, Springer International Publishing Switzerland, 2018.
- [9] Wollschlaeger, M., Sauter, T., and Jasperneite, J., "The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0", *IEEE Industrial Electronics Magazine*, vol. 11, no. 1, pp. 17-27, 2017.
- [10] Yan, J., Meng, Y., Lu, L., and L. Li, "Industrial Big Data in an Industry 4.0 Environment: Challenges, Schemes, and Applications for Predictive Maintenance", *IEEE Access*, vol. 5, pp. 23484-23491, 2017.
- [11] Jeschke, S., *Industrial Internet of Things and Cyber Manufacturing Systems*, Jeschke, S., Brecher, C., Song, H., Rawat, D.B. (Eds.), *Industrial Internet of Things: Cybermanufacturing Systems*, Springer International Publishing Switzerland, 2017.
- [12] Schoder, D., *Introduction to the Internet of Things*, Hassan, Q.F. (Ed.), *Internet of Things A to Z*, John Wiley & Sons, Inc., 2018.
- [13] Yashiro, T., Kobayashi, S., Koshizuka, N., and Sakamura, K. "An Internet of Things (IoT) architecture for embedded appliances", *IEEE Region 10 Humanitarian Technology Conference*, pp. 314-319, Sendai, 2013.
- [14] Mendes, C., Osaki, R., and Costa, C., "Internet of Things in Automated Production", *European Journal of Engineering Research and Science (EJERS)*, vol. 2, no. 10, pp. 13-16, 2017.
- [15] Wang, L., Wang, X.V., *Cloud-Based Cyber-Physical Systems in Manufacturing*, Springer International Publishing AG, 2018.
- [16] Oks, S. J., Fritzsche, A., and Möslin, K. M., "An Application Map for Industrial Cyber-Physical Systems", Jeschke, S., Brecher, C., Song, H., Rawat, D.B. (Eds.), *Industrial Internet of Things: Cybermanufacturing Systems*, Springer International Publishing Switzerland, 2017.
- [17] Boubiche, D.E., Pathan, A.-S. K., Lloret, J., Zhou, H., Hong, S., Amin, S.O., and Feki, M.A., "Advanced Industrial Wireless Sensor Networks and Intelligent IoT", *IEEE Communications Magazine*, vol. 56, no. 2, pp. 14-15, 2018.
- [18] Alcaraz, C., Najera, P., Lopez, J., and Roman, R., "Wireless Sensor Networks and the Internet of Things: Do We Need a Complete Integration?", *1st International Workshop on the Security of the Internet of Things (SecIoT10)*, 2010.
- [19] Gilchrist, A., *Industry 4.0: The Industrial Internet of Things*, Apress, 2016.
- [20] <https://www.st.com/resource/en/datasheet/stm32f103c8.pdf>
- [21] <https://ae-bst.resource.bosch.com/media/tech/media/datasheets/BST-BME280-DS002.pdf>
- [22] https://infocenter.nordicsemi.com/pdf/nRF24L01P_PS_v1.0.pdf
- [23] <https://www.espressif.com/en/products/hardware/esp8266ex/resources>
- [24] <https://medium.com/@benjamindavidfraser/arduino-nrf24l01-communications-947e1acb33fb>
- [25] Gharghan, S.K., Nordin, R., and Ismail, M., "Energy Efficiency of Ultra-Low-Power Bicycle Wireless Sensor Networks Based on a Combination of Power Reduction Techniques", *Journal of Sensors*, vol. 2016, Article ID 7314207, 21 pages, 2016.
- [26] <https://github.com/stevenleeyongwah/NRF24L01>
- [27] <https://github.com/stevenleeyongwah/ESP8266>

3D Scene Reconstruction using a Swarm of Drones

Metehan Aydın

Department of Computer
Engineering
Ankara University, Faculty of
Engineering
Ankara, Turkey
e-mail: eng.metehan.aydin@gmail.com

Erkan Bostancı

Department of Computer
Engineering
Ankara University, Faculty of
Engineering
Ankara, Turkey
e-mail: ebostanci@ankara.edu.tr

Mehmet Serdar Güzel

Department of Computer
Engineering
Ankara University, Faculty of
Engineering
Ankara, Turkey
e-mail: mguzel@ankara.edu.tr

Abstract— 3D models of scenes are widely used but reconstructing them in 3D is a challenging task. The task generally starts from data acquisition and ends with 3D model of the scene. There are various methods to achieve the task but most of them don't offer a suitable solution for large scenes and implementing them is laborious even in the case of a small scene. The document is prepared to propose a solution to these problems and explain a project which implements the solution. The solution benefits from two approaches; a multi-agent drone system and an image-based 3D reconstruction pipeline. These approaches and how the project implements them are explained in details.

Keywords- 3D reconstruction; multi-agent; swarm; dense; computer vision; virtual reality; image-based; drone; uav

I. INTRODUCTION

Three-dimensional models of scenes are needed in various applications including 3D mapping, online shopping, video games, cultural heritage archival, virtual reality applications and crime scene investigation. There are three main techniques to model the scenes. These techniques are Computer Aided Design (CAD), optical range imaging sensors [1] and image-based 3D modelling [2].

3D modelling using a CAD tool takes much time to model, it is laborious and the model of scene is not realistic. Optical range imaging sensors can model a realistic scene and they don't take much time but they are very expensive and they can't be used for some surfaces which don't reflect the laser beams back. Image-based 3D modelling offers a cheaper, realistic and generic solution. But, taking photos of the scene manually by a hand-held camera restricts the capabilities of the image-based 3D modelling. Because, areas which have buildings or which is large can't be photographed using a hand-held camera and taking photos of the area in the desired format for image-based 3D modelling is very difficult.

The purpose of the project is removing these restrictions and using the whole capabilities of image-based 3D modelling. The project removes the restrictions using a multi-agent drone system. The multi-agent system captures the videos of the scene. Then, an image-based 3D modelling pipeline is applied to the video to reconstruct the scene in 3D.

The rest of the document explains how the project implements the multi-agent drone system and the image-based 3D modelling pipeline. In Section 2, the multi-agent system used in the project is explained. Section 3 presents the coverage algorithm of the project and 3D reconstruction process is explained in Section 4. The results of the project is in Section 5 and there is an conclusion in Section 6.

II. MULTI-AGENT DRONE SYSTEM

The project is implemented using two drones but it is designed to be able to have four, two and one drones whose models should be DJI Phantom Professional 3 [3] but the system is suitable for different model drones. DJI Phantom Professional 3 must be controlled with a remote controller and the remote controller must be connected to a device whose operating system must be Android or iOS. In order to benefit some of the drone features which include programming the drone, DJI provides a Software Development Kit (SDK) [4]. There are two version of the SDK. These are Android SDK and iOS SDK. The project is programmed using Android SDK. The SDK provides Java Programming languages so the project is implemented using Java Programming Language. As mentioned above, the system can have different number of drones. In response to mission requirements, desired number of agents can be used. The multi-agent system provides us flexibility, robustness and scalability [5]. Flexibility gives the ability of generating effective solutions to different problems which varies in size. Robustness gives the ability of completing the task despite of failures on the environment and the agents. Scalability gives the ability of adding new agents easily to the system and working properly with the new agents. These features give multi-agent systems some advantages against a single system. These advantages [6][7] are listed below.

- Multi-agent system is cheaper and consumes less energy.
- Multi-agent system is faster.
- Multi-agent system is more reliable.

In order to share the task between agents, the agents must communicate. In a multi-agent system, the agents can communicate using different methods [8]. One of the them is direct communication. With direct communication, the agents can communicate directly by sending and getting

messages. In the project, these messages are delivered using a Wireless Local Area Network (WLAN). The network is provided from Android devices. Almost all Android devices can be used as a modem, the system can use the Android device of an agent as a modem and can create a WLAN. The messages are transferred over the WLAN by sockets provided by Java Programming Language. Transmission Control Protocol (TCP) is used as transmission protocol. There are three types of messages in the system. These types are connection messages, command messages and mission completed messages. The connection messages have two types. These types are add agent and confirm agent. They are actually strings and they have following format.

- ADDAGENT#ip_address
- CONFIRMAGENT#ip_address

When the task is wanted to be shared among agents, the IP addresses of the agents must be known. In the user interface of the program, there is a button to show the IP address of the agent. The IP address of an agent can be learnt using the button. There is an agent list in each agent. The agents are added there and the task is shared among them. When an agent wants to add the other agents, it follows the following algorithm.

1. Check if the agent hasn't been added yet.
2. If it hasn't been added, send an add message.
3. Wait for the confirm message.
4. When the confirm message is taken, add it to the agent list.

The agent which receives the add message just send a confirm message to the sender and adds the sender to the agent list. There is no need to check whether the agent has already been in the list because the algorithm ensures that if an agent gets an add message, the sender of the add message is not in the agent list of the receiver. The algorithm for the receiver side is as follows.

1. If there is an add message, add the sender of the message to the agent list
2. Send a confirm message to the sender.

Fig. 1 shows the whole process briefly. In a multi-agent system, the agents can be controlled in two ways [9]. One of the ways is centralized control. In centralized control, one of the agents becomes the master agent and it controls the other agents. The project is implemented using centralized control but the master agent is not fixed. Instead, all the agents can be a master. It means that all the agents can control the other members. Fig. 2 can give an idea about the control mechanism. When an agent is given a mission, it first divides the mission for others and defines which parts the others will cover. it sends this information using an instance of a class whose name is MissionInformation. A command message is actually this instance. In response to each agent in its agent list, it creates an instance and sends it

over the network. When the other agents get the instance, they stop their work and start the given mission.

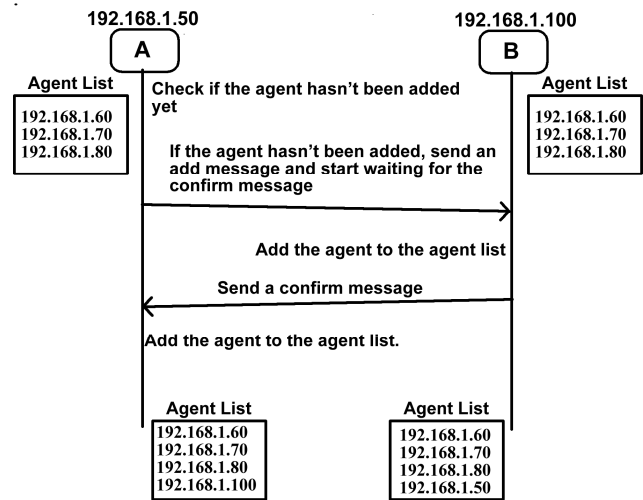


Figure 1. Adding an agent to the system

In order to ensure the mission is completed successfully, the agents use mission completed messages. There are three types of mission completed messages. These are success, not completed, error, state. These are actually string and they have following format.

- SUCCESS#ip_address
- NOTCOMPLETED#ip_address
- ERROR#ip_address
- STATE#ip_address

When an agent completes the task, it sends a success message to the master. If the master completes the mission earlier than the other agents, it checks if there are any agents which don't complete the mission. If there are, it sends a state message to the agents. When the agents receive the message, they send a not completed message if there is not any error and they can complete the task. If there is an error and they can't complete the task, they send an error message. When the master agent receives an error message, it also completes the mission of the sender of the message.

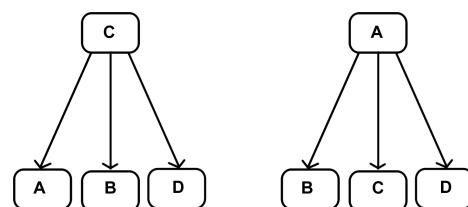


Figure 2. Selection of master

III. COVERAGE

In the project, an area is defined with four points. These points are selected using a map. The area inside the points is covered to reconstruct in 3D. Fig. 3 shows a selected area.

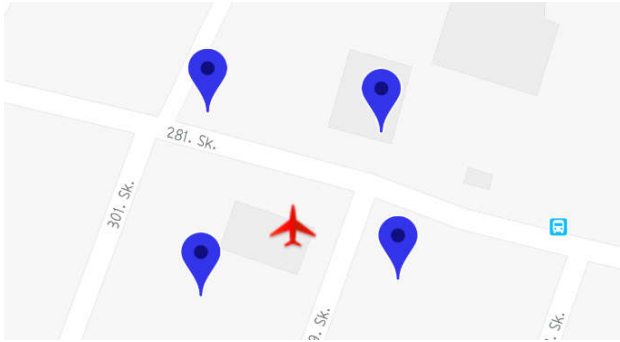


Figure 3. Selected area

In order to reconstruct an object in 3D, the object must be photographed in all the viewpoints. Following figure can give an idea.

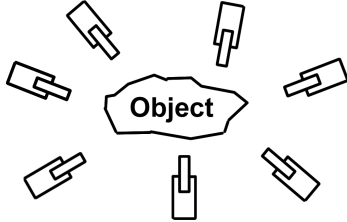


Figure 4. Photographing an object

In order to take photos in all the viewpoints of the objects inside the area, the agents must rotate around it. Fig. 5 shows the rotation of the drones around an area.

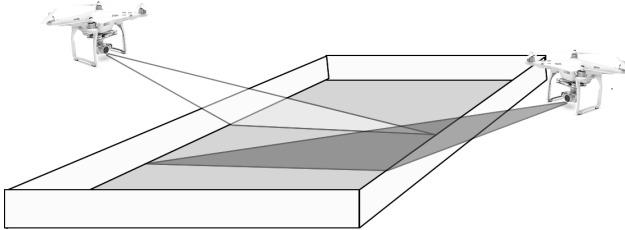


Figure 5. Coverage of an area by agents

As it can be seen in Fig. 5, while the agents are rotating, the cameras must be directed the area. This means the heading direction of the drones must always be directed to area. This can be achieved by rotating the agent in every point with respect to the angle between the edges which are connected to the point. The angle and the edges are shown in Fig. 6. The altitude of the agents must be determined according to two variables. These variables are field of view angle of the camera on the agent and the longest edge of the area. Selecting the longest edge as a variable to determine the altitude results covering a larger area. But, when the scene is reconstructed, the undesired parts can be deleted. In the project, the field of view angle of the agents is 92 degrees. Initially, the camera of an agent is parallel to ground. When a task is given and the camera starts the record, the angle of the camera is increased 45 degrees in the clockwise direction. This allows to record right under

the drone and there are 47 degrees angle left to record the given area. The altitude of the drone is selected same as the length of the longest edge. This results covering a little larger area but as we stated above, when the scene is constructed, the undesired parts can be deleted.

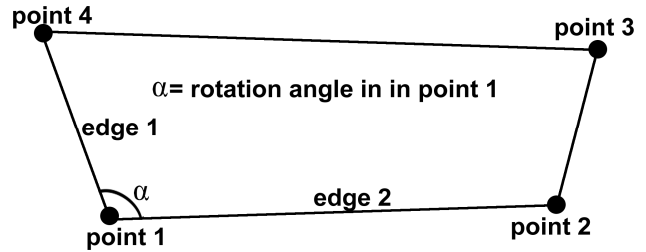


Figure 6. Rotation angle

IV. 3D RECONSTRUCTION

When the coverage is over and we get the videos of the scene, the videos are divided into frames. In a video, there are a lot of frames which are almost same and these frames do not contribute reconstruction. They are wasting time and performance. It is better to determine the frames which represent these frames. This process is known as key framing. A program is designed as a part of the project to extract the key frames. It simply takes the video as an input and gives the key frames of the video as an output. In order to determine a key frame, the program follows the following algorithm.

1. Store the first frame as a key frame and select it as the last key frame.
2. For the frames after the last key frame, compute the similarity between the key frame and the frames.
3. If the similarity is above a threshold, eliminate the frame
4. If the similarity is under a threshold, store it and select as the last key frame.

The program is designed to be able to work with multiple threads. That means the key framing process can be divided into threads. The purpose of this design is increasing performance and decreasing the process time. The program can have only one thread or it can also have N number of threads. Unfortunately, increasing in the number of threads does not result in an increase in the performance of the program. We tried the program with a variety number of threads to determine the number with which the performance is maximized. Table 1 shows the relation between the thread number and performance.

Table 1. Relation between number of threads and number of frames.

Number of Threads	Processed Number of Frames
1 Thread	10 Frames
3 Threads	19 Frames
5 Threads	26 Frames
10 Threads	25 Frames

As it is seen in the table, it can be better to work with five threads. When the key frames are extracted, the scene is ready to reconstruct. In order to reconstruct the scene; first, the camera parameters and 3D point cloud of the scene must be known. In the project, these are provided with a technique whose name is Structure from Motion (SfM) [10]. There are several Structure from Motion algorithms and tools which implement these algorithms. Although some of the algorithms can have different techniques, most of them follow a general work flow. The work flow is shown in Fig. 7.

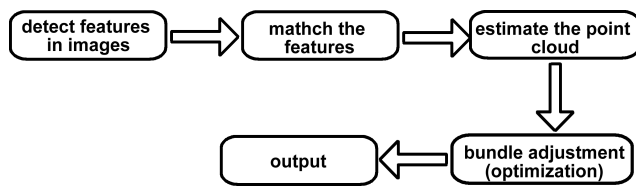


Figure 7. Work flow of Structure from Motion algorithms

In the project, a software whose name is VisualSFM [11] is used as an SfM tool. The software takes the images as input then gives the camera parameters and the point cloud of the scene as output. Then, the output is used for dense reconstruction. Dense reconstruction is the last step of 3D reconstruction process, it gives us 3D model of the scene. In the project, the dense reconstruction is provided by a software whose name is Clustering Views for Multi-view Stereo (CMVS) [12]. CMVS is a software which takes the camera parameters, the point cloud of the scene and the images as input then produces 3D dense of the scene as output. The software can be integrated with VisualSFM and they can work together properly. One of the main reason why we choose CMVS tool is its compatibility with VisualSFM. There is button in VisualSFM to run directly CMVS software. When the button is clicked, the output of the VisualSFM becomes the input of the CMVS software. Thus, the scene is reconstructed using just one interface. The output of the CMVS is a 3D dense of the scene and it can be displayed using a software. In the project, MeshLab [13] software is used to display the dense.

V. RESULTS

The project had been purposed to reconstruct the area in 3D with two reliable and flexible approaches and it met the objective successfully. The project can cover, thus, reconstruct larger areas using these approaches. These two approaches also made the project easy to use. Fig. 8 shows an output from the project.

VI. CONCLUSION

The work presents a 3D reconstruction way using two flexible and automated approaches. Most of the researches uses image based 3D reconstruction approach with an hand held camera. The work showed that how a multi-agent

system can be used to reconstruct an area in 3D using an image based approach. The work also showed how powerful a multi-agent system and image based 3D reconstruction.



Figure 8. An output from the project

VII. REFERENCES

- [1] Besl, Paul J. "Active optical range imaging sensors." *Advances in machine vision*. Springer, New York, NY, 1989. 1-63.
- [2] Remondino, Fabio, and Sabry El - Hakim. "Image - based 3D modelling: a review." *The photogrammetric record* 21.115 (2006): 269-291.
- [3] <https://www.dji.com/phantom-3-pro>
- [4] <https://developer.dji.com/mobile-sdk/>
- [5] Iñaki Navarro and Fernando Matia, "An Introduction to Swarm Robotics," *ISRN Robotics*, vol. 2013, Article ID 608164, 10 pages, 2013
- [6] Tan, Ying, and Zhong-yang Zheng. "Research advance in swarm robotics." *Defence Technology* 9.1 (2013): 18-39
- [7] Jan Carlo Barca and Y. Ahmet Sekercioglu (2013). *Swarm robotics reviewed*. Robotica, 31, pp 345359 doi:10.1017/S026357471200032X
- [8] Cao, Y. Uny, Alex S. Fukunaga, and Andrew Kahng. "Cooperative mobile robotics: Antecedents and directions." *Autonomous robots* 4.1 (1997): 7-27.
- [9] Mohan, Yogeswaran, and S. G. Ponnambalam. "An extensive review of research in swarm robotics." *Nature & Biologically Inspired Computing*, 2009. NaBIC 2009. World Congress on. IEEE, 2009.
- [10] Westoby, Matthew J., et al. "'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications." *Geomorphology* 179 (2012): 300-314.
- [11] Wu, Changchang. "VisualSFM: A visual structure from motion system." (2011).
- [12] Furukawa, Yasutaka, et al. "Towards internet-scale multi-view stereo." *Computer Vision and Pattern Recognition (CVPR)*, 2010 IEEE Conference on. IEEE, 2010.
- [13] <http://www.meshlab.net/>

Development of a Smart Helmet for Digital Data Collection and Applications on Construction Sites

Eray Varyeter², Kevin Jachmann¹, Mario Rahal², Simon Nicklas¹, Konstantin Krahtov³, Prof. Dr. -Ing. Sven Rogalski^{1,2}

¹Darmstadt University of Applied Sciences, Darmstadt/Germany

²GFTN – Gesellschaft zur Förderung technischen Nachwuchses – Darmstadt e.V., Darmstadt/Germany

³Open Experience GmbH, Karlsruhe/Germany

Abstract – This paper presents the development and implementation of a smart helmet to be used on construction sites in order to collect digital data, such as 360 degree images and 3D point clouds, by using two embedded cameras with fisheye lenses and a stereovision camera connected to an embedded processing unit. Collected digital data is processed and used for recording the actual state, inspecting the construction site, doing automatic measurements, creating an indoor navigation system and generating a built-in documentation.

Keywords – Digitalization, Construction 4.0, 360 Degree Imaging, Indoor Navigation, Point Cloud Processing.

I. INTRODUCTION

Digitalization can be described as the usage of information and communication technologies to improve and transform business operations, models and processes. All enterprises are aware and affected by digitalization from many aspects, however, the construction industry still lacks implementations of digitalization, especially in Germany, Austria and Switzerland [1]. The aim of digiBau is to develop a modular and adaptable system on construction helmets, as well as on drones, in order to collect 360 degree images and 3D point clouds that will be used for recording and inspecting the current state of the construction site, doing automatic measurements, creating an indoor navigation system and generating built-in documentation. This paper explains the concept, development, implementation and verification of this system on construction helmets.

In section II, analysis, concept and development of 360 degree image capturing, and face detection algorithms are explained. Section III explains the analysis and evaluation of different 3D technologies as well as the development of the indoor navigation system. Section IV covers the developed 3D applications and tools. Section V explains the mechanical design and the power management system. Finally, the paper provides a summary conclusion in section VI.

II. 360 DEGREE IMAGE CAPTURING

360 degree imaging has been very popular and used recently in many different applications since it covers a wide field of view and provides the details of the surroundings. The conventional method for creating 360

degree images is using systems that contain multiple wide angle cameras and stitch all captured images by finding and matching the feature points. However, there are some challenges and disadvantages of the conventional methods. This algorithm often fails if there are not a sufficient number of objects that can be used as common feature points in the overlapping regions of the captured images. Since the helmet will be used on construction sites, overlapping regions will have only empty walls or windows, which will lead the stitching algorithm to fail because of insufficient number of common points. Additionally, this conventional method requires many cameras in order to run without problem. Therefore, it was decided to use a different method for creating a 360 degree image, which will decrease the cost and complexity, and also guarantee that the software will run without depending on the existence of objects that could be used as common feature points in the overlapping regions [2].

The 360 degree image capturing algorithm in this paper uses 2 embedded cameras with fisheye lenses, which enable capturing 195 degree field of view images from the cameras placed on the left and the right side of the helmet. A fisheye lens is an ultra-wide angle lens that generates strong visual distortion and achieves extremely wide angles of view, which is intended to create a wide panoramic image [3]. After researching possible processing embedded units, it was decided to use NVIDIA Jetson TX2, which is an embedded system on module that provides access GPU usage for computation together with CPU, which makes it capable of deploying complex computer vision and deep learning algorithms easily and quickly.

The first step in this algorithm, after capturing images with fisheye lenses, is intensity compensation. The intensity of fisheye images is reduced at the edges compared to the center because of the strong distortion. This effect is called the vignetting effect and can be seen in figure 1.

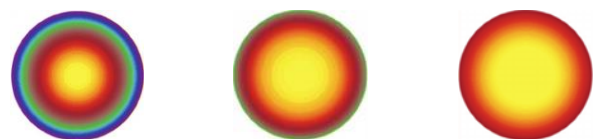


Figure 1: Example intensities of (a) fisheye lens, (b) wide-angle lens, (c) normal lens

It can be clearly seen from the figure above that this effect will cause the final 360 degree image to have a nonequal intensity profile. In order to compensate this effect, an image of a large white blank paper was captured, and the intensity profile was extracted, which can be seen in figure 2 below [2].

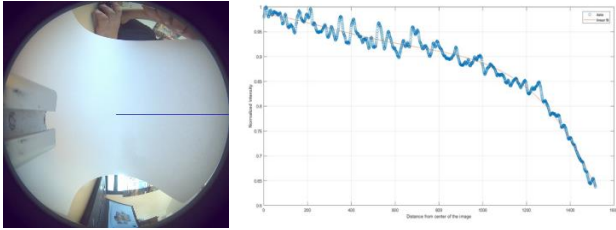


Figure 2: (a) Intensity extraction, (b) Extracted intensity profile

A polynomial function was fit to get an approximation of how the intensity changes over the radius. Then, each pixel was compensated by multiplying with a coefficient that was calculated based on the output value of the polynomial function [2].

The second step in the algorithm is image dewarping, which is the process of geometric transformation of an image to reverse the effects of visual distortions caused by the fisheye lenses. Image dewarping consists of two operations. The first projects each pixel from the input 2D image to a 3D point in a unit sphere. The second maps each 3D point to spherical 2D point that is compatible with the 360 degree image viewer. These steps can be seen in figure 3 [2].

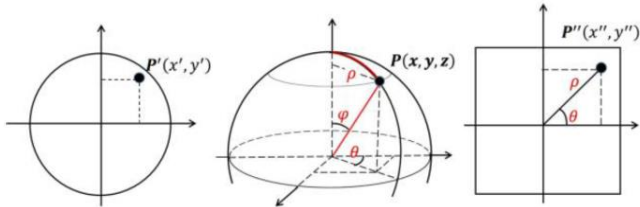


Figure 3: (a) 2D point in fisheye image, (b) 3D point in unit sphere, (c) Spherical 2D point

$P'(x', y')$ is the example pixel in the fisheye image. Width, height and field of view in degrees of the image are shown as W , H and f , respectively. To calculate pitch and yaw, see equation 1 below.

$$\theta_s = f \left(\frac{x'}{W} - 0.5 \right) \quad \varphi_s = f \left(\frac{y'}{H} - 0.5 \right) \quad (1)$$

$P(x, y, z)$ is the projected 3D point, where x , y and z are the 3D coordinates and calculated with equation 2 below.

$$x = \cos(\varphi_s) \sin(\theta_s), y = \cos(\varphi_s) \cos(\theta_s), z = \cos(\varphi_s) \quad (2)$$

The distance between the projected center and the 3D point is shown as p and derived by equation 3 below.

$$p = \frac{H}{f} \tan^{-1} \left(\frac{\sqrt{x^2 + z^2}}{y} \right) \quad (3)$$

Then, the 2D spherical (equirectangular) point $P''(x'', y'')$ is constructed. The calculation of spherical projected point can be seen in equation 4 and 5 below.

$$\theta = \tan^{-1}(z/x) \quad (4)$$

$$x'' = 0.5W + p \cos(\theta), y'' = 0.5H + p \sin(\theta) \quad (5)$$

Note that dewarping the left and right images does not guarantee alignment between them. In order to align the images, checkerboard patterns are placed in the sights of overlapping regions. Two mosaic images are prepared, one of which has the overlapping regions from the left image and the other has them from the right image. Control points are selected manually on both of the mosaic images and used to estimate an affine transformation matrix, which is then applied on the middle image for alignment [2].

After the affine transformation, overlapping regions of two images are blended using a ramp function to create the final 360 degree image. Blended pixel $B(r, c)$ at row r and column c is computed by the equation below.

$$B(r, c) = a_1 x L(r, c) + a_2 x R(r, c) \quad (6)$$

$L(r, c)$ and $R(r, c)$ are the pixels at position (r, c) that are taken from the left and the right overlapping regions. When the left overlapping region is blended, the coefficients of a_1 and a_2 are computed by using the equation below.

$$a_1 = (n - c + 1)/n, \quad a_2 = c/n \quad (7)$$

where n is the width of the overlapping region and c is the current processed column in the loop. Computations of the coefficients are reversed when the right overlapping region is blended [2]. Blending can be illustrated in figure 4.

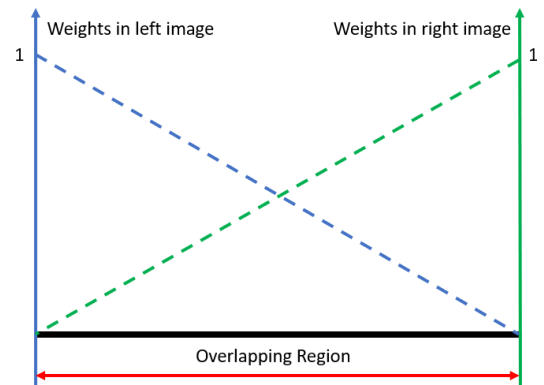


Figure 4: Blending of the overlapping region

The result of blending is the output 360 degree image that can be viewed with a 360 degree viewer. After creating the 360 degree image, face detection is applied to protect the privacy of workers on the construction site. A pre-trained deep neural network caffe (Convolutional Architecture for Fast Feature Embedding) model is used for facial detection. Caffe works by using different classes such as blobs, layers,

net and solver. Storing the information is done in blobs. Each layer in caffe model transforms the information stored in bottom blobs to top blobs. Net consists of many layers, which compute gradients forward and backward. The job of solver class is updating the weights by using gradients [4]. An example 360 degree image captured on a construction site can be seen below.



Figure 5: 360 degree image

III. INDOOR NAVIGATION

The remaining goals of the project, after creating a 360 degree image, can only be achieved by using 3D technology. There are many different technologies with some advantages and disadvantages. Pulsed time of flight cameras (LiDAR) provide wide field of view thanks to the scanning process and they don't need light for operating, however, they are highly dependent on the reflective properties of the material, and they have bulky size and high cost, which makes them not suitable for the project because this system will be used on different surfaces that have different materials. Continuous wave time of flight cameras have the same problem as LiDARs. Structured light cameras don't need light for operating, they have high accuracy at short range, and they are not dependent on the reflective properties, however, the setup of the system is complex and installation cost is very high. Stereovision cameras have traditional and simple components, low cost, high accuracy at short range and no dependency on the reflective properties. They have some disadvantages such as requiring high processing power to generate a depth map and not performing well in low light conditions, but these disadvantages can be discarded since Jetson TX2 has a high processing power and the helmet will be used only in the construction sites where the environment is provided by good light conditions for workers [5].

After analyzing different stereovision cameras, it was realized that most appropriate option was using ZED mini camera from Stereolabs, which is a partner company of Nvidia in vision applications. ZED mini API is designed in a way that uses GPU power automatically when capturing images and point clouds. Other big advantage of ZED mini is that it is able to estimate its relative position to the world

around it with a position accuracy of ± 1 mm and orientation accuracy of 0.1° by using both motion sensors (gyroscope and accelerometer) and visual odometry technique [6]. This feature is very useful since one of the goals of the project is to develop an indoor navigation system.

ZED mini combines the information from motion sensors and visual odometry to track its relative position to the starting point. Since the obtained information is only relative change to its starting point, it is important to have a pre-defined reference position and orientation for each floor in the buildings. Some other features of floors and buildings are also necessary to be known to handle captured images and point clouds easily, which makes it essential to create a structure. This structure can be seen in table below.

Table 1: Building and Floor Feature Structure

Building Features	Floor Features
Building ID	Floor number
Name	Width of floor plan image
Street	Height of floor plan image
Number	Actual width of the floor
City	Actual height of the floor
State	QR code reference pos in X
Country	QR code reference pos in Y
ZIP Code	Path of floor plan image
	Reference orientation

Building ID and floor number are used as the folder names to save the data. All these features are saved in a text file under their relative folder. Each building folder contains all taught floors by the user and a BuildingFeatures.txt file, which building features structure elements are stored. Each floor folder contains a floor plan image, captured images and point clouds, and a FloorFeatures.txt file, which floor features structure elements are stored. Every feature element of the building and floor structure is decoded in QR codes, which are printed and hung on the walls on the construction sites. Users are provided with an option to scan the QR code with ZED mini, which automatically loads the floor plan image to the user interface, display the captured images and point clouds on the floor plan image with their relative capturing positions and display the pre-trained reference position and orientation. Users are provided with options such as starting and stopping of the tracking and the navigation, selecting a destination point on the floor plan image which then estimates the distance and the time to reach to the destination point. Users can also start capturing 360 degree images and 3D point clouds from the scene and save them with their position information under the specific building and floor folder. Captured images and point clouds are shown with letters B and W respectively, which stand for Bild (image) and Wolke (cloud) in German. Example indoor navigation application can be seen in figure 6 below.

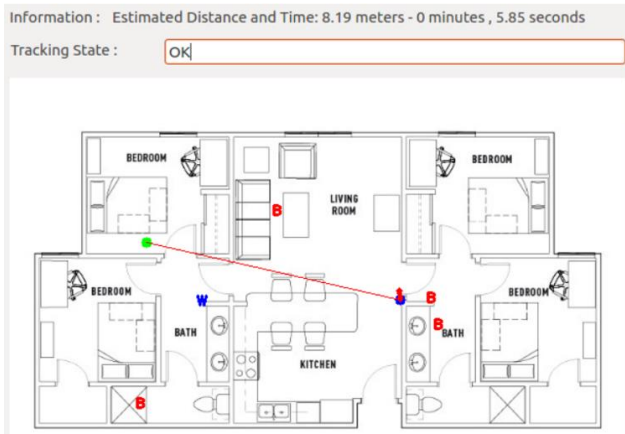


Figure 6: Indoor navigation

Two additional tools are designed to handle QR codes and floor plan structures, first of which is for teaching a new floor plan structure. Users are provided with a user interface where they can enter all the feature information, load a floor plan image and then teach the new folder structure in the file system. Scanning an existing QR code from the construction site, loading all information automatically and then modifying is also provided to avoid writing everything manually when most of the features are already same with an existing QR code. Second tool is developed for generating QR codes of new floors to print and hang them on the walls in the construction site. Users are again provided with a user interface where they can enter the information or load existing QR code structure from floor plans, and then generate and save new QR code to the system, which can then be used for printing and hanging on a wall in the construction site.

IV. 3D APPLICATIONS AND TOOLS

After implementing indoor navigation system, some other 3D tools are developed for users, such as depth sensing, distance measurement, importing clouds and point cloud processing with Point Cloud Library (PCL).

A. Depth Sensing

Depth map is a representation of distance values in the direction of camera's sight. Generated depth map is grayscale image where each pixel is represented with 8 bits and has values varying between 0 and 255 for the closest and the furthest depth, respectively. Depth sensing tool is developed for users to easily measure the distance between any object they move the mouse cursor in the scene and the camera automatically. Stereovision cameras calculate the depth of each pixel depends on focal length, baseline and disparity. Calculation can be seen in the following equation.

$$\text{Depth} = (\text{Focal Length} \times \text{Baseline}) / \text{Disparity} \quad (8)$$

Focal length can be described as the measure of how strongly the light is converged or diverged by the system. Baseline is the distance between the centers of each lens in the stereovision camera. Disparity is the pixel location difference of an object from left and right images [7]. Concept of depth calculation can be seen in figure 7 below.

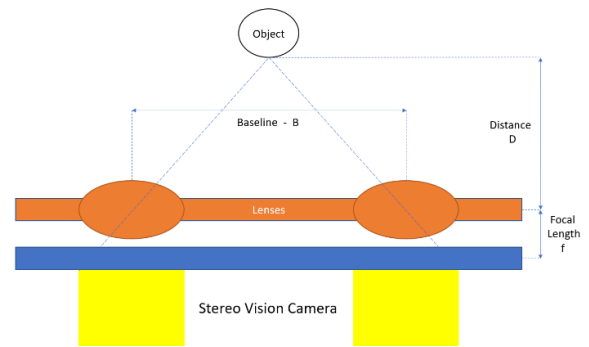


Figure 7: Measuring depth with stereovision camera

ZED mini also provides a confidence map, which represents the confidence of measured distance as values varying between 0 and 100. Surface normal map is also provided for measuring the normals that are used for estimating the traversability and real time lighting. Depth sensing tool can be seen in figure 8 below.

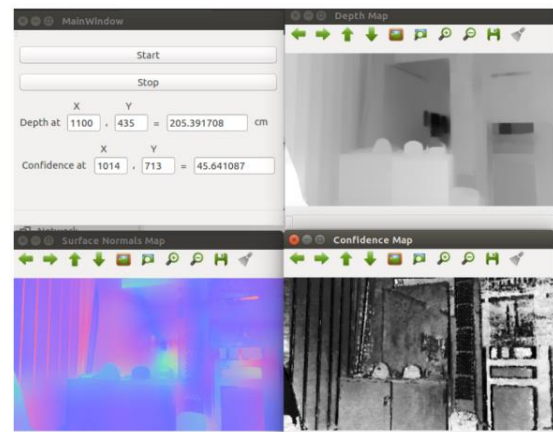


Figure 8: Depth sensing tool: (a) GUI, (b) depth map, (c) surface normals map, (d) confidence map

B. Distance Measurement

Another developed tool for users is distance measurement between the user selected points on the live image. This tool is very useful when a worker on a construction site wants to measure a length of an object without having to go there physically. 3D positions of the selected points have to be known in order to measure the distance between them, which requires to use point clouds. Users are provided with

a user interface where they can start and stop capturing live images from ZED mini and draw a line on the image.

3D points of selected pixels are used to calculate the distance. Euclidean distance calculation can be seen from the equation below.

$$\sqrt{(P1.x - P2.x)^2 + (P1.y - P2.y)^2 + (P1.z - P2.z)^2} \quad (9)$$

where P1 and P2 are the selected points. Distance measurement tool can be seen in figure 9 below.

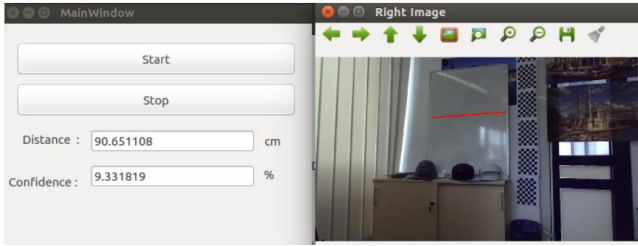


Figure 9: Distance measurement: (a) GUI, (b) live image

C. Importing Clouds

This tool is developed for users to import a cloud to the scene in order to check manually if it fits inside of the scene cloud. An example for this application can be testing whether a furniture is placeable in a room. Users are provided with a user interface where they can translate and rotate the imported cloud in all directions. Translation and rotation are the geometric transformations that can be defined as mappings from one coordinate system onto itself. PCL is using transformation matrix, which is a 4x4 size matrix, to apply translation and rotation on point clouds. Example transformation matrix can be seen in table 2, with blue parts containing information about rotation and green parts containing information about translation [8]. Tool for importing clouds can be seen in figure 10 below.

Table 2: Rotation in X with β , and translation in Y with α

1	0	0	0
0	$\cos(\beta)$	$\sin(\beta)$	α
0	$-\sin(\beta)$	$\cos(\beta)$	0
0	0	0	1



Figure 10: Importing clouds: (a) scene and imported cloud, (b) imported cloud

D. Point Cloud Processing

This tool is developed for users to have the option to apply different functionalities of point cloud library. Users are provided with a user interface where they can set parameters, apply and see the output cloud. First functionality is translation and rotation in any direction. Concept of transformation is already explained in the previous section. This tool can be seen in figure 11 below.

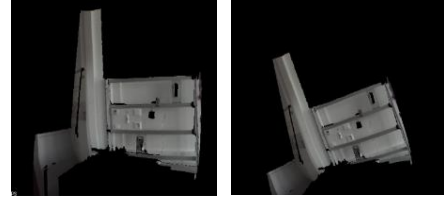


Figure 11: Transformation: (a) input, (b) output

Second functionality is searching points within certain centimeters of centroid. This tool is developed for searching points in a range of defined radius. Point cloud library has a tree based data structure, called octree, in order to manage, search and compress 3D points. Each node of octree is called voxel, which has either eight children or no children. Searching algorithm is using the octree structure to store the 3D points. Octree object keeps a vector of point indices in its nodes and runs a radius based search and extract the indices of the points which are in the specified radius. Points with extracted indices are combined to create the final cloud [9]. This tool can be seen in figure 12 below.

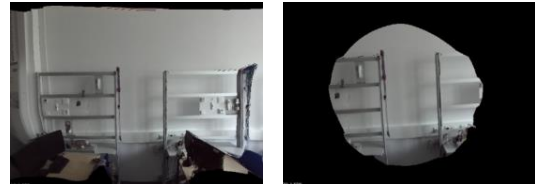


Figure 12: Searching points in a range, (a) input, (b) output

Next functionality is passthrough filtering, which is used for removing the points from a cloud that are not in the user given range. This tool is useful when the users want to discard all the unneeded objects from the cloud. Users are provided with a user interface where they can select the axis and the range of the passthrough filter. This tool can be seen in figure 13 below.

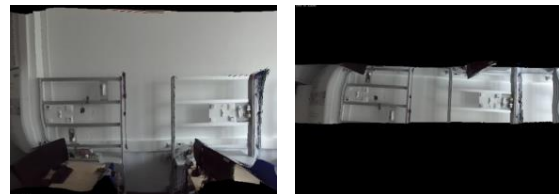


Figure 13: Passthrough filter, (a) input, (b) output

Another developed functionality is outlier removal. Outliers can be described as the undesired noises caused by the 3D sensor's inaccuracy. Outliers result in calculation errors because the inaccuracy of 3D sensor registers measurements where there should not be any measurement. There are different advantages of removing outliers, such as speeding up the computation and obtaining better and more precise depth values [10]. This tool can be seen in figure 14.



Figure 14: Outlier removal, (a) input, (b) output

Downsampling is another developed functionality of point cloud processing tool. One of the biggest problems with point cloud processing is that clouds have a lot of information, which makes it very hard to process because it requires a lot of computation power and time. Downsampling is one of the most efficient methods to reduce the complexity and the number of points of a cloud. PCL provides voxel grid structure for this purpose, which can be considered as cubes, whose sizes are defined by the user. Centroid of each voxel grid is calculated and only one point per each cubic voxel grid survives, others are removed. By this way, the cloud still preserves its geometric and shape characteristics and has much less points than its original one. This tool can be seen in figure 15 below.

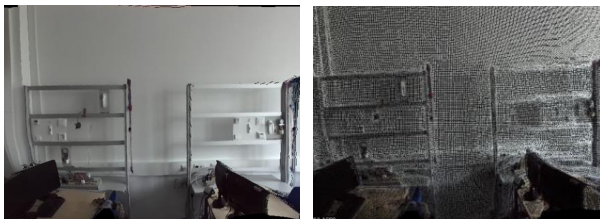


Figure 15: Downsampling, (a) input, (b) output

Next functionality is called upsampling, which can be considered as the opposite operation of downsampling. This operation is basically a surface reconstruction algorithm which works by integrating the existing points in the cloud. PCL is using Principle Component Analysis (PCA) and Moving Least Squares (MLS) method to implement upsampling. PCA method is applied to reduce the number of elements by creating linearly independent features that represent the whole data best while keeping the loss of information minimum. Extracted features, also called principal components, give the idea about where local

planes can be in the point cloud [11]. After the algorithm fits these planes, PCL is using MLS to generate a polynomial function in the set of distances from the points to the actual surface, which locally minimizes the least square error. After fitting this function, points are projected back to the surface. Then, algorithm samples the extracted functions of the surface to perform upsampling. Users are provided with a user interface where they can enter radius and step size. Sampling radius is the radius around each point where the local plane is sampled. Step size is used when the algorithm reconstructs the surface. Bigger values of step size result in less new point in the output cloud [12]. This tool can be seen in figure 16 below.

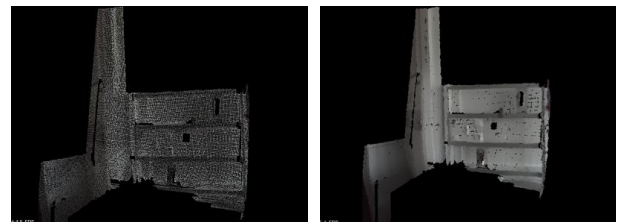


Figure 16: Upsampling, (a) input, (b) output

Another functionality is surface smoothing. All depth sensor might produce inaccurate results and have measurement errors. In order to compensate these errors, PCL provides surface smoothing algorithm, which iterates all points and interpolates by using MLS algorithm. As a result of this interpolation, original surface is reconstructed. Resulting cloud has more accurate normals than the original cloud. Users are provided with a user interface where they can set the radius of search. Algorithm is approximating the surface and computing the normals by estimating a polynomial function. This tool can be seen in figure 17.

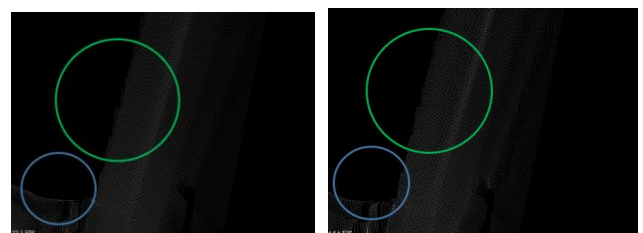


Figure 17: Surface smoothing, (a) input, (b) output

Next functionality is called euclidean segmentation. Segmentation of point clouds is the operation of breaking it apart into different pieces, which are called clusters. Main goal of segmentation is being able to process them individually. Segmentation provides different clusters, each of which the users can select, obtain the model and process individually. Concept of euclidean segmentation and this tool can be seen in table 3 and figure 18 below [13].

Table 3: Steps of euclidean segmentation

Set up an empty list of clusters C and an empty queue Q
For every point p_i of P , perform the following steps <ul style="list-style-type: none"> - Add p_i to the Q - Calculate euclidean distances d_k for all p_k points to the current point p_i - Search for all p_k points that are in a sphere with d_k < d_{th} - Add p_k points to the queue Q only if <ol style="list-style-type: none"> 1. They were not processed before 2. Maximum number of points that Q can have (defined by the user from GUI) are not exceeded - Add Q to the list of clusters C if it has more points that the minimum number of points Q must have (defined by the user from GUI)
Terminate the algorithm when all points p_i of P are processed and placed in C

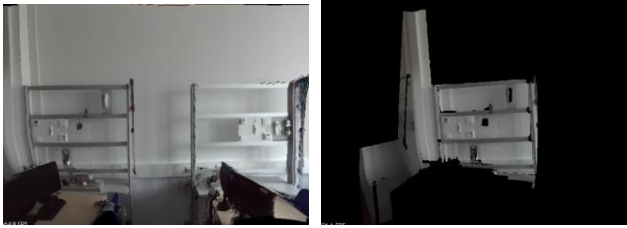


Figure 18: Euclidean segmentation, (a) input, (b) output

Last functionality of this tool is plane segmentation. PCL uses RANSAC (Random Sample Consensus) algorithm to fit a set of points to a model, which can be described as an iterative method that is used for robust model fitting. RANSAC selects a sample of set of three random points and computes the model and its parameters of the corresponding plane. Model computation is done by using the plane equation below.

$$a(x - x_0) + b(y - y_0) + c(z - z_0) + d = 0 \quad (10)$$

Then, the algorithm detects all points of the original cloud that belongs to the computed plane model by using the distance threshold value obtained from the developed user interface. RANSAC checks if the distance between the points and the plane is less than the specified threshold. If it is less, the algorithm counts them as inlier points. This process is repeated until a sufficient confidence is achieved. The algorithm compares the obtained result with the last saved one and replaces the saved result with the new one if it is better. If the algorithm can't achieve this confidence within the maximum number of iterations, then the output of the algorithm will be empty. This tool can be seen in figure 19 below [14-16].

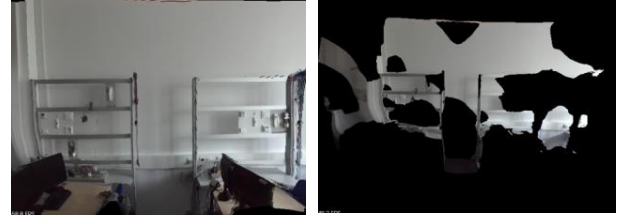


Figure 19: Plane segmentation, (a) input, (b) output

V. MECHANICAL DESIGN AND POWER MANAGEMENT SYSTEM

Another important aspect of this project is to develop a solution for mechanical design of the construction helmet. Firstly, requirements are clearly defined. Helmet should be wearable by anyone on the construction site, which means it is necessary to have an adjustment tool to adjust the helmet size depending on their head size. Cameras for 360 degree image capturing should be fixed in a way that they will not move even if it's being pushed forward or pulled backward accidentally. Since the embedded processing unit will be used all the time, it will generate a lot of heat. In order to allow heat dissipation, there should be some holes on the top cover of the helmet. All electrical parts should be fixed in a way that they will not be affected by mechanical movements, shocks and vibration. They should be isolated from direct contact to the head of the user. Weight and center of mass of the helmet should be considered that it is carryable by a human. Proper place for DC power jack should be arranged to power up the helmet. Since the positions of the left and the right cameras should be adjusted carefully to get the best 360 degree image, there should be a fixture inside the helmet that can adjust the positions of the cameras. Design was completed according to the defined requirements and printed using 3D printer. Final design and printed helmet can be seen below.



Figure 20: Final design, (a) front view, (b) back view

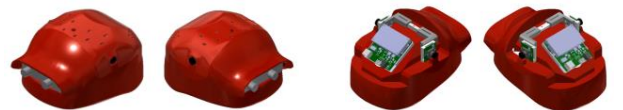


Figure 21: Final design cross view, (a) with, (b) without top cover



Figure 22: (a) Final design bottom view, (b) adjustment tool



Figure 23: Printed helmet, (a) top, (b) front, (c) bottom view



Figure 24: Printed helmet, (a) top view, (b) side view

Power management system was also designed and implemented to supply the necessary power to the helmet. System consists of main battery pack with regulator board, connector and interface board, and monitoring and output board. Main battery pack consists of 6 Lithium-Ion cells, each of which is 4.2 Vdc and has 3000 mAh power capacity. Output voltage is regulated through a dedicated battery management circuitry to insure proper output voltage. Connector and interface board provide a way to connect other devices to the battery, as well as power on/off switch. Monitoring and output board is the part which performs the task of delivering the needed power to Jetson and its carrier board. It measures the voltage taken from the interface board and displays it by using an LCD module. Monitoring circuit is monitoring the battery capacity level through voltage measurement. In order to inform the user about the status, optic and sound alarms are implemented. Solution is implemented by simulation using Proteus UDE, testing with experimental test board and implementation on PCB.

VI. CONCLUSION

The need of digitalization in the construction industry motivated this project to be initiated to develop a system on the construction helmets to collect 360 degree images and 3D point cloud data which would be used for recording and inspecting the state of the construction site, doing automatic measurements, generating 3D built-in documentations and creating indoor navigation system. All proposed concepts were implemented, tested and proven to be working as expected. First prototype of the helmet was introduced in the fair "BAU 2019" in Munich, Germany. Prototype was also tested on several construction sites in Germany.

VII. ACKNOWLEDGEMENTS

This research project was financed by ZIM (Zentrales Innovationsprogramm Mittelstand – Central Innovation

Program) funding program of BMWi (Bundesministeriums für Wirtschaft und Energie – Federal Ministry of Economics and Energy) and completed in a collaboration with Open Experience GmbH. We thank our colleagues from Open Experience GmbH who provided insight and expertise that greatly assisted us.

We also would like to show our gratitude to Mr. Deniz Acet for his great and collaborative work in the mechanical design of the construction helmet.

VIII. REFERENCES

- [1] Roland Berge, Digitalisierung der Bauwirtschaft, Available: https://www.rolandberger.com/publications/publication_pdf/roland_berge_digitalisierung_bauwirtschaft_final.pdf.
- [2] H. Tuan; B. Madhukar, Dual Fisheye Stitching for 360 Degree Imaging, 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, 2017.
- [3] K. Rudolf, Reversed Telephoto Lenses: II. The Fish-Eye Lens, A History of the Photographic Lens, 1989.
- [4] A. Peter, A Practical Introduction to Deep Learning with Caffe, Australian Center for Robotic Vision.
- [5] Basler AG, 2D or 3D Camera? Which 3D Camera Technology Fits Your Application? Available: https://www.baslerweb.com/wp-1481127658/media/en/downloads/documents/white_papers/BAS160_8_White_Paper_3D_Technologies.pdf.
- [6] ZED, Meet ZED mini, the world's first camera for mixed reality, Available: <https://www.stereolabs.com/zed-mini/>.
- [7] C. Lakis, 3D Stereo Vision Camera-sensors-systems, *Advancements and Technologies*, 2013.
- [8] R. William, Rotation Matrices, Available: http://www1.udel.edu/biology/rosewc/kaap427627/notes/matrices_rotations.pdf, 2014.
- [9] E. Jan; B. Dorit; N. Andreas, One billion points in the cloud – an octree for efficient processing of 3D laser scans, Jacobs University Bremen GmbH, Automation Group, School of Engineering and Science, Bremen, 2012.
- [10] N. Xiaojuan; L. Fan; T. Ge; W. Yinghui, An efficient outlier removal method for scattered point cloud data, PLoS ONE 13(8): e0201280, <https://doi.org/10.1371/journal.pone.0201280>, 2018.
- [11] C. Yu; X. Guo; A. Zhang; X. Pan, An Improvement Algorithm of Principal Component Analysis, 8th International Conference on Electronic Measurement and Instruments, 2007.
- [12] N. Andrew, An As-Short-As Possible Introduction to the Least Squares, Weighted Least Squares and Moving Least Squares Methods for Scattered Data Approximation and Interpolation, TU Darmstadt, 2004.
- [13] R. Radu Bogdan, Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments, Dissertation, TU München, 2010.
- [14] F. Martin; B. Robert, Random Sample Consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, 1981.
- [15] L. Lin; Y. Fan; Z. Haihong; L. Dalin; T. Lei, An Improved RANSAC for 3D Point Cloud Plane Segmentation Based on Normal Distribution Transformation Cells, Remote Sensing, 2017.
- [16] Y. Michael Ying; F. Wolfgang, Plane Detection in Point Cloud Data, Available: <http://www.ipb.uni-bonn.de/pdfs/Yang2010Plane.pdf>, 2010.

Android Secure Camera Application

K. Ö. DUMAN¹ and H. T. SENCAR²

¹ TOBB University of Economics and Technology, Ankara/Turkey, kemalozgurduman@gmail.com

² TOBB University of Economics and Technology, Ankara/Turkey, htsencar@etu.edu.tr

Abstract - Source Camera Identification is a technique that allows to identify the source camera which is used for taking an image. This technique is mainly used in digital forensics for proving the authenticity of an image which can be used as an evidence in a criminal case. However, it also enables illegal identity tracking and fingerprint copy attacks by using Photo-Response Non-Uniformity Noise (PRNU). Even though there are image source anonymization techniques to avoid these attacks, they are only theoretically available to users. Starting from this point of view, we provide a Secure Camera Application to provide such anonymization out of the box. We demonstrate the availability and anonymization success of the application for different android mobile devices. In this process we also provide better source camera identification method for High Dynamic Range (HDR) images.

Keywords - Source Camera Identification, Image Anonymization, High Dynamic Range.

I. INTRODUCTION

Nowadays, in the all aspects of life, images are used to make the moment persistent. Since all smartphones have the ability of taking photographs, there are enormous image data generated by cell phones or any other camera. In accommodation with social media and data sharing technologies these photographs are uploaded on the internet continuously. Previous studies show that if there are enough photographs to create a Photo Response Non-Uniformity Noise (PRNU) based fingerprint, it is possible to identify source camera of the image successfully [1]. However, it is not that successful when it comes to High Dynamic Range (HDR) photos since the HDR photograph generation requires some extra processes on the image. Apparently, these extra processes affect the original PRNU noise of the image as it does not match with the source camera fingerprint in many cases. In this paper we provide a new implementation that provides better matching results of HDR photos with its source camera.

Source camera identification techniques are very useful in terms of proving authenticity of the image [2] but it also enables illegal identity tracking and fingerprint copy attacks [3]. Based on this information, there are PRNU noise removal techniques called Image Source Anonymization that removes the PRNU noise from the taken image. When the noise is removed it is not possible to extract the noise correctly to identify the source of the image. However, these techniques are only theoretically available and there is no such application which uses the image source anonymization technique to provide PRNU noise-free images. It would be more convenient and secure for users if such a tool existed on the same device which is used for taking

the image. The reason for unavailability of such tool is the fact that it is not straightforward to run such a process on a mobile device since these devices have limited system resources. Android mobile devices have some precautions in order to run the devices smoothly. Such a precaution is killing an application which uses high amount of system resource to prevent a memory shortage which could lead to a system failure. In this paper, we provide an implementation to make it possible to run the image source anonymization process on the mobile devices with the limited system resources along with the fully functional android application that enables users to create anonymized (PRNU noise-free) images.

The organization of the paper is as follows; Section 2 explains how the PRNU based source camera identification and PRNU noise removal techniques work. Section 3 describes what the HDR images are and explains the reason behind the source camera identification problems for HDR images and our method to make it possible. In Section 4, we explain our application and method to run it on the devices with limited system resources. Finally, Section 5 concludes the paper.

II. SOURCE CAMERA IDENTIFICATION AND IMAGE SOURCE ANONYMIZATION

PRNU based source camera identification leverages the imperfection in the imaging sensors that have different reaction to light. This behavior causes each camera sensor to have a specific noise in the taken images. This noise creates a unique fingerprint which can be extracted and used for source camera identification for the images taken from the same camera [1]. Each image includes a specific PRNU noise and actual image data which can be found by using a denoising filter, WDF , as pointed out in [4]. I_e being the noiseless image, this can be expressed as:

$$I_e = WDF(I) \quad (1)$$

Then PRNU noise can be extracted from the image by subtracting the I_e from the original image I which will give the sensor noise, N_I , can be computed as:

$$N_I = I - I_e \quad (2)$$

After extracting the sensor noise estimates from different images, fingerprint F_C of the source camera C created using the maximum likelihood estimator [5]. In order to identify image's source camera by using different fingerprints a measurement of

correlation is required. Such a cross correlation between the fingerprint F_C and the PRNU noise of image, N_I , can be computed using the Peek-To-Correlation Energy (PCE) [6]. Deciding whether an image noise is correlated to fingerprint by examining the PCE value, requires a decision threshold. Decision threshold p , defines the minimum value for positive identification result. Thus, any image that is tested with specific camera's fingerprint is said to be taken from the same camera if the value obtained from the PCE function greater than p . Previous researches show that this threshold value can be set to 50 [7, 8].

Since the PRNU noise can be extracted from the image as described in equation (1) and equation (2), it is also possible to remove the noise from the image. Removing the PRNU noise from the image makes it impossible to identify the source camera since identification method requires PRNU noise to be existed in the image. Without this noise, image becomes anonymous and this is called image source anonymization. This anonymization technique is achieved by subtracting the PRNU noise multiplied by a constant from the image and repeating the process until a PCE value that is lower than the decision threshold is reached as described in [8].

III. SOURCE CAMERA IDENTIFICATION WITH HIGH DYNAMIC RANGE IMAGES

A. HDR Images and Source Camera Identification Problems

High Dynamic Range (HDR) is a technique that produces images which have dynamic range of luminosity. Rather than normal image production, one HDR image requires additional images that are taken in different luminance levels and then combined into one image. Purpose of this is to produce more realistic images close to human eye visual system. However, taking multiple images - usually three - in a row leads to problems depending on the environment which image is taken because even those images taken in a short span of time may end up with the displacement of the objects in the scene. Thus, it requires additional alignment step of the different images before the combination process. There are various scene alignments [9, 10] and de-ghosting methods [11] to solve this problem. Because of the existence of this alignment process, aligning displaced objects can cause the PRNU noise to shift or get lost. This is causing negative result in identification of the HDR image I_{HDR} and Fingerprint F . Since the wide variety of devices has the HDR technology today, a method to make this identification would provide better results in PRNU based source camera identification. In order to achieve this, we provide a new identification implementation to match I_{HDR} to its source F .

B. HDR Images Source Camera Identification Method

This method requires the segmentation of the image noise into m number of matrices, M_{rxc} where r refers to row size and c refers to column size. Result of identification is found by searching matrices in the fingerprint up to a shift threshold until a valid PCE value θ is found. First, a fingerprint consisting of

n number of non-HDR images F_n is created and an HDR image I_{HDR} from the same source is taken. After PRNU noise, N_{HDR} is extracted and segmented into m number of matrices, N_{HDR_i} , $i \in [1, m]$. Instead of trying to match matrices with the same index, for each noise matrix in the list, a list of fingerprints search parts are constructed with the same segmentation algorithm in order to detect any displacement of PRNU noise. For each N_{HDR_i} , multiple parts of F_n is created by shifting the corresponding matrix by one pixel up to a shift threshold ρ . Shift threshold defines the boundaries for the search area to eliminate unnecessary searches. Therefore, we obtain a list of fingerprint parts for each noise part by equation (3).

$$F_{n_i} = [F_{n_{rxc}}, F_{n_{(r+j)xc}}, \dots, F_{n_{(r+p)xc}}, F_{n_{rx(c+k)}}, \dots, F_{n_{rx(c+p)}}] \quad (3)$$

Matching function M with parameters N_{HDR_i} and $F_{n_{i_j}}$ is

$$M(N_{HDR_i}, F_{n_{i_j}}) = \begin{cases} 1 & \text{if } PCE(N_{HDR_i}, F_{n_{i_j}}) > \theta \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

C. Test Results

Test results show that using image segmentation technique to identify HDR image source camera provides better results than direct identification technique. Table 1 shows the match rates acquired with the method with direct matching versus method with image segmentation. Test results are collected with the fingerprint consist of 100 non-HDR images and 50 HDR test images. Matching accuracy is calculated by dividing sum of match results to total number of images. Image segmentation method implementation provides 66% better match results.

Table 1: HDR image identification results

	Direct Matching Technique	Image Segmentation Technique
Match Accuracy	24%	90%

IV. SECURE CAMERA APPLICATION

A. Insecurity of Publicly Available Images

Current image capturing technologies advances on smartphones lead mobile phones to replace cameras. This makes people use smartphones for taking images at any moment of life. Wide spread usage of social media makes people to publish these images on public platforms without any security concern. However, because of the existence of source camera identification techniques these images may be used as source for PRNU noise extraction. This noise data then can be used for creating fingerprint or possibly as a fake PRNU noise in another image that is not taken by the user. Existence of PRNU noise causes an insecurity in publicly available images. To eliminate this insecurity, the PRNU noise should be removed from the images to prevent any possible PRNU based

source camera identification procedure on the images. In order to create secure images from smartphones, a camera application should apply the noise removal procedure to captured or given images. Images without any PRNU noise become anonymous since there is no data that can be used to track the source camera. In this paper, we present a secure camera application for android devices to provide such anonymization out of the box.

B. Secure Camera Application Details

Providing an image anonymization feature for the smartphones would eliminate the security issue we have described in the previous section. However, the image anonymization process requires a huge amount of system resources, especially the memory (RAM). It is not straightforward to run these processes in the smartphones with the limited resources since android devices has a protective feature to kill the processes that use an excessive amount of memory. In order to overcome this issue, we present a new implementation which is using only certain amount of memory without compromising the anonymization success. Since every camera has a different resolution and thus different image sizes, regular techniques of full image anonymization may not work on every mobile device because of the memory problems. Instead of feeding the anonymization and fingerprint creation algorithm with the image with original size, we segment the image into smaller data chunks before these processes. In the first step of image anonymization, fingerprint of the camera will be created, this process requires a certain number of images to load into memory and put into fingerprint creation method which significantly increases the memory usage due to the algorithm's nature. Before this process begins images that are subject to fingerprint are put into image segmentation process that splits the image into n number of parts without changing the resolution. So, let the image I_i be an i^{th} image to be used in the fingerprint creation, segmentation method takes I_i and returns the list of pointers that point to specific regions of the image as $M \times N$ matrix. After all images are put into segmentation process, we create m number of fingerprints where m is the number of parts extracted from the image according to segmentation criteria M and N . Fingerprints are then combined into one single fingerprint with the same size as the original image and saved into internal memory. Since the fingerprints created with the segmentation algorithm to overcome memory problems anonymization process also requires the same procedure. This process is very similar to the previous, but this time requires to load both the fingerprint and the image to anonymize, segmenting both into same $M \times N$ chunks. After, having the both list of fingerprints and image chunks, same indexed chunks are put into image anonymization process. After anonymization of m image chunks are completed original image needs to be combined into one image to be presented to user. *Figure 1* explains how secure camera application works. First, user selects to load or take an image to anonymize, since it is required to have fingerprint to start the anonymization process, application asks user to create a fingerprint first.

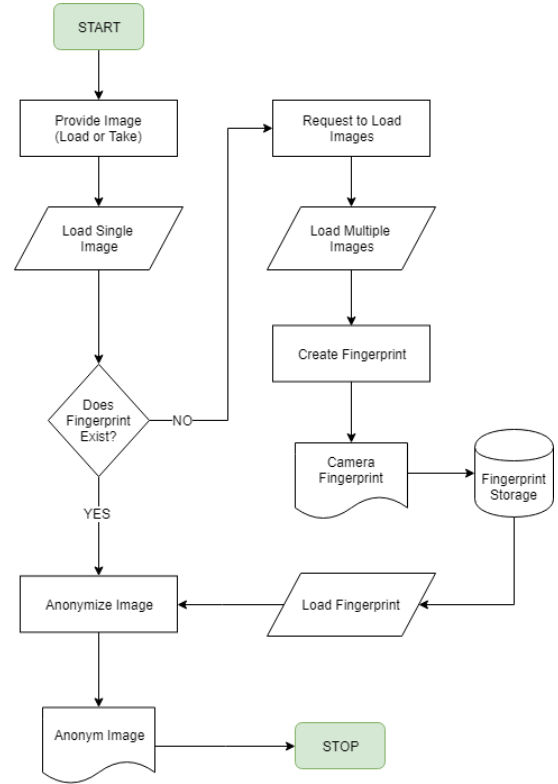


Figure 1: Secure Camera Application flow diagram

Then, user selects multiple images to create a fingerprint and immediately after this process is completed, user is asked to load or take an image to anonymize. Note that consecutive anonymization requests will not repeat fingerprint creation process as it is previously saved into internal memory. However, user can create new fingerprint with new images at any time.

C. Test Results

In order to evaluate average memory usage and anonymization rate of the application we conducted the fingerprint creation and anonymization process on different devices with different specs. To evaluate real world scenarios, we run the application without stopping any other background processes.

Table 2: Fingerprint creation process average memory usage

Device	Resolution	Avg. Memory Usage
Xiomi MI 5	3456 x 4608	680 MB
Samsung Note 4	5312 x 2988	590 MB
Samsung S8	4032 x 3024	650 MB
Samsung Note 5	5312 x 2988	580 MB
Samsung J7 Pro	2322 x 4128	480 MB

Memory usage taken from the Android Profiling Tool of the fingerprint creation and anonymization process in different smartphones are given in *Table 2* and *Table 3*. During fingerprint creation process the application memory consumption varies between 480 MB and 680 MB. Since the

test devices cover the top and average system specifications in the market, we conclude that the application can work on any device that has at least 700 MB available memory for the given image resolutions.

Table 3: Anonymization process average memory usage

Device	Average memory usage
Xiomi MI 5	450 MB
Samsung Note 4	490 MB
Samsung S8	425 MB
Samsung Note 5	485 MB
Samsung J7 Pro	285 MB

For anonymization process a maximum of 490 MB memory is used. Note that, anonymization process requires less memory than fingerprint creation since this process loads only one image with existing fingerprint into memory. Therefore, we conclude that 500 MB of available memory is enough to complete this process for the given image resolutions.

In order to evaluate anonymization success of the application an anonymization rate is calculated for each smartphone by dividing the total number of images that are successfully anonymized to total number of images. For each device under testing we collected 100 images different than the anonymizing fingerprint and 20 anonymized images. Clearly, images anonymized by the application are completely anonymized against the fingerprint that is created by 100 images.

V. CONCLUSION

In this paper, we provide secure camera application for android smartphones that allows users to create secure images that do not contain any PRNU noise which can be used for illegal identity tracking or fingerprint copy attacks. Our proposed implementation technique makes it possible to run such a secure camera application on smartphones which have limited system resources. Test results show that we provide complete anonymization against the fingerprints constructed with 100 images. It has a critical importance to have such a secure camera application in smartphones, especially in an era where a wide variety of people use online image sharing platforms extensively and continuously. As future work, in order to provide better user experience, we plan to improve the application by adding automatic fingerprint creation and anonymization process as a hidden background service.

REFERENCES

- [1] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205-214, June 2006.
- [2] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Determining Image Origin and Integrity Using Sensor Noise," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 74-90, March 2008.
- [3] M. Steinebach, H. Liu, P. Fan, and S. Katzenbeisser, "Cell phone camera ballistics: attacks and countermeasures," in *Proc. of SPIE: Multimedia on Mobile Devices*, vol. 7542, pp. 0B-0C, 2010.

- [4] J. Lukáš, J. Fridrich, and M. Goljan, "Detecting digital image forgeries using sensor pattern noise," in *Security, Steganography, and Watermarking of Multimedia Contents VIII*, 2006.
- [5] M. Chen, J. Fridrich, and M. Goljan, "Digital imaging sensor identification (further study)," in *Security, Steganography, and Watermarking of Multimedia Contents IX*, 2007.
- [6] M. Goljan, "Digital Camera Identification from Images – Estimating False Acceptance Probability," in *Digital Watermarking*, Springer Berlin Heidelberg, pp. 454-468, 2009.
- [7] M. Goljan, J. Fridrich, and T. Filler, "Large scale test of sensor fingerprint camera identification," in *Media Forensics and Security*, 2009.
- [8] A. E. Dirik and A. Karaküçük, "Forensic use of photo response non-uniformity of imaging sensors and a counter method," *Optics Express*, vol. 22, no. 1, p. 470, Jan. 2014.
- [9] G. Ward, "Fast, Robust Image Registration for Compositing High Dynamic Range Photographs from Hand-Held Exposures," *Journal of Graphics Tools*, vol. 8, no. 2, pp. 17-30, Jan. 2003.
- [10] A. Tomaszewska and R. Mantiuk, "Image Registration for Multi-exposure High Dynamic Range Image Acquisition," in *WSCG*, 2007.
- [11] M. Granados, K. I. Kim, J. Tompkin, and C. Theobalt, "Automatic noise modeling for ghost-free HDR reconstruction," *ACM Transactions on Graphics*, vol. 32, no. 6, pp. 1-10, Nov. 2013.

A Recommendation System for Seattle Public Library Using Naïve Bayes Classifier

A.KARAMANLIOĞLU¹, A. ÇETİNKAYA², A.DALKIRAN², M. KOÇA³, and
A.KARAMANLIOĞLU²

¹Trakya University, Edirne/Turkey, akaramanlioglu@trakya.edu.tr

²Middle East Technical University, Ankara/Turkey, cetinkaya@ceng.metu.edu.tr

²Middle East Technical University, Ankara/Turkey, dalkiran@ceng.metu.edu.tr

³Istanbul Technical University, Istanbul/Turkey, koca19@itu.edu.tr

²Middle East Technical University, Ankara/Turkey, alperk@ceng.metu.edu.tr

Abstract— In this study, we intended to recommend possible books to be checked out from Seattle Public Library (SPL) within a month. While it seems possible to make daily estimates from existing data, it will be more useful to make monthly data forecasts since SPL provides fresh data every month. The information obtained here will contribute to logistic modelling. Since some assumptions on the location of the books and storage capacity of each location are required for the better management of the resources, this problem can be considered as a warehouse resource allocation problem. In order to perform the required predictions, Naïve Bayes (NB) algorithm is applied on the Seattle Public Library Dataset (SPLD), which contains all of the checkout records between 2005 and 2017 in the SPL.

Keywords - Big Data, Warehouse Resource Management, Naïve Bayes, Book Recommendation, Recommender Systems

I. INTRODUCTION

The warehousing problem is an important problem that can be encountered in many areas. Failure to properly manage warehouse allocations can lead to many problems. These problems may vary depending on the area: In some areas, these problems can be overtaken, while other area resources cannot be properly managed, resulting in very serious consequences.

Data mining can be defined as the whole process of determining models, solving problems related to data analysis, and forecasting. Data mining tools enable businesses in various domains to predict future trends in different subjects. In order to do this, several methods have been designed and developed. These methods have many advantages and disadvantages compared to each other, and success rates may vary depending on the applied datasets. While some methods may be successful in some datasets, another method may be able to yield better results in others.

In this study, it is intended to recommend the books to be checked out from the Seattle Public Library (SPL) within a month. Although it is possible to make daily estimates from the current data, it makes more sense to make monthly data estimates because SPL provides up-to-date data every month.

The information obtained here is thought to be useful in modeling logistics processes. This problem can be considered as a warehouse resource allocation problem, as there are some assumptions about the location of the books and the storage capacity of each location are needed for better resource management. Recommendations are carried out by applying Naïve Bayes (NB) algorithm on The Seattle Public Library Data Set (SPLD), which includes all transactions in SPL between 2005 and 2017. The success rate of the method was evaluated according to the accuracy performance measure.

II. LITERATURE REVIEW

A detailed comparison of the related studies in terms of the dataset name, size of the dataset, algorithms and purpose is presented in Table 1.

In Karno et al [1], loan registration records between 2005 and 2010 from the University of Toronto Mississauga Library were used. The aim of the study was to propose a tool to assist the management of the library. For this purpose, the bibliomining for the library data warehouse has been implemented. The study conducted over a dataset that includes around 1 million checkout records.

Renaud et al [2] used Library of Congress (LC) Classification over more than 6.2 million transactions from the university library data to find a correlation between library use and student achievement.

In the study performed by Rajagopal et al. [3], patron and circulation data from the library of University of Hong Kong were used to design a model for a recommendation system by applying K-means clustering. These algorithms provided to make book estimates by setting cluster centers with subject headers. Before K-means was implemented, the data were cleaned, transformed and reconciled in a specially designed data warehouse, which is essentially a collection of small Data Marts, by systematically processing data relating to the storage of relevant data in their operational databases.

In the study performed by Lina and Zhiyong [4], Association Rule Mining was used along with K-means clustering in order

to provide book recommendation for students based on the data acquired from the Liao Ning Technical University library.

Chen et al. in their study [5] examined user behaviour in a university library's large-scale book-loan records, then applied two suggestion algorithms: Collaborative filtering and probability-based. They have shown that the test results of the probability-based algorithm are better.

Silverstein and Shieber [6] conducted a study on data taken from Harvard College Library and were aimed to determine the low-demand items and to transport them from shelves to a remote storage. This experiment was performed on 6 million instances of transaction data and decision trees were used to train the model.

In the study conducted by Wagstaff and Liu [7], six different machine learning algorithms were used to classify weeding candidates. These algorithms were applied on a large-scale dataset containing library data from 2011 to 2014 at Wesleyan University. Each weeding candidate was represented with the information of the number of years between publication year and 2012, number of U.S. libraries and peer libraries, number of checkouts, number of Wesleyan faculty member's and librarian's votes. The decision was defined as a Boolean variable as "Keep" or "Weed". The highest precision was obtained with K-Nearest Neighbours algorithm. The study showed that NB and Linear Support Vector Machine (Linear SVM) algorithms provide the best recall value.

In the study Whitney et al. [8], two sets of data acquired from the University of Los Angeles Library were used with nearly 9.3 million total transactions. The data were processed with content based, circulation based and FRBR methods in order to provide individual recommendations to library users based on their previous loan histories.

III. DATASET

A. Seattle Public Library Dataset

Seattle Public Library Dataset (SPLD) [9] includes a log of all physical item checkouts that took place in Seattle Public Library. The dataset begins with checkouts occurring in April 2005, and was regularly updated until the end of 2016. This dataset does not contain data for the renewals. It includes more than 90 Million instances of transaction records.

The dataset contains several types of files including raw data that contains checkout records, a data dictionary allows you to decode the "ItemType" column from the checkout records and library collection inventory that is a dataset in its own right and stores important metadata about each title, such as the author names and subjects.

The fields included in the dataset and their respected meanings are given in Table 2.

B. Preprocessing

Since the number of instances included in the dataset is over the 90 million and the total size of the data is almost 7 Gigabytes, we had to perform data mining techniques and applied some preprocessing methods to the data in order to surpass the complexities caused by the data size. Also, because of the physical restrictions, we couldn't be able to process the data as a whole. So, we have performed the required tasks on

the preprocessed data that are distributed into 13 pieces for every year between 2005 and 2017, respectively. Then for each of the yearly distributed data, "ItemBarcode" field is deleted since every transaction id is unique and this field wouldn't be useful for the classification purposes. Distribution of the data over the years is given in Figure 1.

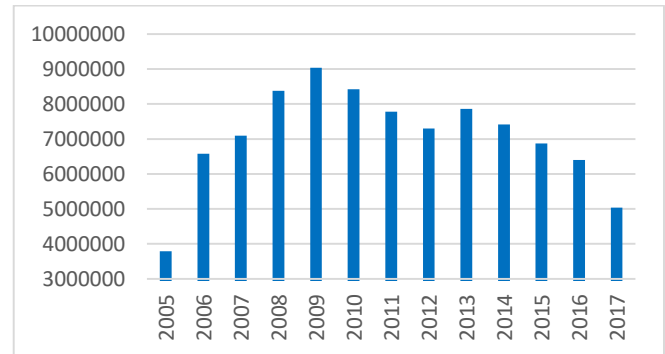


Figure 1: Distribution of the data instances over the years.

Then we have converted the string type instances to integer type for all unique items. All unique items are calculated with a Java program in order to find the occurrence of each item from all of the transactions that are held over the years.

Finally, we have performed sampling based on the frequency of the items and removed the items which have less than 0.0001 frequency rate (The items that have less than 9000 instances) are eliminated from the dataset. This process helped us to get rid of the items that are rarely rented. Thus, classification accuracies are greatly improved after this process since the number of items to be predicted is reduced. In the end, we applied a classification algorithm on the data that includes 1.5 Million instances for 153 unique items.

IV. METHOD

The NB classification is a technique that is based on Bayes theorem and has the assumption of independence among the predictors [10]. A NB classifier basically assumes that the presence of a particular property in a class is inconsequential to the presence of other properties. While these properties are dependent on each other or the presence of other properties, all these properties contribute independently of the probability.

The NB is a supervised learning technique and allows a multivariate distribution to be reduced to multiple univariate distributions. All of the features are considered to be equally important. In this technique, the classes are predetermined. It is also prior information that the sample data belong to which classes. According to these classes, the conditional probability calculation is performed.

The construction of the NB model is easy and is particularly useful for large data sets. The disadvantage of this algorithm is that it depends on the assumption of independence and perform badly if this assumption is not met. Spam filtering and document classification are the most common application domains of this method. In many areas, the results of the method can compete with methods such as Artificial Neural Networks or Decision Trees.

Table 1: Comparison of the related studies.

First Author-Year	Used Dataset	Size of the Dataset	Algorithm(s)	Purpose
Karno-2012 [1]	Library of Universiti Teknologi Malaysia	1 million transactions	Cluster analysis Data mining	Book recommendation based on previous borrowing data
Renaud-2015 [2]	UCI Library	6.2 million transactions	LC Classification	Correlation between library use and student achievement
Rajagopal-2012 [3]	University of Hong Kong Libraries	NA	K-means clustering	Book recommendation based on previous borrowing data
Lina-2013 [4]	Library of Liao Ning Technical University	62500 transactions	K-means clustering and association rules mining	Book recommendation based on previous borrowing data
Chen-2012 [5]	Library of UIBE (University of International Business and Economics)	571815 transactions	Collaborative filtering and probability-based algorithms	Book recommendation based on previous borrowing data
Silverstein-1996 [6]	Harvard College Library	6 million transactions	Decision trees	Predict books to eliminate from off-site storage
Wagstaff-2017 [7]	Wesleyan University library data	80346 transactions	6 different algorithms applied (k -NN, NB, Decision Tree, Random Forest, Linear Kernel SVM, Gaussian)	Human weeding decisions and predictions
Whitney-2006 [8]	University of California Library	9.3 Million transactions	Kernel SVM Content-based, circulation based and FRBR methods	Book recommendation based on previous borrowing data

Table 2: Fields included in the SPLD.

Field	Data	Type
BibNumber	Unique Item ID for the Items	Integer
ItemBarcode	Unique transaction ID	String
ItemType	Type of the Item	String
Collection	Collection that the Item belongs	String
CallNumber	Location of the book in the library.	String
CheckoutDateTime	Transaction date and time in dd/mm/yyyy hour AM/PM format	String

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this study, we have applied NB to predict the books from SPLD that includes more than 1.5 Million transaction records of 153 unique items. The input data was read from a CSV file. Since the scale of the data is quite large, a big data interpreter is needed. For this purpose, the Spark Python API (PySpark) was used. Thus, NB classification algorithm could be implemented via Python programming language using the big data related features of Spark. Dataset is divided into training (80%) and test (20%) sets. NB algorithm was applied to the training dataset and the trained model was extracted. Then, on the test set, we have predicted “BibNumber”, which is a unique attribute for all of the items. Different settings were applied for the training phase and the best parameters for smoothing and model type were determined. The best results were obtained when the value of the lambda smoothing parameter was 1,0. In addition, selecting the model type parameter as multinomial affected the results positively. Using these parameters, we have accomplished 71.28% accuracy. Since it is challenging to predict the correct label among 153 classes, the classification accuracy seems promising.

From Table 1, the studies given in [1, 3, 4, 5, 8] provide book recommendation based on the user’s previous borrowing data. Also, in [6] unused items have been tried to be predicted in order to manage the shelves on the library. K-means clustering, Association Rule Mining, Collaborative filtering, probability-based algorithms, Content-based, circulation based and FRBR methods have been tested on checkout records for similar intentions with our study. Whereas the characteristics of the used datasets and used attributes are different, they may not be accurately comparable. The literature study also supports that more studies are needed to obtain more useful conclusions about the prediction approaches.

Prediction of future trends is very important for a business in various domains in order for them to provide better services. Library management is one of the most important domains in this aspect. Providing useful recommendations to users can improve their customers’ experiences. It will most certainly help the libraries to provide more efficiency in their operations thus cutting their expenses. The findings of this study can be used to conduct further research on the estimation approaches for library data.

VI. FUTURE WORK

We aim to make recommendations with various algorithms on SPLD and evaluate the results. Conceptual design that is planned to be realized for the proposed system is illustrated in Figure 2. According to this design, preprocessed data will be stored in HDFS file system since the scale of the data is large and usage of a distributed data system would be useful for our purposes. In order to make desired predictions, various machine learning approaches are planned to be implemented on Spark MLlib. HDFS file system will be linked to Spark MLlib by

utilizing YARN. Random Forests, K-Nearest Neighbours and NB approaches are planned to be applied for the classification.

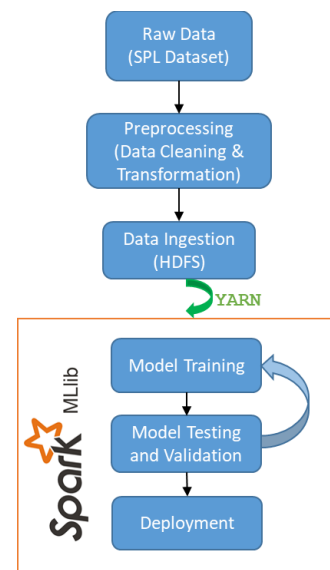


Figure 2: Conceptual Design to be realized.

Based on the success rates, training methods can be changed and the training phase will be repeated accordingly. In the end, we will be selecting the most successful approaches to finalize the predictions.

REFERENCES

- [1] R. Karno, S. A. Noordin, S. A. Rahman, and A. Talib, “Facilitating resource allocation decision through data mining: The case of UTM library,” in *18th IBIMA Conference on Innovation and Sustainable Competitive Advantage: From Regional Development to World Economies*, January 2012.
- [2] J. Renaud, S. Britton, D. Wang, and M. Ogihara. (2015). Mining library and university data to understand library use patterns. *The Electronic Library*. 33(3). pp. 355—372. Available: <https://www.emeraldinsight.com/doi/pdfplus/10.1108/EL-07-2013-0136>
- [3] S. Rajagopal, and A. C. M. Kwan. (2012). Book Recommendation System using Data Mining for the University of Hong Kong Libraries. *Information, Technology and Educational Change*.
- [4] J. Lina, and M. Zhiyong. (2013). The Application of Book Intelligent Recommendation Based on the Association Rule Mining of Clementine. *Journal of Software Engineering and Applications*. 6(7B). pp. 30—33. Available: <http://dx.doi.org/10.4236/jsea.2013.67B006>
- [5] C. Chen, L. Zhang, H. Qiao, S. Wang, Y. Liu, and X. Qiu, “Book Recommendation Based on Book-Loan Logs,” in *International Conference on Asian Digital Libraries*, pp. 269-278, November 2012.
- [6] C. Silverstein, and S. M. Shieber. (1996). Predicting individual book use for off-site storage using decision trees. *The Library Quarterly*. 66(3). pp. 266—293.
- [7] K. L. Wagstaff, “Automated Classification to Improve the Efficiency of Weeding Library Collections,” Doctoral dissertation, San Jose State University, 2017.
- [8] C. Whitney, and L. R. Schiff. (2006). The Melvyl recommender project: Developing library recommendation services.
- [9] Seattle Public Library Dataset. Available: <https://www.kaggle.com/seattle-public-library/seattle-library-checkout-records>
- [10] Y. D. Rubinstein, and T. Hastie, “Discriminative vs Informative Learning,” in *KDD*, vol. 5, pp. 49-53, August 1997.

Server Based Indoor Location and Navigation Using Beacon Devices

A. BAYAR¹ and H. KUTUCU²

¹ Karabuk University, Department of Computer Engineering, Karabuk/Turkey, adembyar@gmail.com

²Karabuk University, Department of Computer Engineering, Karabuk/Turkey, hakankutucu@karabuk.edu.tr

Abstract- Positioning indoors / outdoors has gained importance in recent years for the architecture, engineering, construction and marketing industries. Because global positioning system (GPS) data is not available indoors. Identification of user presence in the building, pursuit persons and goods in factories, direction of people in emergency situations, description of how to go to a pavilion in big fairs, direction setting according to location at which to enter from the airport, direction to the mall or offering a product promotion and discount opportunity in order to increase sales according to the location of the person in the shopping center is a few of the places where indoor location determination can be used. In this project, using the Beacon devices in the indoor space, an application will be shown instantaneously by detecting the position of the person within the target area with high sensitivity and performing the navigation functions afterwards. In this work, it is aimed to detect user location with high accuracy and to enable real-time monitoring of location by developing an android application that can work on mobile platforms. In addition, a web application that works simultaneously with the mobile application will be developed and used for remote user monitoring, routing and early detection in emergency response.

Keywords – Indoor Positioning, BLE, Beacon, Navigation.

INTRODUCTION

NOWADAYS, people spend most of their time in enclosed spaces. In modern urbanization, these enclosed spaces are built in a complex way to resemble the streets and streets of a city. In particular, people are lost in hospitals, airports and big shopping malls. Increasing human needs, an intense and fast-paced lifestyle have increased the importance of time management which is one of the most important problems of daily life.

The most important of the location estimation applications in the outdoor environment is undoubtedly the global positioning system (GPS). It is possible to estimate the position mathematically by GPS. This system, which is used in many fields such as transportation, geographic information systems, security, product tracking and navigation, is not able to show its performance in indoor environment with high precision. There are a lot of methods developed for locating in the indoor environment. Some of them are Rfid, Wi-Fi, BLE, UWB, Ultrasounds and Zigbee [1,2].

Radio frequency identification (RFID) tags include battery and have the capability of powering the circuit. By this way, it

is possible to send high level signals as a response of low level reader signals. Received Signal Strength Indicator (RSSI) is the measure of distance in this technology. RFID provides wireless transmission via radio signals in a single direction using a label to reach location information of people or objects. There are various problems about radio wave propagation in indoor environments like absorption, diffraction, path loss and reflection [3-6].

Bluetooth estimates the position in similar ways with Wi-Fi, but it uses different frequency band (2.4 Ghz). Nowadays most of the mobile phones have Bluetooth sensors. For better accuracy reference points need to be installed. Bluetooth technology has better resolution than Wi-Fi [7].

Ultra wideband (UWB) technology determine positions of objects or person using Radio frequency (RF) signals. This technology is based on short-term pulse and uses broad RF band to transmit data. The system takes its power from high time resolution and short wavelength. In this way, high precision performance under 20 cm can be achieved in indoor positioning systems. [2,8,10].

Positioning systems with ZigBee technology work at 2.4Ghz frequency band and based mainly on RSSI measurements. Angelis et al. proposed that average absolute error can be under 3m with dynamic calibration [11]. Since RSSI value can vary considerably with environmental factors, it is better to use for home and industrial automation, smart monitoring applications because of the power consumption rate, network size and network architecture.

In the ultrasonic sound waves method, the signal waves sent by a transmitter and collected by a receiver node. Time-of-Flight (TOF) data is used for distance estimation. Measurement with ultrasonic waves can achieve millimeter accuracy with a range of 6m. But, undesirable phenomena can affect ultrasonic waves such as crosstalk, echo and jamming [12, 13].

Using different radio frequencies in indoor area positioning technologies, methods such as Wi-Fi, ZigBee and UWB can be applied. ZigBee technology is useful for wireless control and monitoring solutions without requiring extensive infrastructure. It is possible to use ZigBee technology in positioning systems that do not require high precision accuracy. Wi-Fi technology provides longer communication distance than ZigBee and

UWB technology. Complexity and energy consumption rates of Wi-Fi are much higher than that of ZigBee and UWB. However, both Wi-Fi and ZigBee technology have several meters accuracy when they applied for positioning. Compared to ZigBee and Wi-Fi technology, UWB offers significant advantages with respect to robustness, energy consumption and location accuracy [14-16].

In this study, it is aimed to determine the users' instantaneous position with Beacon devices based on the Bluetooth Low Energy (BLE) technology in a closed environment by applying the Kalman filter to the measured RSSI to remove noise in signals. The signals obtained as a result of the Kalman filter will be used to find position by the trilateration method.

When a person enters the selected indoor environment, the current location of him/her and rooms, corridor, office, etc. of the environment will be seen on the map that is loaded on a mobile application. Furthermore, the person is navigated by selecting the target point on the map. For navigation, the shortest route will be found and the person will be directed to that point.

INDOOR NAVIGATION SPECIFICS

A. Ble and Beacon Devices

The BLE technology, which operates at 2.4 GHz frequency band such as the Bluetooth technology, uses the Frequency Hopping Wide Spectrum technique and performs a communication with 1 Mbit/s data transmission rate. Compared to other Bluetooth versions in terms of power consumption, BLE is energy efficient. They are easy and practical to use in many fields such as home, logistics, security and personalized entertainment [17].

Beacon devices with increased use after BLE 4.0 technology can also be used for some other purposes such as measuring temperature, acceleration, etc. by customizing with various equipment. In general, it has a simple hardware card rather than a complex structure [16].

Beacon devices emit a radio signal containing hexadecimal numbers. In this signal, there are Universally Unique Identifier (UUID), major and minor values belongs to the beacon device. The power of the incoming signal is measured in dBm and various operations can be performed with the help of this signal. They are used for several purposes by emitting signals with a specific time interval. These signals emitted by beacons are detected by mobile phones. A software on mobile phones is triggered by these signals. In other words, an application on mobile phones starts to work as soon as the phone enters the range of beacon markers. This interaction allows to prepare customized applications for users [18].

METHOD

A. Detecting and removing noise

There are a number of factors that affect the RSSI signal in indoor positioning technologies. These are deviations in radio signal, signal reflective and absorbing factors (indoor traffic, floor plan, building construction material), direction of mobile device and beacon, temperature, damp. For an effective system design, it is necessary to work with multiple measuring sets.

The Kalman filter [19] offers an efficient and optimal solution for positioning and tracking where the system has a normal distribution and linear error of measurement. Generally, this filter can be applied to systems modeled in real life without linear structure. [20]. The Kalman filter tries to guess the status of the system with input and output of the previous steps. The Kalman filter is very powerful and capable to predict the system's non-measurable condition despite the filtering feature as in traditional estimators.

The algorithm works by recursive real-time on noisy data, filtering the errors by fitting them to the least-squares curve and optimizes the mathematical estimation of the future situation produced by modeling the physical characteristics of the system. The RSSI signals from the beacons devices can be cleaned with high rate by eliminating the environment noise by applying the Kalman filter.

We applied the Kalman filter to the RSSI signals which measured from different points given in Table 1. These values are measured from 1 meter away from the mobile phone. In our tests, we used Samsung Galaxy S8. RSSI signal values taken from the beacon devices placed in a room and a corridor are given in Table 1 and these values are processed by the Kalman filter.

Table 1: Application of Kalman Filter

Measures	In the room		In the corridor	
	Original RSSI	After The Kalman Filter	Original RSSI	After The Kalman Filter
1	-81	-83	-86	-85
2	-83	-82	-81	-81
3	-85	-83	-83	-81
4	-76	-81	-90	-82
5	-75	-82	-91	-83
6	-83	-82	-80	-81
7	-84	-83	-84	-81
8	-82	-82	-89	-83
9	-76	-83	-86	-82
10	-79	-81	-81	-80

RSSI signals are easily affected by factors such as walls, furniture, Wi-Fi signals. These factors can be listed as deviation and absorption from objects in the environment, the direction of the beacon device. Sometimes it can be reflected signals even from the source of RSSI and reduces the accuracy of the system. All factors in the environment must be considered in order to capture high sensitivity and instant position information in indoor positioning. Sometimes it may be necessary to perform an average value calculation on multiple data sets.

Two places were selected for testing. The first one is the corridor of a house and the other is the engineering floor of Karabuk University. The maps of the environments were scaled and transferred to the virtual environment. Beacon devices were placed in the environment so that the user could enter the coverage area of at least three of them and their positions are defined as x, y coordinates.

The flow chart of the system is given in Figure 1.

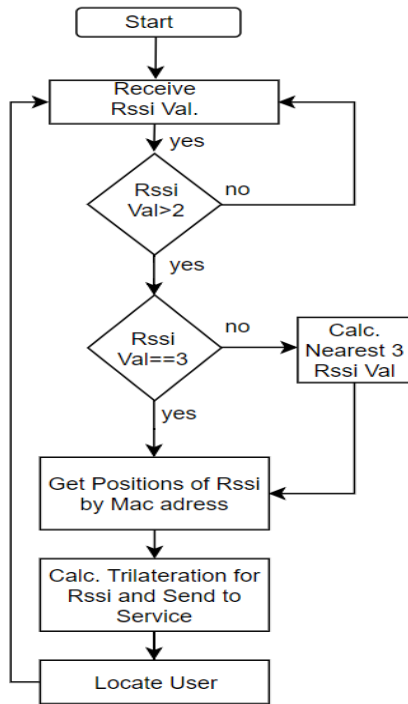


Figure 1: Flow Chart of the system.

B. Application of Trilateration Method

This method, known as distance measurement, is an algorithm that calculates the position using coordinates obtained from known nodes. Calculated distance is the intersection point of circles formed by distances between reference nodes. If the unknown point is (x, y), known points are (x_i, y_i) and d_i represents the radius of the circle. Therefore, we have three sets of circle equations with known radius parameter and anchor locations (Eq. 1,2,3).

In Figure 2, (V₁, V₂, V₃) are reference nodes and d₁, d₂, d₃ are distances from the target points to the reference nodes. The intersection area of the d₁, d₂, d₃ radiused circles gives the position information of the target point [16].

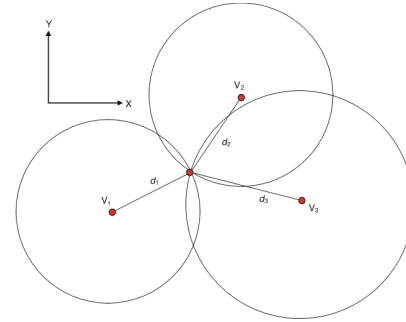


Figure 2: Trilateration Method

General formula of the trilateration method.

$$d_1^2 = (x-x_1)^2 + (y-y_1)^2 \quad (1)$$

$$d_2^2 = (x-x_2)^2 + (y-y_2)^2 \quad (2)$$

$$d_3^2 = (x-x_3)^2 + (y-y_3)^2 \quad (3)$$

x₁, x₂, x₃, y₁, y₂, y₃ coordinates of reference nodes and (x, y) are unknown positions of target points. The solution of these equations gives points of circles intersection providing an area of indoor localization (Fig. 2).

The RSSI distance conversion formula is calculated by the following equation (4).

$$d = 10^{\frac{TxPower - RSSI}{10n}} \quad (4)$$

The TxPower value is the average RSSI value that is calculated at a 1 meter distance from the beacon devices. n is expressed as the noise ratio in the environment. As a result of the tests, the value n for the selected corridor was taken as 2.63 and this value can be changed dynamically in practice.

At the beginning of the application, the system starts to read RSSI values from beacon devices in the environment. As soon as at least 3 beacon devices are within range the positions of beacon devices which defined in the system are found by mac address in the RSSI signal. These positions are given as input with the RSSI values to the trilateration algorithm. Calculated x and y position information marks user position on mobile application and sent web application over the C# Rest Services.

Table 2 : Position calculation with RSSI signals from beacon devices at different points.

	Beac. A	Beac. B	Beac. C	Pos(x,y)
1	-77	-76	-74	169,266
2	-78	-77	-72	146,315
3	-85	-80	-73	240,302
4	-83	-72	-67	123,214
5	-91	-75	-72	107,221

Figure 3 shows the results measured in the corridor. In addition to the instantaneous position information, the user's previous positions are also temporarily stored in database and used for position accuracy.

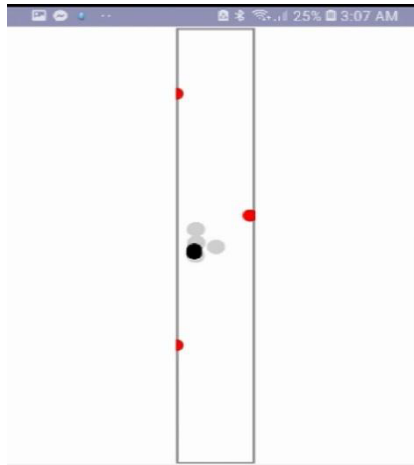


Figure 3: Test in the corridor

The red dots in Figure 3 show the positions of the beacon devices placed in the corridor. The black point is the position of the user in the corridor and the gray points are the previous positions of the user.

As soon as the user enters the environment, the application starts to scan the signals from the beacons. The application calculates the position by taking the nearest three of the beacon signals that are broadcasting 10 times per second. The calculated position is compared with the previous position of the user. A deviation from the environment noise is checked and the difference is calculated separately over x and y . If the difference is more than 1 meter, there is a possible deviation and in such a case, a 50-60 cm value is added to the positions. During this check-out time, new operation is repeated. In this way, the user position is calculated with a minimum difference.

In addition to the mobile application, position of the user is also displayed instantly in a web (Fig. 4) application and the user can be monitored in the environment. This method aims to provide monitoring and safety control in a factory or in a dangerous work environment.

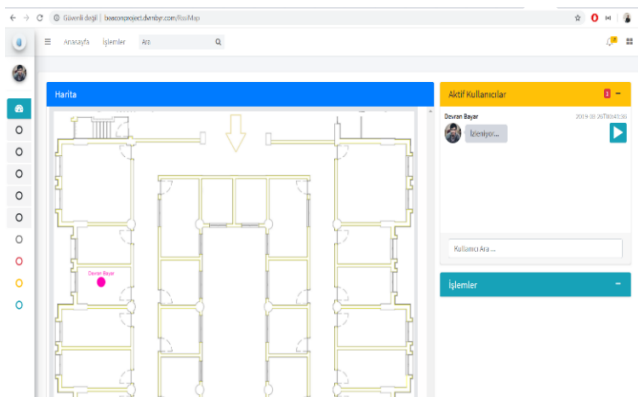


Figure 4: Monitoring of active users on the web application.

In addition, position of users stored in database and provide

simulate and monitoring. An application that can be used in many areas such as security, tracking was designed.

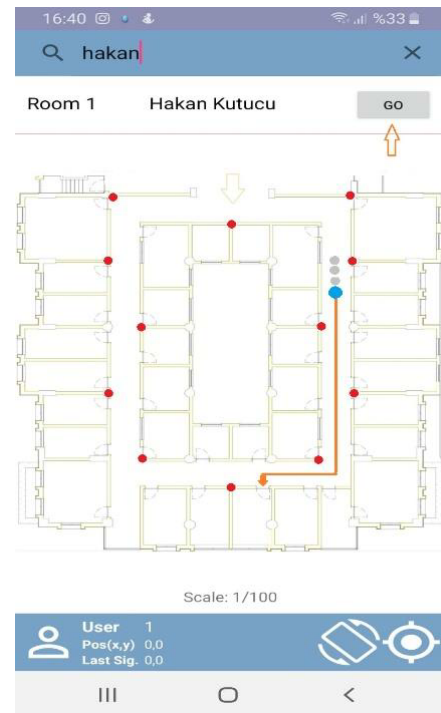


Figure 5: Mobile application interface

Figure 5 shows the user positioning interface of the mobile application. The red dots are beacon devices placed in the corridor and placed in such a way that at least 3 of the users can enter the coverage area. After the user enters the selected area, the user position is shown on the map in the application. Then the user can get navigation service wherever he wants.

CONCLUSION

In this study, a mobile application has been developed with the effective algorithms to minimize the signals noise with beacon devices placed in a rectangular environment. In the environment, noise sources such as walls and Wi-Fi, which cause RSSI signals corruption, were filtered and the location was determined with minimum error and user location successfully detected and simulated.

In addition, we designed a web application for monitoring user instantaneously within a selected environment. With web application it has become an alternative to many camera-based applications where application costs are high. Cost of the whole system is relatively cheaper than other technologies.

REFERENCES

- [1] K. Benkic, M. Malajner, P. Planinšic, Ž. Cucej, "Using RSSI value for distance estimation in Wireless sensor

- networks based on ZigBee”, in Proc. IWSSIP Conf., Jun. 2008, pp. 303-306.
- [2] Y. Gu, A. Lo and I. Niemegeers, "A Survey of Indoor Positioning Systems for Wireless Personal Networks," *IEEE Commun. Surveys Tutorials*, vol. 11, no. 1, pp. 13 - 32, January 2009
 - [3] Bouet, M. and Dos Santos, A. L., "RFID tags: Positioning principles and localization techniques", 2008 1st IFIP Wireless Days, WD 2008, (2008).
 - [4] Toplan, E, Ersoy, C., 2012, "Yaşlı Takibi Amaçlı Rfid Tabanlı Bina İçi Konum Belirleme", 20th Signal Processing and Communications Applications Conference (SIU)", 20, S.1-5
 - [5] He, S. and Chan, S. H. G., "Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons", *IEEE Communications Surveys And Tutorials*, 18 (1): 466–490 (2016).
 - [6] Kaemarungsi, K., Krishnamurthy, P., 2004., "Modeling of indoor positioning systems based on location fingerprinting" Twentythird Annual Joint Conf. of the IEEE Computer and Comm. Societies, 2, S. 1012–1022
 - [7] Neburka, J., Tlamsa, Z., Benes, V., Polak, L., Kaller, O., Bolecek, L., Sebesta, J., and Kratochvil, T., "Study of the performance of RSSI based bluetooth smart indoor positioning", 2016 26th International Conference Radioelektronika, Radioelektronika 2016, 121–125 (2016).
 - [8] Quyum, A., "Guidelines for Indoor Positioning", M.S. Thesis, Lulea University of Technology, Lulea, Sweden, (2013).
 - [9] U. Hatthasin, S. Thainimit, K. Vibhatavanij, N. Premasathian and D. Worasawate, "An Improvement of an RFID Indoor Positioning System using One Base Station", *Proceedings of the 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information*
 - [10] Larranaga, J., Muguira, L., Lopez-Garde, J. M., and Vazquez, J. I., "An environment adaptive ZigBee-based indoor positioning algorithm", 2010 International Conference On Indoor Positioning And Indoor Navigation, IPIN 2010 - Conference Proceedings, (September): 15–17 (2010).
 - [11] De Angelis, G., De Angelis, A., Moschitta, A., and Carbone, P., "Ultrasound based positioning using Time of Flight measurements and crosstalk mitigation", *Conference Record - IEEE Instrumentation And Measurement Technology Conference*, 2015–July 1865–1870 (2015).
 - [12] Medina, C., Segura, J. C., and De la Torre, Á., "Ultrasound indoor positioning system based on a low-power wireless sensor network providing subcentimeter accuracy", *Sensors (Switzerland)*, 13 (3): 3501–3526 (2013)
 - [13] Zhang, J., Sahinoglu, Z., Kinney, P., "UWB Systems for Wireless Sensor Networks", *Proceedings of the IEEE*, 313-331 (2009).
 - [14] Chawathe, S., S., 2008. Beacon placement for indoor localization using bluetooth. 11th International IEEE Conference on Intelligent Transportation Systems, Beijing, pp. 980-985.
 - [15] O. Oguejiofor, V. Okorogu, A. Adewale, B. Osuesu, "Outdoor Localization System Using RSSI Measurement of Wireless Sensor Network", *International Journal of Innovative Technology and Exploring Engineering*, vol. 2, Jan.2013, pp. 1-6.
 - [16] M. Köhne, J. Sieck, "Location-based Services with iBeacon Technology", *Second International Conference on Artificial Intelligence, Modelling and Simulation*, Madrid, Spain, 315-321, 18-20 November 2014
 - [17] Mori, T., Kajioka, S., Uchiya, T., Takumi, I., Matsuo, H., 2015., "Experiments of position estimation by BLE beacons on actual situations, " in Proc. IEEE 4th Global Conference, S. 638.
 - [18] Farid, Z., Nordin, R. and Ismail, M., 2013. Recent Advances in Wireless Indoor Localization Techniques and System. *Journal of Computer Networks and Communications*, 2013, 1-12
 - [19] R. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
 - [20] G. F. Welch and G. Bishop, "An introduction to the kalman filter," *University of North Carolina, Chapel Hill, NC, USA, Tech. Rep.*, 1995

A Conceptual Design for Managing Internet of Things Devices in Emergency Situations

B.KAYMAZ¹ and T.ERCAN¹

¹ Yasar University, İzmir/Turkey, kymzbrk@gmail.com

¹Yasar University, İzmir/Turkey, tuncay.ercan@yasar.edu.tr

Abstract - Internet of Things has entered our lives as many forms and devices and enabled us to connect even the smallest-size devices to the internet. This progression aided us by deploying sensors on the workspace we need, such as hospitals, buildings, all sort of vehicles, logistics, even automated manufacturing and factories. In terms of quantity, around 8 billion of sensors are connected to IoT systems by 2017. These sensors may be utilized in many ways, such as emergency reporting and data collecting. IoT mesh networks function not only in normal conditions, even in emergency situations such as climate disasters, fire, flood, earthquake, tsunami, war or terrorism-related nuclear, biological or chemical attacks or conventional attacks. Such circumstances cause mass panic and chaos resulting in serious consequences unless contained and controlled. In this study, we propose a conceptual design of a functioning prototype which has the operational ability to provide an auxiliary network connection, as well as detecting Internet of Things devices in the area of disaster, collecting IoT device info and providing a mobile WAN access to the user.

Keywords – IoT, Emergency Situations, Disaster Control System, Wireless Network.

I. INTRODUCTION

As Moore stated that while processing power increases and the cost of this power decreases, technology becomes more abundant and cheaper by time, in Electronics Magazine issued in 1965. This law can be observed today and leads us to a degree of freedom where any device could be connected to the internet. Internet of things can be considered as a new way of revolution just before the industry 4.0. As the automation replaces labor, internet of things makes us able to connect to the cloud by even our refrigerators. This shift in paradigm provides us with more flexibility in connectivity with our wireless computer networks.

Deployed sensors produce data such as thermal, visual, sonic based and even electromagnetic. Refineries, dams, and factories, for instance, utilizes these sensors to obtain information about critical systems, while decision-making systems are dependent on these sensors [1].

The exact definition of the IoT is a bit blurred just because there are many definitions and details that define this concept. It can be defined as connecting the virtual things such as information and data with physical things that we interact with every day. As one can easily understand, there are many technologies that include the IoT in their core, subsequently,

this leads to a blurred line between these technologies, smart dust, wireless sensor networks, smart grids, smart cities are considered to be in the scope of the IoT as well [2]. It could be better to understand that all these technologies have sensors and internet connection is delivered via secondary modules such as gateways or additional helper nodes depending on the severity of the system. So there are many definitions around the Internet of Things. It is better to mention about these definitions from multiple researchers.

The name "IoT" first used by Kevin Aston in 1999, Kevin states that it is better to have data gathered by the Internet of Things which eliminates the human fault on collected data, consequently management of our devices and systems could be more accurate and manageable. Such as periodic maintenance for a critical system that affects many individuals. IoT integrates the digital world of information technology to our physical world with the help of sensors and actuators [3]. Rob van Kranenberg states that IoT delivers a global connection of dynamic entities via the help of standard computer network protocols, using the sensory interfaces actively and contributing the data gathering [4].

In light of this knowledge, IoT devices will be around us in the near future, shaping our daily routines and affecting our lives in any field. It is essential to achieve utilization of these devices not just providing data for mega-corporations but providing a safe environment for people. Almost every year, a natural disaster occurs without a linear timing, location and may differ as well. It can be assumed that early response to an emergency could decrease the fatalities consequently reducing the mass panic and disorder among people. It is critical to note that, natural disasters may be significantly colossal in magnitude so the aftermath is perpetuated by the infrastructure which rendered uselessly. Moreover, infrastructure has limitations since it is used by anyone panicked nearby consequently congestion affects the quality and reliability of infrastructure.

There are many applications of management of emergency situations differentiated by an approach such as software-defined networks, emergency IoT protocols, auxiliary communication gateways, unmanned drone disaster discovery, emergency data provided by mobile phone usage, smart detection systems that provide analysis on network traffic and more. Since IoT can be applied to many different systems such as healthcare or smart cities, collected data have a variety about the data type. These varieties make it possible to approach from

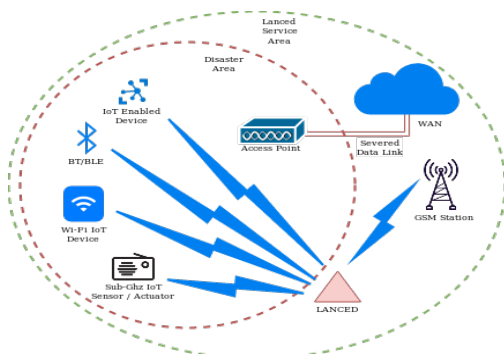
many angles which also differentiated by data type and application. In this study, we tried to develop a functioning prototype of a smart control box which has the operational capability;

- to provide an auxiliary network connection,
- to detect IoT devices in the disaster environment,
- to collect their device information,
- to provide a mobile WAN access to the user as the first emergency responders.

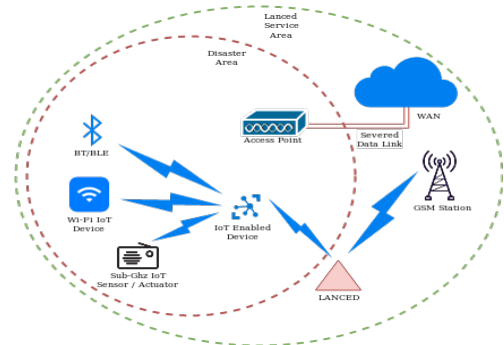
The remaining part of the paper is organized as follows; the overall system design of our prototype is presented in section II. Section III contains the system setup and how the functions in smart box work. Discussions for HW and SW components are given in section IV. Section V concludes.

II. SYSTEM DESIGN AND RELATED LITERATURE

There are two scenarios for the proposed device, shown in Figure 1. In Figure 1a, IoT devices in the area are still capable of data transmission, while infrastructure is not available for WAN access. The LANCED device is able to detect various IoT devices, providing a brief status for encrypted data transmission while providing actual data for devices which are broadcasting messages not encrypted. The LANCED device can be previously authorized by an operator to act as an auxiliary network connection which can be supplied via GPRS connection, as a result device in the disaster area is able to resume their operation, shown in Figure 1b.



(a) IoT device data in disaster site are collected by LANCED then routed via GPRS connection via nearby GSM station.



(b) IoT device data in disaster site are routed via LANCED as an auxiliary Access Point, previously set by the authorized operator.

Figure 1: LANCED operation

The prototype for LANCED is consisted of sensing and communication functionalities, provides detection for Wi-Fi, Bluetooth and Bluetooth Low Energy and Sub-GHz IoT nodes. The prototype also can provide internet connection via GPRS module, as well capable of delivering Access Point service for operators in the disaster area, as well as authorized users in the vicinity.

Proposed device prototype handles the tertiary modules interfaced via a microcomputer with an Operating System. Raspberry Pi 3 B+ is preferred as a base board, supplying required interfaces for one or two external Network Interface Card, a Global Positioning System module, a Software Defined Radio module. On-board Bluetooth, Bluetooth Low Energy module is preferred, also can be extended by any external BT/BLE modules. General Purpose Input Output Pins on the Raspberry Pi board are also used for interfacing with GPRS shield attached on the board. On-board Network Interface Card is used for providing a mobile Access Point, can be used to supply network connection for both operator and users. The prototype device is contained within a chassis made of Poly Lactic Acid, produced by 3D printing. The chassis holds the Raspberry Pi board, as well as external modules bolted on the chassis, while allowing for quick release for each external module, components of chassis, are displayed in Figure 2.

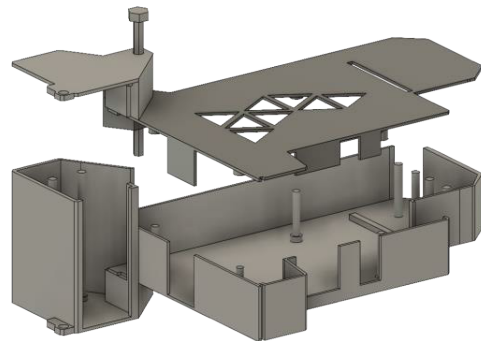


Figure 2: L.A.N.C.E.D. Poly Lactic Acid Chassis

In terms of a computer network, congestion is an issue in case of emergency situations which is a point of failure of many interconnected systems. To overcome this issue, there are several protocols based SDN methods. The article by Qiu et al. discusses real-time data transmission control which provides less packet loss and energy consumption. Authors provide two main methods which are Delay Iterative Method (DIM) and Residual Energy Probability Choice (REPC). DIM calculates the delays between responses for the nodes then re-configures the neighbors providing a solution for ignored valid paths. REPC is a novel method presented by this research, exchanges data about the energy of the node between other nodes resulting in the decision of becoming the forwarding node [5].

Minimal designs in IoT provides mobility and accessibility which can be used in wearable technologies designed for any situations such as firefight and healthcare. Data collected though sensors from the service personnel provides live feed on

the situation. In this matter researchers proposed a wearable technology for detecting and containing chikungunya virus [6]. The proposed system provides real time data collected from IoT sensor layer then passed through fog layer where classification and alert generation are handled. The last step is cloud layer which consisted of storage and indexes for government agencies and hospitals. Since the virus is spread via female mosquitoes, provided solution is fast, reliable and provide accuracy since live feed is provided.

Another wearable technology named CROW (Critical and Rescue Operations using Wearable Wireless sensors networks) is proposed by Ben Arbia et al. in [7]. Rescue personnel on the disaster site is connected through network connection which is handled by Optimized Routing Approach for Critical and Emergency Networks routing protocol. CROW system allows users to generate and feed data using different types of hardware which are Raspberry Pi, sensors and smart phones.

It is essential to note that IoT provides communication between vast numbers of devices with various types which enables researchers to use hybrid systems. These hybrid systems are able to collect different data and derive an approach to the problem with multiple angles. Ref [8] proposes a hybrid system for providing emergency management in case of a natural disaster. The scenario is based on a river with the hazard of flooding, the rescue team is mobilized after using data collected from a server called GeoData Collector which provides social network data. Augmented reality (AR) is aiding the team via displaying related information about the site such as sensors in close approximation while the disaster is scaled by mapping which is provided by Analytic Dashboard. Smart cities are covering a wide topic of IoT devices, a vast number of infrastructure elements are included for sensing and data collecting. Such elements are also in the scope of disaster recovery since lack of management would result in civilian casualties. Another proposed system that brings a solution for tunnels provides automated sensing in times of hazard following a warning signal pattern for the civilians in the area [9]. The system is fed by multiple data sources such are temperature, pressure, wind, humidity, light, which are then processed by the system micro-controller. Energy consumption is also balanced via the controller.

Communication mediums have a variety of types in the result of different devices and methods. Different types of communication mediums provide reliability as well as options of use. It is critical to state that in times of recovery or disaster, different communication methods could be lifesaving, in case of an infrastructure failure of any of these mediums. The research in [10], proposes a backup communication medium in case of an infrastructure failure. Main feature of the system are situational awareness and relay communication medium to recover from severed data communication. The system is monitoring the deployment area via sensors for emergency, then triggering the backup communication medium if the base station is down or is congested. Integrated relay node is able to boost the signal of the base station or replacing it to recover from severed connection.

Among alternatives, our approach brings a novel system

designed to provide mobility, ease of use and situational enumeration for first responders, providing a communication hub and WAN access for nearby users.

Our prototype is mainly designed to detect Wi-Fi devices in close proximity with data collected by analyzing broadcast messages, manufacturer data, and signal strength. Bluetooth devices are also in scope of targeted devices so any devices in the subject of matter disaster site shall be recovered via analysis of collected data during our session. The last requirement is Sub-GHz modules which are collection nodes reporting data to other IoT devices in the vicinity. L.A.N.C.E.D. (Lightweight Automated Network Component Enumeration Device) will be organizing several modules simultaneously to achieve data collection of IoT devices with Wi-Fi, Bluetooth/BLE and sub-GHz capabilities while providing WAN access as an auxiliary connection for maintaining data access. Figure 3 depicts our prototype design.



Figure 3: L.A.N.C.E.D. Mark I device mounted with GPS, GPRS, SDR and two NICs

III. SYSTEM SETUP

The system will be utilizing several open-source programs as well as essential hardware devices for operation. Lanced device will be providing multiple interfaces for the user to operate the device covering monitored and monitorless usage options, can be seen in Figure 4.

LANCED device will be using a BASH program for automated installation (lanced_installer) and another one (lanced) for GUI and operation. BASH language will be used to implement almost all of the program, to complement the capabilities of BASH, Python language may be needed, provides wide range of libraries also easy to implement and use with other programs. Main operations will be handled by BASH, interfacing with hardware and software, storing data in respective folders, also installation will be done by another part of the main program.

LANCED is designed to handle automation with aim of reducing user interaction resulting in reduced response times that is critical for managing emergency situations. Operation of the device lets the user to have an elevated view of disaster site that contains IoT devices with Wi-Fi, Bluetooth and sub-GHz variations. LANCED menu offers minimal design for the user while allowing operation and diagnostics for the hardware and

software modules for troubleshoot. Operation of the proposed system is starts after powering up the device. There are three modes of operation are available for the user for operation.

First method allow user for quick deployment, after the OS boot process, LANCED is able to start the monitoring process.

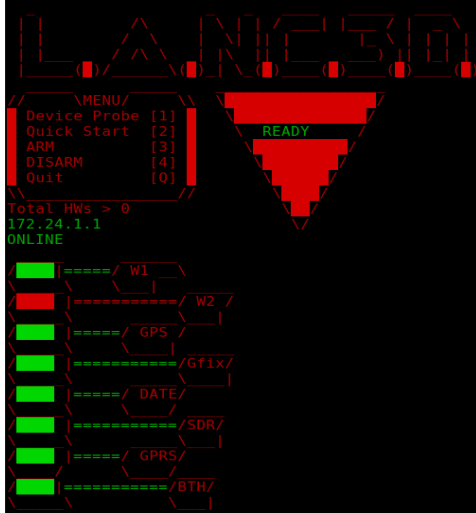


Figure 4: L.A.N.C.E.D. Graphical User Interface

Second method allow the user to interact with the proposed system via multiple ways. HDMI display port on RPi board is accessible for the user to connect any type of monitor to provide graphical feedback. The user also is able to connect any keyboard and mouse via any of unused USB ports on the proposed system. Terminal variation allows user to interact with CLI screen. Compability for certain ANSI characters have constraints in different terminal types, to overcome compability issues, Linux TTY terminal supported ANSI characters preferred for user display. Linux tty encoded characters allow LANCED to be displayed with any other terminal consequently GUI is supported by any platform with different types of terminals.

Third method of use is possible with connection via SSH through any mobile phone with installed terminal application, or a laptop with UNIX based OS installed. RPi internal Wi-Fi NIC is used as an AP to provide Wi-Fi connection for one or more users, after connection is established with the LANCED AP, terminal access grants user to run required programs for operation. AP ip is previously defined while setting up Hostapd and is displayed under MENU section (see Figure 4). Other LANCED functions are given in Table 1.

Table 1: LANCED functions example

Function name	Tasks/Description
components_chk()	hardware and software module diagnostics is performed for user GUI report
start_ks()	if overall status of the device is ready, start device discovery
stop_ks()	End processes of discovery programs
ln_quit()	End Point-to-Point protocol daemon then power off GPRS module

Capability of AP of the proposed system is achieved with program called Hostapd [12], with parameters set by LANCED installer. Setting of the AP can be modified as desired after installation of the software modules. Hostapd is a user space daemon for access point and authentication servers, under the terms of BSD license. It implements IEEE 802.11 access point management, IEEE 802.1X/WPA/WPA2/EAP Authenticators, RADIUS client, EAP server, and RADIUS authentication server.

Organization of the hardware modules and software programs are handled with LANCED.sh. GUI is consisted of sections with minimal design. Ease of use based on minimal user selection with five keys at most displayed in MENU section on GUI (see Figure 4). After boot process is completed, user is able to run the following functions in Figure 5, by pressing keys given in MENU.

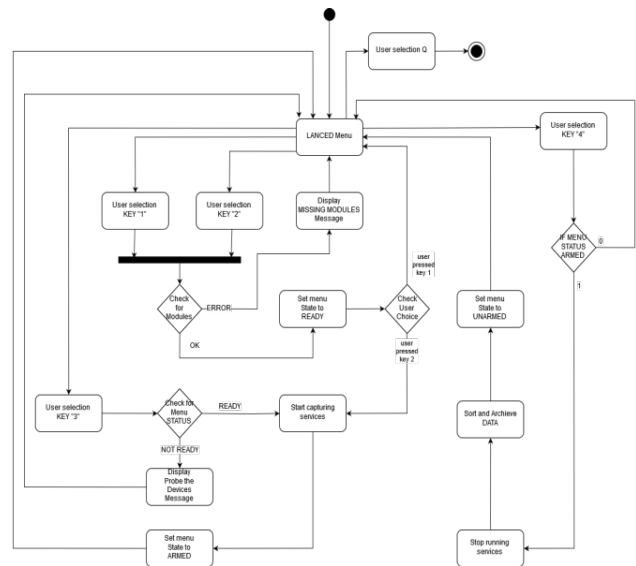


Figure 5: LANCED program activity diagram

The LANCED system achieves Wi-Fi AP tracking capability with using program called kismet and one or two Wi-Fi NICs. Kismet is an 802.11 wireless network detector, sniffer, and intrusion detection system, under the terms of GPL-2.0 license. Kismet also supports a plugin architecture allowing for additional non-802.11 protocols to be decoded. Kismet identifies networks by passively collecting packets and detecting networks, which allows it to detect (and given time, expose the names of) hidden networks and the presence of non-beaconing networks via data traffic [13].

Broadcast messages from GPS satellites containing one-way range data of the GPS satellite locations are correlated with three or more GPS satellites with the replica signal that is created by the user. Provided four GPS satellites, latitude, longitude, altitude and the correction to the user's clock are determined [14]. Prototype uses a GlobalSat G-STAR IV BU-353S4 is a GPS receiver that uses SiRF Star IV GSD4e chipset, running with frequency of 1575.42 MHZ. GPS acquisition also provides Raspian OS time correction since RPi board does not

contain a RTC module on the board.

In addition the Android smartphone operating system (from version 4.0 onward) uses GPSD to monitor the phone's on-board GPS, so every location-aware Android app is indirectly a GPSD client [15]. LANCED program is able to detect if there is a mobile phone acting as a GPSD server, in case of absence of GPS hardware module, the user is not constrained to use a GPS hardware module connected via USB port. After connection of the module is detected, interface connection bar is displayed in green text is changed from red (Figure 4).

Radios are consisted of modulators, demodulators and tuners. Implementation of these hardware is possible with the software based applications with the modern computing. Software defined radios are the result of modern computing, is available to everyone providing functionality of a wide band radio scanner. Received signals are filtered with band selection filters with a processor resulting in desired sampled data output [16]. SDRs have many functions, radio scanning, tracking or receiving messages from weather balloons, communicating with amateur radios, watching TV broadcasts, receiving GSM signals, listening to satellites, listening to FM radio, providing a high quality entropy source for random number generation are just the tip of vast number of functionalities. SDR module installed on the proposed system is probed for connection interface followed by status check, if succeeded with probing then LANCED program indicates the progress on the GUI (Figure 4).

SubGhz devices those are subjected for monitoring on the disaster site are scanned via SDR module. Devices on the site broadcasting radio messages with frequency of below 1 GHz are have standards which are not entirely mapped for detection, in some cases some devices are in requirement of RF signal to be mapped. Open-source program named rtl_433 under GPL-2.0 license, supplied with required capabilities is able to scan for designated devices. Rtl_433 is a generic data receiver mainly for the 433.92 MHz, 868 MHz, 315 MHz and 915 MHz ISM bands [17].

GPRS technology is aimed to provide efficiency on the data sources utilized by GSM license holders, ensuring physical resource sharing between GSM services. GPRS module installed on the RPi board is interfaced with GPIO pins, and mounted on top of the board. LANCED program initializes the GPRS shield after checking the power status of the GPRS shield.

The Point-to-Point Protocol (PPP) provides a method for transmitting datagrams over serial point-to-point links. Internet connection then is shared with devices connected via AP. Raspian OS then is able to deliver WAN access via iptables program which forwards requests through ppp0 interface. Internet connection status is displayed under AP IP, with a green ONLINE text for WAN access and red OFFLINE text if the connection is not supplied.

Detection of Bluetooth and BLE supported IoT devices are supplied with onboard Bluetooth. LANCED program probes for Bluetooth device status for status indication via GUI. Discovery of Bluetooth and BLE devices depending on the program named BlueHydra, built on bluez library [18].

GUI indication of overall status indicated in the triangular display with green READY message after system integrity is verified. The LANCED system is then ready for operation, the user can trigger the monitoring process by pressing 3 (Figure 4). Monitoring is in the process until receiving a halt signal from the user by pressing 4, resulting in the start of DISARM process. Both ARM and DISARM process regulates run and halt signals to respective programs, stopping or starting them in a proper way.

LANCED program regulates collected IoT device data after each operation process, processed data is saved into the respective folder which is created and named by LANCED, and also timestamps are used to sign data for duplicated data. The LANCED program is also keeping track of unique devices to avoid data duplication. A number of collected devices are indicated after each operation informing the user (Figure 4).

IV. DISCUSSION OF OVERALL SYSTEM

Our prototype LANCED is a multi-purpose smart control box. Its main body is based on Raspberry Pi 3 microcomputer. It includes some additional components required by different functionalities like G-STAR IV GPS module with USB connection, Software Defined Radio module with USB connection, RTL-SDR, Raspberry Pi GSM/GPRS Shield from Sixfab, Wi-Fi Network Interface Card, TP-LINK TL-WN722N. They bring together a functioning prototype which has operational capabilities for different emergency scenarios. They enable auxiliary network connectivity, can detect the IoT devices in use, and collect their data to discover the ownerships of different devices in the disaster area.

LANCED has a user-friendly GUI and helps user interaction for managing alternative controls in emergency situations. LANCED operation helps users to have an elevated view of the disaster site from the point of IoT and smart devices. The system utilizes open-source programs as well as essential hardware devices for operation by its generic and integrated design. Even though it is designed to detect Wi-Fi and Bluetooth devices, ZigBee and other smart devices working in different Industrial Communications protocols can be covered depending on the interfaces embedded into the system.

V. CONCLUSION

LANCED is designed to handle automation for detecting IoT or smart devices and collecting their ownership/usage information in emergency situations. It reduces user interaction and utilizes critical response times in disaster scenarios. The LANCED menu offers minimal design for the user while allowing operation and diagnostics for the hardware and software modules for troubleshooting.

LANCED supports the users, rescue teams in order to have a monitoring capability for IoT devices of Wi-Fi, Bluetooth, and sub-GHz variations.

REFERENCES

- [1] L. Da Xu, H. Wu, S. Li, "Internet of things in industries: A survey," *IEEE Transactions on industrial informatics*, 2014, 10.4: 2233-2243.

- [2] U. Dieter, H. Mark, M. Florian, "An architectural approach towards the future internet of things," In: *Architecting the internet of things*. Springer, Berlin, Heidelberg, 2011. p. 1-24.
- [3] A. Kevin, et al., "That 'internet of things' thing," *RFID journal*, 2009, 22.7: 97-114.
- [4] K. Rob van, *The Internet of Things: A critique of ambient technology and the all-seeing network of RFID*, 2008.
- [5] T. Qiu et al., "ERGID: An efficient routing protocol for emergency response Internet of Things," *Journal of Network and Computer Applications*, 2016, 72: 104-112.
- [6] SK. Sood, M. Isha, "Wearable IoT sensor based healthcare system for identifying and controlling chikungunya virus," *Computers in Industry*, 2017, 91: 33-44.
- [7] D.B. Arbia et al., "Enhanced IoT-based end-to-end emergency and disaster relief system," *Journal of Sensor and Actuator Networks*, 2017, 6.3: 19.
- [8] G. Luchetti et al., "Whistland: An augmented reality crowd-mapping system for civil protection and emergency management," *ISPRS International Journal of Geo-Information*, 2017, 6.2: 41.
- [9] A. Ángel et al., "Managing emergency situations in the smart city: The smart signal," *Sensors*, 2015, 15.6: 14370-14396.
- [10] C.H. Lee et al., "Implementation of relay-based emergency communication system on software defined radio," In: *2015 IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS)*. IEEE, 2015. p. 787-791.
- [11] P. Shelley et al. *UNIX power tools*. O'Reilly Media, Inc., 2003.
- [12] HOSTAPD (2002). from source: <https://w1.fi/hostapd/>
- [13] KISMET (2002). from source : <https://www.kismetwireless.net/>
- [14] W.P. Bradford et al., *Global positioning system: Theory and applications, Volume II*. American Institute of Aeronautics and Astronautics, 1996.
- [15] GPSD (1995). from source: <http://catb.org/gpsd/>
- [16] K.J. Friedrich, "Software-defined radio: basics and evolution to cognitive radio," *EURASIP journal on wireless communications and networking*, 2005, 2005.3: 275-283.
- [17] RTL_433 (2012). from source : https://github.com/merbanan/rtl_433
- [18] Pwnieexpress. (2016) "Blue Hydra," from source : https://github.com/pwnieexpress/blue_hydra

Can we Fight Social Media with Multimedia Learning Systems?

I. VESELINOV¹ and Z. KOTEVSKI²

¹ “St. Kliment Ohridski” University, Bitola/Macedonia, ivan_veselinov@yahoo.com

² “St. Kliment Ohridski” University, Bitola/Macedonia, zoran.kotevski@uklo.edu.mk

Abstract - The latest information technologies enabled the world to become a place with more and more easily accessible information. Immense quantities of information are exchanged on a daily basis using contemporary communication technologies, especially by the means of social media, but are we becoming more knowledgeable? The latest reports (World Bank report and PISA testing results) reveal that high school students from Macedonia showed degrading results concerning knowledge. Internet statistics about social media usage and the survey we have conducted show that high school students in Macedonia extensively use social media and find multimedia learning quite helpful, but rarely use computerized learning systems in classes. The purpose of this research is to search for an opportunity to partly divert students from social media usage to spend time on Multimedia Learning Systems (MLS). The final goal is to develop a MLS that would be interesting enough to attract students to use it and have the opportunity to gain knowledge, without entirely sacrificing their social networking needs. To develop such system we conducted a survey and collected data about the students' affinities and needs. In this paper we present the results of the survey and propose a framework of prerequisites that a novel MLS should incorporate.

Keywords – Information technologies, Social media, Multimedia learning.

I. INTRODUCTION

THIS research concentrates on three crucial moments: the benefits of multimedia learning, the current situation with knowledge of high school students in Macedonia and the social media usage. In the context of learning by the aid of multimedia, the research in the past few decades inevitably implies that implementing multimedia in education is highly beneficial to learning by providing improvements to students' cognitive abilities, acceleration of memorization and alleviation of understanding of abstract, micro and macro entities. For example, Lee et al. [1], as pioneers in the field of multimedia learning, reported that this method of instruction demonstrates great capacity to enhance traditional teaching. They also show that, when applied correctly, it can bring breadth and depth to the subject, render efficient use of class time, create flexibility in teaching and enhance students' learning. Mayer [2], as one of the greatest researches in the field, also concluded that, under some conditions, multimedia learning can lead to substantial improvements in learning and

that students are better able to make sense of scientific or mathematical explanations when they are able to hold relevant visual and verbal representations in working memory at the same time. Garcia et al. [3] showed that, in some specific case of descriptive geometry, the use of flash animations accelerates the development of students' spatial perception. Lin et al. [4] investigated the potential benefits of using animation, visual cueing, and their combination in a multimedia environment designed to support learners' acquisition and retention of scientific concepts and processes, concluding that participants provided with animations retained significantly more concepts than their colleague peers provided with static graphics. Kim et al. [5] explored the effects of gamification of learning and concluded that, after some period of adaptation, the learning curve is quite steeper than with regular learning. Similar conclusions are presented by many other researches, such as: Arcelli et al. [6], Mayer et al. [7], Milovanovic et al. [8], Pérez-López et al. [9], Barra et al. [10], Surjono et al. [11], Tibbitts et al. [12], Gilakjani [13], Kumar et al. [14], Bittman et al. [15], Mai et al. [16], Chen et al. [17], Marsono and Wu [18], Adesope et al. [19], Ocepek et al. [20], Danielson et al. [21], Leow et al. [22], Smith et al. [23], Mayer et al. [24], Jeong et al. [25], Scheiter et al. [26], Park et al. [27], Schweppe et al. [28], to name a few. In addition, the effort of Almarā'beh et al. [29], that explored the effectiveness of multimedia learning, confirm that multimedia technology empowers the educational processes by means of increased interaction between students and teachers, which makes learning more dynamic and longer lasting and it is applicable in the world outside the classroom. Thus, we know the benefits, but some questions arise. Do we properly implement multimedia technologies in education? Are the students inspired to use multimedia in their learning?

The statistics about social media, on the other hand, reveal that their usage is still getting momentum. According to Brandwatch [30], as of January 2019, there are 3.397 billion active social media users in the world. The number of social media users has grown by 320 million users between September 2017 and October 2018, which implies that there are almost 10 new social media users each second. The average daily time that people spend on social media is almost two hours and 81% of the teenagers feel that social media has a positive effect on their lives.

But, if we analyze knowledge statistics like the World Bank

report on public expenditure for FYR Macedonia (Report No. 93913-MK) from 2015 [31] and the results of the latest PISA testing [32] of the same year of 2015, we can observe some disturbing results. The report from the World Bank [31] explains that *"The quality of primary and secondary education seems to have declined", "In 2013, Macedonia's spending on education of 4.1 percent of GDP was the highest in the region, but its education performance indicators were below the regional average. The quality of primary and secondary education has not improved significantly over the last several years, as measured by international tests. Scores on the Trends in International Mathematics and Science Study (TIMSS) declined significantly between 1999 and 2011, and are below other European countries.", "The levels of skills and knowledge of Macedonian students lag behind those of students in comparable countries. For example, on PIRLS 2006, which assesses children in the fourth year of formal schooling on a range of reading comprehension strategies, fewer children reached the lowest benchmark in FYR Macedonia (55%) than all neighboring and EU countries which participated (the next lowest was 83% for Slovenia). On the TIMSS assessment, which test children in grade 8 in mathematics and science, Macedonia did significantly worse between 1999 and 2011 as the average mathematics score fell from 447 to 426, with no other European country scoring less. The proportion of FYR Macedonian students who achieve the minimum standard (the 'low international benchmark') fell from 70 percent to 61 percent in mathematics and from 73 percent to 53 percent in science".* Regarding the PISA testing [32], the results showed that 62.9% of the Macedonian high school students are on or below level 1 (from max. of 6) for science, and in mathematics and reading that percentage is even higher (just above 70%). This average in OECD countries is below 20%, which indicates quite large difference in knowledge of students in OECD countries compared to Macedonian students. Level 5 for Science and Reading was reached by only 0.2% of Macedonian students where the OECD average is 6.7% and 7.2% respectively, while level 5 for mathematics reach only 0.7% with OECD average of 8.4%.

Putting these three moments together we learn that multimedia technologies and multimedia contents bring multitude benefits to education and learning, that the world population, especially young people, increasingly use Information, Communication and Multimedia (ICM) technologies, but the knowledge they gain stays stable (with slight decline) and for some countries, such as Macedonia, it deteriorates. Consequently, the question "why is this happening" arises and is there something we can do to prevent it? In this manner, we conducted a survey in a municipality in Macedonia, to search for some answers concerning the high school students' behavior regarding the usage of ICM technologies, in or out of the school, for the purpose of their education. The final goal of this research is to propose a framework of prerequisites that a novel Multimedia Learning System (MLS) should incorporate. The system should be

interesting enough to attract students' use and provide them the opportunity to gain knowledge without entirely sacrificing their social networking habits. We believe that this can be accomplished by searching for an opportunity to partly divert students from social media usage to spend certain time on MLSs. The survey provided quite interesting but expected results that we present henceforth.

The rest of this paper is organized as follows. Chapter 2 gives an overview of Learning Management Systems (LMSs) and MLSs, their advantages and the obstacles for their implementation to support education in the context of multimedia learning. This chapter presents the motivation for this research as well. In chapter 3 we present our findings from the conducted survey among the high school students in a single municipality in Macedonia. In chapter 4 we discuss the survey results and we disclose essential prerequisites that a novel MLS should satisfy. Chapter 5 presents the concluding remarks and our projections for future work.

II. RELATED SYSTEMS AND MOTIVATION

A systems' category that offers certain possibilities for multimedia learning is the category of LMSs. Moodle [33], for example, is used for blended learning, distance education, flipped classroom and variety of e-learning projects in schools, universities and other workplaces. With customizable management features, it is used to create private websites with online courses for educators and trainers to achieve certain learning goals. Blackboard Learn [34] is a virtual learning environment and course management system. It is a Web-based server software that features course management, customizable open architecture and scalable design that allows integration with student information systems. Its main purposes are to add online elements to courses that are traditionally face-to-face delivered and to develop complete online courses with a few or no face-to-face meetings. Blackboard Learn provides users with a platform for communication and sharing content. Regarding the communication, the system enables posting announcements for students, students' chats, discussions and e-mail. Edmodo [35] is an educational system that offers collaboration and coaching platform to schools and teachers. The Edmodo network enables teachers to share content, distribute quizzes, assignments, and manage communication among students, colleagues, and parents. Edmodo is very teacher-centric in its design and philosophy, while students and parents can only join Edmodo if invited by a teacher. Many other platforms offer similar features, such as Ilias [36], Google Classroom [37], Canvas LMS [38], Sakai [39], Litmos LMS [40] etc., but what is common for the aforementioned LMSs is that they are mainly course management platforms that offer some features for multimedia learning, but cannot be easily customized to function as an MLS. The presented LMSs are not out of the box solutions and their administration can sometimes be difficult and not very user friendly, especially if someone tries to implement an MLS over an LMS platform. Furthermore, the teachers are usually not able to easily download an LMS and

be up and running the next minute with all the necessary preparations for multimedia learning.

With such situation with the LMSs and the specific requirements for multimedia learning, many efforts present designs of specific MLSs. In this manner, Lock Yen Low et al. [41] in 2003 outline the details of the development and deployment of a novel MLS, starting from content development flow to the implementation stage. The project is launched at the Multimedia University (MMU) in Malaysia. One of the drawbacks from the current point of view is that Adobe (Macromedia) Dreamweaver and Flash were used to develop the interactive environment within two university campuses. Nusir et al. [42] designed an interactive MLS for the children of primary schools in Jordan. After implementation, the authors reported improvement of students' abilities and their learning skills. Huang et al. [43] presented the design of an online MLS for improving students' perceptions of learning English. The system is constructed as a quiz offering video learning content. The experimental results demonstrated that students had a positive view of the functions of the video supported quiz system.

To summarize the characteristics of these systems, as we mentioned earlier, the LMSs are not quite convenient for implementation as MLSs and the specific MLSs are proprietary systems, build to suit certain educational environments. These properties make current LMSs and MLSs highly unsuitable for implementation, which was the main motivation to search for an alternative solution that would fulfill the projected goals of learning enhancements with a novel MLS according to the specific requirements of the stakeholders involved.

III. THE SURVEY

To explore the students' affinities and needs we conducted a survey among the high school students from all study years in eight schools from a single municipality in Macedonia. The survey was concentrated on the usage of ICM devices during or out of classes, for the purpose of learning (the use of multimedia content for learning, exchange of learning materials and communication with colleagues and teachers) and whether those technologies are helpful to students' improvement of acquisition of knowledge. The population of high school students in the municipality is 3.764 and the population of high school students in the country is 71.650. The sample from which we received filled survey responses was 166, thus, regarding the municipality, the widest confidence interval (at 50% of answers) at a confidence level of 95% is 7.44. Consequently, the widest confidence interval regarding the whole population is 7.6. In the following charts we present our survey findings that lead to some quite interesting resolutions.

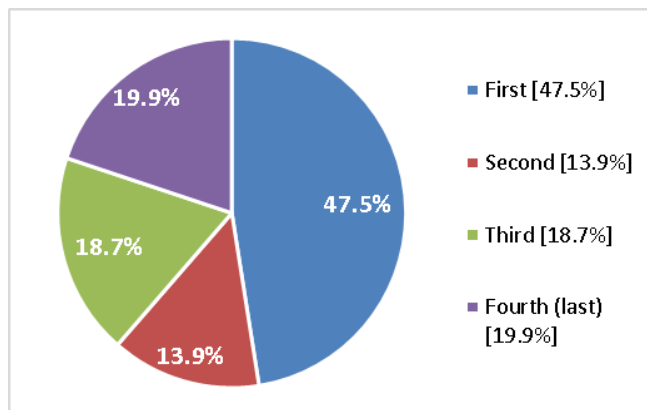


Figure 1: Distribution of survey respondents by year of study.

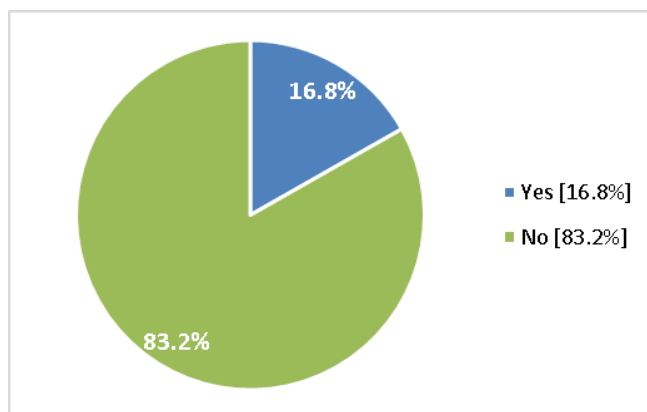


Figure 2: Distribution of subjects in which students use ICM devices and multimedia content.

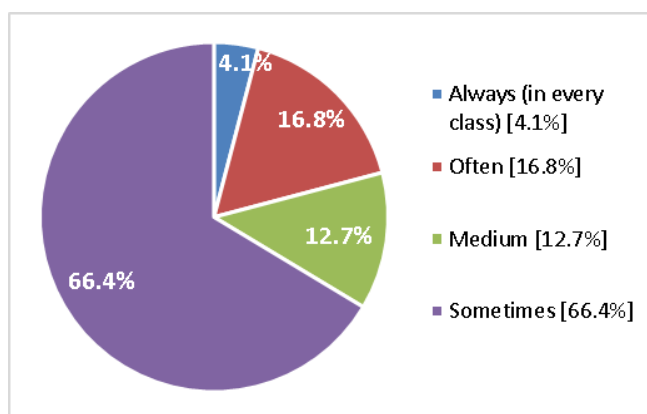


Figure 3: Frequency of usage of ICM technologies in classes of the 16.8% of subjects that implement ICM technologies.

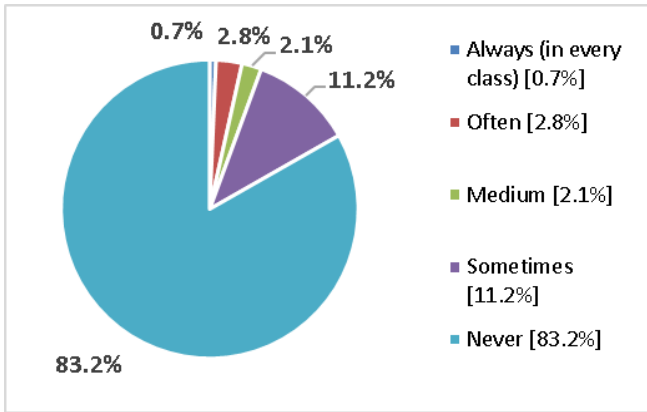


Figure 4: Frequency of usage of ICM technologies in classes of the total number of subjects.

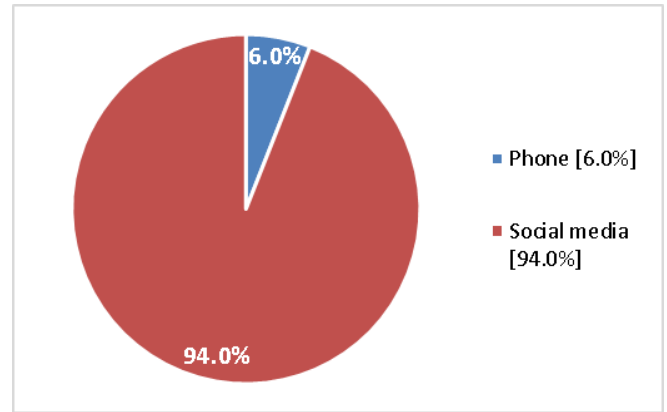


Figure 7: Most popular means for communication between students out of school premises.

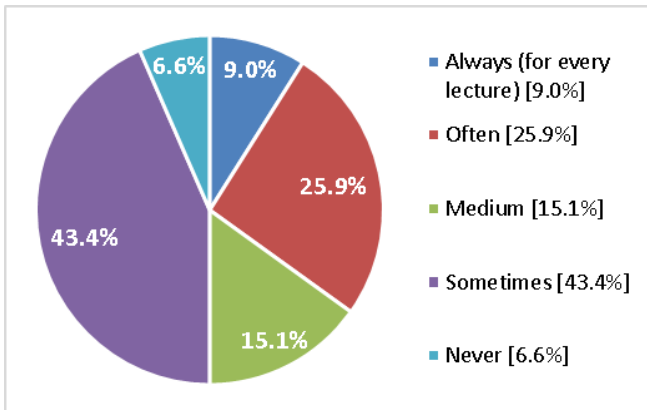


Figure 5: Students' frequency of searching for multimedia content to support their learning while studying at home.

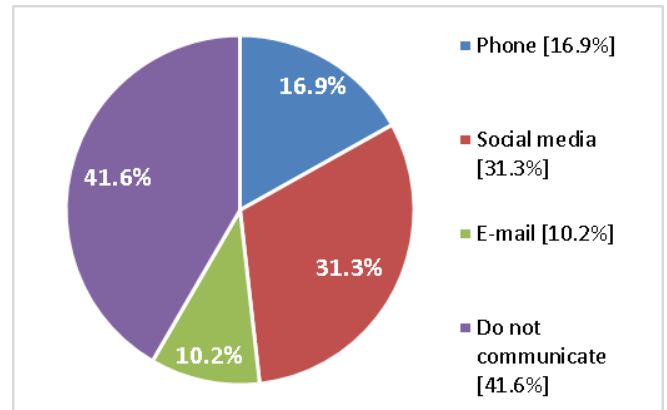


Figure 8: Most popular means for communication between students and teachers out of school premises.

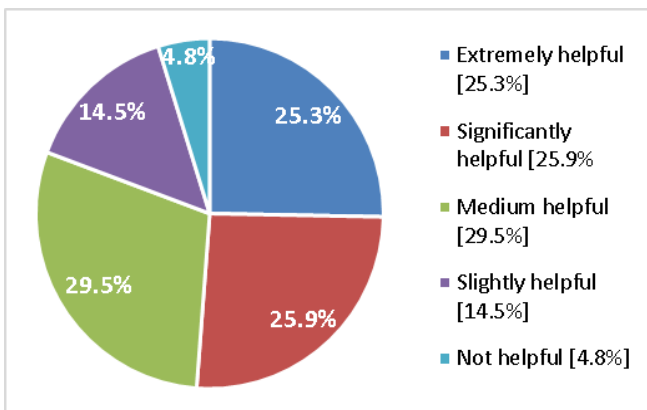


Figure 6: Helpfulness of multimedia learning according to student respondents.

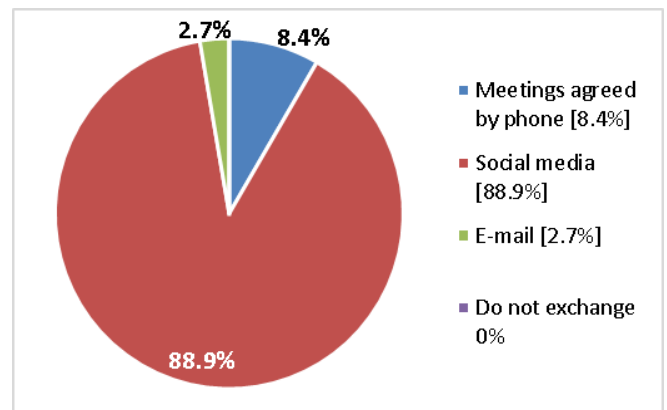


Figure 9: Most popular means for exchange of learning materials between students out of school premises.

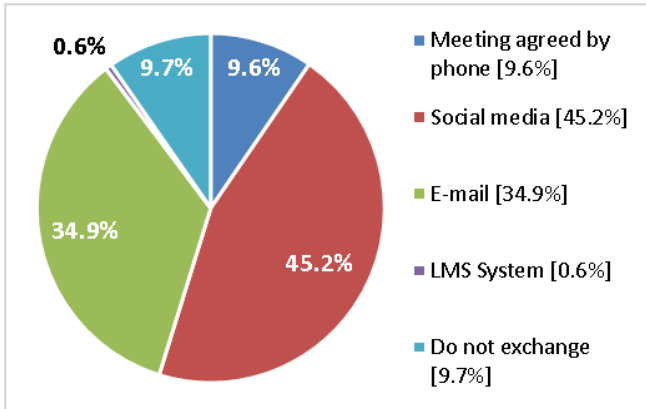


Figure 10: How do students most often receive learning materials from teachers out of school premises?

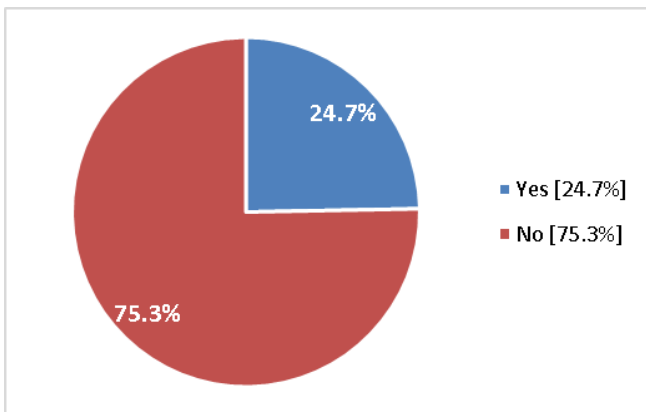


Figure 11: Students' usage of some LMS platform, such as Moodle, Edmodo, ATutor, Google Scholar etc.

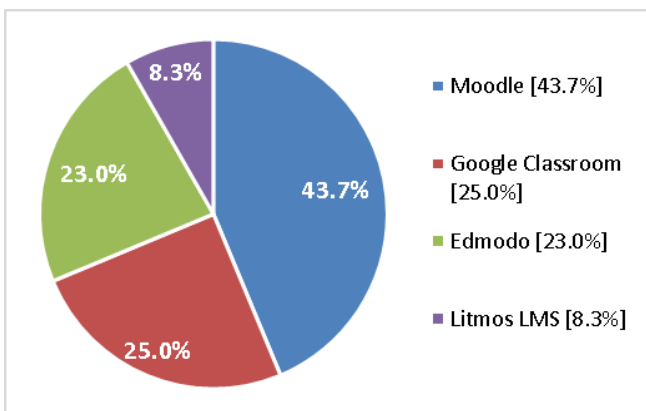


Figure 12: Distribution of usage of different types of LMSs for the 24.7% of students that sometime use LMS.

IV. DISCUSSION

The first noticeable finding of the survey is the fact that in only 16.8% of all subjects the students are allowed to use ICM devices such as desktop/laptop, tablet or smartphone, for the purposes of learning (Figure 2). This is the basic obstacle that

needs to be overtaken in order to enable the possibility for educational processes to advance using contemporary trends and multimedia technologies. It doesn't necessarily mean that ICM devices and multimedia content should be extensively used in every class lecture, but it is surely recommended that each subject implements some multimedia learning. The frequency of usage of ICM devices and multimedia learning content for the aforementioned 16.8% of all subjects is another unsatisfactory indicator (Figure 3). With only 20.9% of often to always to 66.4% of only sometimes we can conclude that the usage of multimedia content in classes is pretty scarce. If we represent this frequency against the total number of subjects and classes (Figure 4) we come to a tiny percentage of 3.5% for often and always, 2.1% of medium usage and only 11.2% of occasional usage.

But, is this what students need? Are they satisfied with these percentages of multimedia learning? Figure 5 and Figure 6 strongly indicate that the answers to these questions are negative. Almost 35% of the students often to always search for multimedia content on Internet to support their learning at home, and even 50% of them search for multimedia content for about the half of the lectures. Only tiny 6.6% of the students are not interested in learning with an aid of multimedia. What is even more promising is the fact that students find multimedia learning quite helpful. 51.2% of the students find multimedia learning significantly to extremely helpful and 29.5 answered that multimedia content is of medium help to their learning. Only 4.8% find it not helpful.

One of the most noticeable observations of this survey is the way that students communicate with each other out of the school premises. Figure 7 shows that 94% of the students use social media (Facebook, Twitter, LinkedIn etc.). Social media are the most popular for exchange of learning materials as well with 88.9% (Figure 9). Similarly, social media are also the most popular means for teachers to deliver learning materials to their students with 45.2%, while 34.9% of teachers prefer e-mail delivery (Figure 10). Even the communication between students and teachers out of school is basically conveyed via social media, even though more than 40% of the students have no communication with the teachers out of their schools (Figure 8).

Besides the usage of multimedia learning content and communication technologies, we were interested in the usage of Learning Management Systems (LMS). The majority of LMSs are basically class management systems and offer little options for multimedia learning, but it is an indication of the current situation in the schools and their interest in implementation of new technologies. It turned out that only 24.7% of students use LMS (Figure 11) and the most popular ones are Moodle, Google Classroom and Edmodo with 91.7% of combined coverage (Figure 12).

If we summarize the discussion we can compile several critical resolutions:

(i) High school students extensively use social media for multiple communication purposes, including exchange of

learning materials with their colleagues. As the statistics of Brandwatch [30] revealed, even 81% of the teenagers feel that social media has a positive effect on their lives. Through the social media they exchange colossal amount of information, but the sad part of such information exchange is that it rarely contributes to their learning and knowledge.

(ii) High school students are quite fond of using multimedia content to support their learning and find it very helpful. Many of them search for such content on their own because they do not receive multimedia learning support during their classes at school.

(iii) LMSs are not tailored for multimedia learning and are not easily deployable, while the rare MLSs are proprietary systems with features specific to the environment of their deployment.

(iv) Multimedia learning resources are scarce and have to be made for the specific requirements of the lectures.

These resolutions unavoidably lead us to a conclusion that a novel MLS may be the solution that will enhance learning and provide help to provision of knowledge. But, if we want to introduce a new system for students that are already used to working with popular social media, we must carefully plan its development. We believe that if we could divert some of the time that students spend working of social media to work on multimedia learning to gain knowledge, without them sacrificing their habits and communications, we may have a tangible solution.

As an outcome, we postulated the following essential prerequisites that a novel MLS should satisfy. Thus, the system should:

- be specifically tailored to the requirements of the educational system of implementation in compliance with the official learning materials issued by the government for the students in each year of study;
- enable all users to create and share multimedia content and connect the content to a specific lecture, but the introduction of newly created multimedia content should be controlled by one of the teachers that teach the corresponding subject;
- incorporate communications and incentives similar to the ones offered by the contemporary social media, in order to provide the users with some features that they are quite familiar with, and fond of as well;
- offer interconnection with the most popular social media services, such as Facebook, Twitter etc.;
- promote grading awards for the most productive students of multimedia content;
- be built simple, intuitive and user friendly, designed with a simple and clean interface;
- offer easily searchable content.

In our previous research [44] we already proposed an architecture for MLS with social networking features, based on a series of interviews conducted with high school teachers and this research supports our belief that such system will foster the provision of knowledge.

V. CONCLUSION

In this research we explore the possibilities to support learning and knowledge for high school students by development proposal of a novel MLS. In the present world, vast amounts of information are exchanged every minute and social media are becoming more popular every day. Such situation easily misleads that the more information available would mean more knowledge, but the latest international reports and testing results about the secondary education in Macedonia speak otherwise. The World Bank report [31] and the PISA testing results [32] reveal that the knowledge of Macedonian high school students is declining and lags quite behind the average of OECD countries. On the other hand, world statistics show that social media are extremely popular among teenagers and they even believe that social media have a positive effect on their lives. Taking into account the popularity of social media and the benefits of multimedia learning we conducted a survey among the high school students in a municipality in Macedonia to seek for an answer whether a novel MLS that will include social media features would present a solid support to learning. The survey revealed some interesting insights that students are quite fond of multimedia learning and find it highly helpful, but struggle to find appropriate multimedia content. Furthermore, their most popular means for communication and exchange of learning materials are the social media like Facebook, Twitter, LinkedIn, and e-mail communication has certain popularity as well. Since well-established LMSs are mainly built for class management, they are not easily implementable for multimedia learning, at least not for the specific requirements of a certain educational system. For this reason we propose several important prerequisites for a novel MLS that would support learning and knowledge. The idea is to partly divert student to use a MLS instead of just social media and therefore social media features must be included in the new learning system. In a previous research we already proposed an architecture for such system and this research of the behavior of high school students from Macedonia supports the proposed idea. Our future work will include activities for planning and development of the novel MLS and analyses after a certain period of usage. The idea is to achieve construction of a system in which every student can cooperate in order to facilitate and improve the effectiveness and efficiency of the MLS.

REFERENCES

- [1] Lee, Pui Mun, and William G. Sullivan (1996). Developing and implementing interactive multimedia in education. *IEEE Transactions on Education*, 39(3), 430-435.
- [2] Mayer, Richard E. (1999). Multimedia aids to problem-solving transfer. *International Journal of Educational Research*, 31(7), 611-623.
- [3] García, Ramón Rubio, Javier Suárez Quirós, Ramón Gallego Santos, Santiago Martín González, and Samuel Morán Fernanz (2007). Interactive multimedia animation with macromedia flash in descriptive geometry teaching. *Computers & Education*, 49(3), 615-639.
- [4] Lin, Lijia, and Robert K. Atkinson (2011). Using animations and visual cueing to support learning of scientific concepts and processes. *Computers & Education*, 56(3), 650-658.

- [5] Kim, Jung Tae, and Won-Hyung Lee (2015). Dynamical model for gamification of learning (DMGL). *Multimedia Tools and Applications*, 74(19), 8483-8493.
- [6] Arcelli, Francesca, and Massimo De Santo (2002). Multimedia distributed learning environments: Evolution towards intelligent communications. *Multimedia Tools and Applications*, 16(3), 187-206.
- [7] Mayer, Richard E., and Roxana Moreno (2002). Animation as an aid to multimedia learning. *Educational psychology review*, Plenum publishing corporation, 14(1), 87-99.
- [8] Milovanovic, Marina, Jasmina Obradovic, and Aleksandar Milajic (2013). Application of Interactive Multimedia Tools in Teaching Mathematics-Examples of Lessons from Geometry. *TOJET: The Turkish Online Journal of Educational Technology*, 12(1).
- [9] Pérez-López, David, and Manuel Contero (2013). Delivering educational multimedia contents through an augmented reality application: A case study on its impact on knowledge acquisition and retention. *TOJET: The Turkish Online Journal of Educational Technology*, 12(4).
- [10] Barra, Enrique, Sandra Aguirre Herrera, Jose Ygnacio Pastor Caño, and Juan Quemada Vives (2014). Using multimedia and peer assessment to promote collaborative e-learning. *New Review of Hypermedia and Multimedia*, 20(2), 103-121.
- [11] Surjono, Herman D. (2015). The Effects of Multimedia and Learning Style on Student Achievement in Online Electronics Course. *Turkish Online Journal of Educational Technology-TOJET*, 14(1), 116-122.
- [12] Tibbitts, George E., John R. Nicholas, and Ronald J. McKay (1978). Five Teaching Methods: A Comparative Study in Nurse Education. *Australian Journal of Education*, 22(1), 91-92.
- [13] Gilakjani, Abbas Pourhosein. (2012). The significant role of multimedia in motivating EFL learners' interest in English language learning. *International Journal of Modern Education and Computer Science*, 4(4), 57-66.
- [14] Kumar, Jeya Amantha, Balakrishnan Muniandy, Wan Ahmad Jaafar Wan Yahaya, and Wan Ahmad Jaafar (2016). Emotional Design in Multimedia Learning: How Emotional Intelligence Moderates Learning Outcomes. *International Journal of Modern Education & Computer Science*, 8(5), 54-63.
- [15] Bittman, Michael, Leonie Rutherford, Jude Brown, and Lens Unsworth (2011). Digital natives? New and old media and children's outcomes. *Australian journal of education*, 55(2), 161-175.
- [16] Mai, N. E. O., Ken Tse-Kian Neo, and Heidi Yeen-Ju Tan (2012). Applying authentic learning strategies in a multimedia and web learning environment (MWLE): Malaysian students' perspective. *TOJET: The Turkish Online Journal of Educational Technology*, 11(3).
- [17] Chen, Chiu-Jung, and Pei-Lin Liu (2012). Comparisons of Learner-Generated versus Instructor-Provided Multimedia Annotations. *Turkish Online Journal of Educational Technology-TOJET*, 11(4), 72-83.
- [18] Marsono, Wu, Mingchang (2016). Designing A Digital Multimedia Interactive Book for Industrial Metrology Measurement Learning. *International Journal of Modern Education and Computer Science (IJMECS)*, 8(5), 39-46.
- [19] Adesope, Olusola O., and John C. Nesbit (2013). Animated and static concept maps enhance learning from spoken narration. *Learning and Instruction*, 27, 1-10.
- [20] Ocepek, Uroš, Zoran Bosnić, Irena Nančovska Šerbec, and Jože Rugelj (2013). Exploring the relation between learning style models and preferred multimedia types. *Computers & Education*, 69, 343-355.
- [21] Danielson, Jared, Vanessa Preast, Holly Bender, and Lesya Hassall (2014). Is the effectiveness of lecture capture related to teaching approach or content type? *Computers & Education*, 72, 121-131.
- [22] Leow, Fui-Theng, and Mai Neo (2014). Interactive Multimedia Learning: Innovating Classroom Education in a Malaysian University. *Turkish Online Journal of Educational Technology-TOJET*, 13(2), 99-110.
- [23] Smith, J. G., and S. Suzuki (2015). Embedded blended learning within an Algebra classroom: a multimedia capture experiment. *Journal of Computer Assisted Learning*, 31(2), 133-147.
- [24] Mayer, Richard E., and Gabriel Estrella (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction*, 33, 12-18.
- [25] Jeong, Hwa-Young, and Sang-Soo Yeo (2014). The quality model for e-learning system with multimedia contents: a pairwise comparison approach. *Multimedia Tools and Applications*, 73(2), 887-900.
- [26] Scheiter, Katharina, and Alexander Eitel (2015). Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learning and Instruction*, 36, 11-26.
- [27] Park, Babette, Lisa Knörzer, Jan L. Plass, and Roland Brünken (2015). Emotional design and positive emotions in multimedia learning: An eyetracking study on the use of anthropomorphisms. *Computers & Education*, 86, 30-42.
- [28] Schweppe, Judith, Alexander Eitel, and Ralf Rummer (2015). The multimedia effect and its stability over time. *Learning and Instruction*, 38, 24-33.
- [29] Almara'beh, H., Amer, E. F., & Sulieman, A. (2015). The Effectiveness of Multimedia Learning Tools in Education. *International Journal of Advanced Research in Computer Science and Software Engineering*, 5(12).
- [30] Brandwatch (www.brandwatch.com) – April 2019.
- [31] World Bank Report. FYR Macedonia Public Expenditure Review – Fiscal Policy for Growth. Report No. 93913 – MK, July 2015.
- [32] OECD Programme for International Student Assessment (PISA). Results from 2015 International Testing.
- [33] Moodle LMS (<https://moodle.org/>).
- [34] Blackboard Learn (www.blackboard.com).
- [35] Edmodo (www.edmodo.com).
- [36] Ilias (www.ilias.de).
- [37] Google Classroom (classroom.google.com).
- [38] Canvas LMS (www.canvaslms.eu).
- [39] Sakai LMS (www.sakailms.org).
- [40] Litmos LMS (www.litmos.com).
- [41] Low, A. L. Y., Low, K. L. T., & Koo, V. C. (2003). Multimedia learning systems: a future interactive educational tool. *The internet and higher education*, 6(1), 25-40.
- [42] Nusir, S., Alsmadi, I., Al-Kabi, M., & Shardqah, F. (2011, April). Designing an interactive multimedia learning system for the children of primary schools in Jordan. In *Global Engineering Education Conference (EDUCON)*, 2011 IEEE (pp. 45-51). IEEE.
- [43] Huang, Y. T., Yang, T. C., Chen, M. C., Chen, C. M., & Sun, Y. S. (2016, July). Design of an Online Multimedia Learning System for Improving Students' Perceptions of English Language Learning. In *Advanced Learning Technologies (ICALT), 2016 IEEE 16th International Conference on* (pp. 327-331). IEEE.
- [44] Veselinov I., Kotevski Z. Rendeovski N., (2018). Architecture for Multimedia Learning System with Social Networking Features. *International Conference on Applied Internet and Information Technologies (AIIT 2018)*, Bitola, Macedonia.

Increasing the use of Digital Banking Applications for the Baby Boomer Generation by Easy Interface

U.OZDEN¹, A.ADIL SEPICI² B.KARADAĞ³.

¹Vakıf Katılım, İstanbul/Turkey, ufuk.ozden@vakifkatilim.com.tr

²Vakıf Katılım, İstanbul/Turkey, ahmetadil.sepici@vakifkatilim.com.tr

³Vakıf Katılım, İstanbul/Turkey, bulut.karadag@vakifkatilim.com.tr

Abstract - Today, the development of technological devices and applications has increased their use at the same rate. Especially the mobile and web applications with social media and entertainment content are actively used by the individuals X (1965 -1979), Y (1980 - 1999), and Z (1999 - 2012) generations. The pioneers of these applications are Facebook, Twitter, Instagram, Pinterest and YouTube. These generations use not only social media applications but also other digital ones. Especially, mobile and internet (digital) banking applications are mostly used by these generations. The generation before the generation X is called the Baby Boomer generation born between 1946 and 1964. It is seen that individuals belonging to this generation used social media applications but did not use digital banking applications at the same rate. In this study, easy interfaces will be designed to enable the individuals in the Baby Boomer generation to use digital banking applications that are easier to use. Social media-like digital banking applications will be exemplified. This will allow individuals in the Baby Boomer generation to use their digital banking applications more actively.

Keywords - User experience, Usability, Dijital Banking for Elderly, Easy Interface

I. INTRODUCTION

With the widespread use of smart phones and access to the internet everywhere, daily work is carried out through mobile applications. Especially young people follow mobile and web technologies closely. Social media, shopping, education, banking transactions etc. applications are actively used by individuals between the ages of 16-44. However, for the age group above 55 years, use of mobile and web technologies is at low rate [1]. Even the rate of use of digital banking applications is even lower. This age group uses social media applications especially on their mobile phones, but banking applications are not used at the same rate.

Some mobile and web applications are used more and some are not used enough. The success of any application is related to the suitability of usability based on user experiences [2].

A questionnaire was prepared for this study. Individuals who over 55 years of age were asked why they do not use digital banking applications. The complexity, lack of need and lack of trust were identified as the main problems. In this study, the

problem of complexity is investigated. A simple and useful front-end is designed for the use of more active digital banking applications for people over 55 who are called Baby Boomers. A kind of social media and entertainment content front-end will be created and complexity problems will be eliminated. In this way, the usage rates of individuals using digital banking applications will be increased.

Human population has exceeded 7.5 billion. Especially in developed countries, while the rate of young population declines, the proportion of elderly population is increasing. According to the population projection of 2050, 40% of the population of developed countries such as Japan, Spain and Portugal will be over 60 years old. 2.1 billion People in the world will consist of people over 60 years of age [3]. Conducting such a study will benefit the use of digital applications by older individuals in the future.

In this study, mobile and internet applications are designed for easy use of individuals over 55 years of age. In this way, it is aimed to increase the number of usage.

II. RELATED WORKS

In this article, it is seen that internet banking is not used by elderly people because of the difficulties in using computer [4].

In this article, it is seen that security and privacy issues are effective in the use of e-banking services [5].

This article has benefited from the factors in the use of internet banking for the elderly [6].

In this article, it is seen that the most important factor in the use of internet banking by elderly people is simplicity [7].

In this study, it is observed that complexity was effective as one of the reasons for not using mobile banking applications [8].

In this study, it is understood that the reason why elders do not use digital technologies is the problem of understanding the digital world [9].

In this study, the elderly people are aware of the technological advantages of mobile devices; however, it has been understood that design diversity and motivation factors prevent their use [10].

In this article, it is understood that the interfaces, which are easily used by the elderly users, will increase the use of digital games [11].

In this study, it is seen that designing easier interfaces for the elderly is an important factor in increasing the use of mobile applications [12].

In this article, it is understood that non-user-friendly applications prevent the use of e-commerce by the elderly [13].

III. METHOD & IMPLEMENTATION

Technology helps to make our lives easier every day. We are able to handle our physical work quickly through technology. Social media, entertainment, banking, shopping and educational applications are actively used on mobile and internet. Young individuals generally use these applications. However, elderly people are not able to use these applications at the same frequency.

The world population is getting older. With longer life expectancy and lower fertility rate, world's population is ageing and it has gained pace in recent years. Between 2015 and 2050 the proportion of the population over 55 years of age is expected to rise from 16.88 percent to 26.87 percent, taking more than a quarter of world's population as shown in Figure 1 [14].

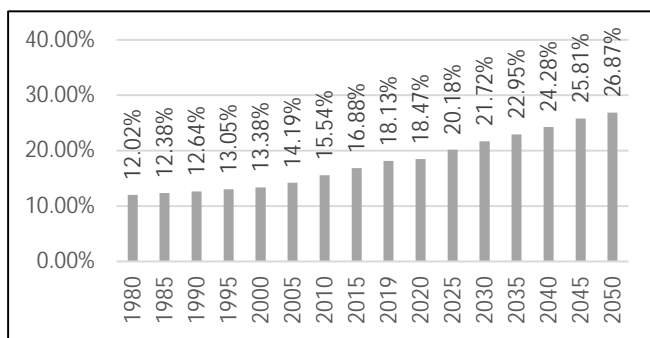


Figure 1: Percentage of Population (55+ Years) of Age

As the population of elderly continues to grow, their use of mobile banking is increasing simultaneously. Between 2011 and 2015 use of mobile banking among 60 years or older of age population increased from 5 percent to 18 percent as shown Table 1 [15]. This also indicates more and more elderly people are using mobile banking applications recently.

Table 1: Use of mobile banking in past 12 months by age

Age group	2011	2012	2013	2014	2015
18-29	45	54	63	60	67
30-44	29	37	43	54	58
45-59	12	21	25	32	34
60+	5	10	9	13	18
Total	22	29	33	39	43
Number of respondents	1859	2180	2187	2437	2151
Percent: Among those with a mobile phone and a bank account.					

In this study, a questionnaire was prepared and it was examined why individuals over 55 years of age do not use mobile and internet banking. The survey was asked the following question; what is the reason if you are not using the Internet and mobile banking application?

Answers: lack of need, complexity, lack of trust

The study focused on how to solve the complexity problem.

Except the reasons regarding security concerns and need of mobile banking the most common reason for not using mobile banking applications is the size of phone screen [15]. This reason is expected to be more common among people aged 55 years or more. As people get older physical constraints such as vision problems, hearing difficulties and reduction of mobility become evident. These constraints effect how elderly people interact with technology. For the use of mobile banking among elderly most challenging constraint is vision problems.

Table 2: Why you do not use mobile banking, (one or more answers are applicable)

My banking needs are being met without mobile banking	87.9
I don't see any reason to use mobile banking	78.3
I'm concerned about the security of mobile banking	72.7
The mobile phone screen is too small	43.1
I don't trust the technology	40.2
I don't have a smartphone	27.4
It's too difficult to use mobile banking	18.2
I don't do the banking in my household	15.3
My bank charges a fee for using mobile banking	6.3
Refused/no to all	2.7
Number of respondents	819

In his master thesis, Piotr Calak has listed vision related problems that elderly people have while using smartphones. He also made implications for software and hardware designs to overcome these problems using guidelines such as iOS Human Interface Guidelines for Apple, Android Accessibility Guidelines for Android Open Source Project, Microsoft User Experience Design Guidelines for Windows Phone and AT&T's uXd Style Guide.

Problem: Decreased Visual Acuity and Light Accommodation.

Description: People lose ability to focus near and far objects by 50 to 55 percent between ages of 8 and 50 [16]. It becomes harder to see details in low lighting conditions for elderly people.

Implications for Software Design:

This design as shown in Figure 2,

- All of the screen should be utilized while designing UI.
- Size of the text and UI elements such as buttons and icons should be expanded.
- Simple controls such as buttons or sliders for users to

adjust the size of the text can be useful.

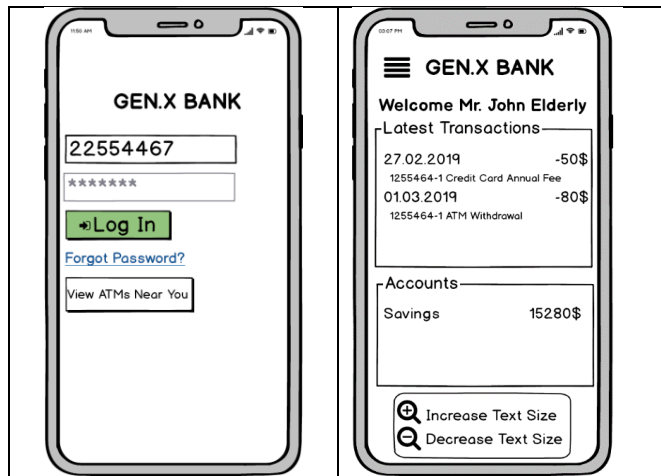


Figure 2: Mobile Banking UI Mock-Up 1

Problem: Lowered Contrast Sensitivity and Loss of Color Sensitivity

Description: Contrast sensitivity of a person diminishes between ages of 20 to 80 years [16]. The decline becomes notable after the age of 40 and diminishes even further after the age of 65 [17].

Also among older adults ability to perceive of short wavelength colors such as blues and yellows are decreased [18].

Implications for Software Design:

This design as shown in Figure 3,

- Graphics with high contrast should be used for text and icons.
- Short-wavelength colors should not be used on UI components.
- Gradients can lead to low contrast areas, so should not be used.

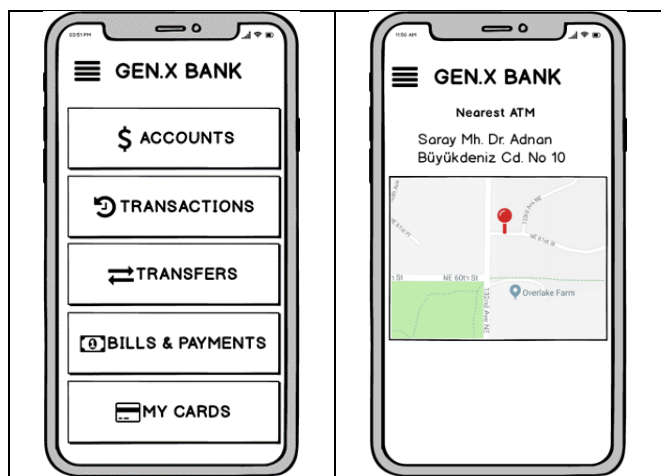


Figure 3: Mobile Banking UI Mock-Up 2

IV. CONCLUSION

Human population is getting older. In 2050, ¼ of the world population will be composed of individuals older than 55 years. When people get older, their perceptions are reduced and complex work becomes difficult. It is therefore difficult to use mobile and internet banking applications. To solve this problem, more simple and useful screens should be designed. In this study, complex design-related problems are tried to be solved. More simple interfaces, larger size buttons and text patterns are designed so that the complexity of use is aimed to be solved. The people who are older than 55 not only have problem with complexity, but also they have problem with trust. For next study, this issue can be examined. Why people whom older than 55 do not trust digital banking applications?

V. REFERENCES

- [1] Tuik, "Turkey in Statistics 2015," Turkish Statistical Institute, Ankara, 2015.
- [2] N. Bevan, "What is The Difference Between The Purpose of Usability and User Experience Evaluation Methods?," no. Interact 2009, pp. 1-4, 2009.
- [3] U. R. a. Migrants, "Ageing Population," United Nations, Vienna, 2017.
- [4] Minna Mattila, Heikki Karjaluto, Tapio Pentto, "Internet banking adoption among mature customers: early majority or laggards?," *Journal of Services Marketing*, vol. 17, no. 5, pp. 514-528, 2003.
- [5] Neha Dixit, Dr. Saroj K. Datta, "Acceptance of E-banking among Adult Customers: An Empirical Investigation in India," *Journal of Internet Banking and Commerce*, vol. 15, no. 2, pp. 1-17, 2010.
- [6] Arenas Gaitán, Jorge, Peral Peral, Begoña, Ramón Jerónimo, M^a Ángeles, "Elderly and internet banking: An application of UTAUT2," *Journal of Internet Banking and Commerce*, vol. 20, no. 1, pp. 1-23, 2015.
- [7] V. M. Kumbhar, "Customers' Demographic Profile and Satisfaction in E-Banking Services: A Study of Indian Banks," *International Journal for Business, Strategy and Management*, vol. 1, no. 1, pp. 1-9, 2011.
- [8] Pedro Cruz, Lineu Barretto Filgueiras Neto, Pablo Muñoz-Gallego, Tommi Laukkanen, "Mobile banking rollout in emerging markets: Evidence from Brazil," *International Journal of Bank Marketing*, vol. 28, no. 5, pp. 342-376, 2010.
- [9] C. W. Olphert, L. Damodaran & A. J. May, "Towards digital inclusion - engaging older people in the 'digital world'," in *Accessible Design in the Digital World Conference 2005*, Dundee, Scotland, 2005.
- [10] Muna Azuddin, Sofianiza Abd Malik, Lili Marziana Abdullah, Mumi Mahmud, "Older people and their use of mobile devices: Issues, purpose and context," in *The 5th International Conference on Information and Communication Technology for The Muslim World (ICT4M)*, Kuching, Malaysia, 2014.
- [11] Wijnand Ijsselstein, Henk Herman Nap, Karolien Poels, Yvonne De Kort, "Digital Game Design for Elderly Users," in *Future Play '07 Proceedings of the 2007 conference on Future Play*, Toronto, Canada, 2007.
- [12] Ana Correia de Barrosa, Roxanne Leitão, Jorge Ribeiro, "Design and evaluation of a mobile user interface for older adults: navigation, interaction and visual design recommendations," *Procedia Computer Science*, vol. 27, p. 369-378, 2014.
- [13] Arthur Tatnall, Jerzy Lepa, "Researching the Adoption of E-commerce and the Internet by Older People," *Logistics Information Management*, vol. 16, no. 1, pp. 56-63, 2003.
- [14] U. Nations, "World Population Prospects," 2017. [Online]. Available: <https://population.un.org/wpp/DataQuery/>. [Accessed 28 03 2019].

- [15] "Consumers and Mobile Financial," Board of Governors of the Federal Reserve System, Washington, DC, 2016.
- [16] Matthew Pattison, Alex Stedmon, "Inclusive design and human factors: designing mobile," *PsychNology Journal*, vol. 4, no. 3, pp. 267-284, 2006.
- [17] Fozard, J. L., Gordon-Salant, Handbook of the Psychology of Aging, San Diego, CA: Elsevier B.V., 2001.
- [18] P. Calak, *Smartphone Evaluation Heuristics for Older Adults*, Guelph, Ontario, Canada, 2013.

Detection of P300 ERP Waves by Different Classification Methods

A.B. AYGUN¹ and A. R. KAVSAOGLU¹

¹ Karabuk University, Karabuk/Turkey, abilalaygun@karabuk.edu.tr

¹Karabuk University, Karabuk/Turkey, kavsaoglu@karabuk.edu.tr

Abstract - The P300 speller is a type of brain computer interface that letters the desired expression to a visual screen using direct brain signals to enable people who cannot use their striped muscles. The P300 wave is generated when the letter that is intended to be printed on a matrix of characters with rows and columns flashes randomly. By detecting this wave, it is estimated which letter to focus on. In this study, the estimation and classification results of LDA, SVM, KNN, LSSVM classification methods were compared to estimate the P300 wave.

Keywords - P300 speller, brain computer interface, classification

I. INTRODUCTION

Brain computer interfaces are a system that performs the commands that are required to be performed as a result of the transfer and processing of the brain signals to the computer. With this system, a paralyzed patient can lift the arm, communicate and even perform all the daily functions in the future. In order for a brain computer interface to work, brain signals are first obtained and digitized and transferred to the computer. Then the data is preprocessed and the feature is extracted and given to the classifier. The classified data are converted into commands to perform the target operation.

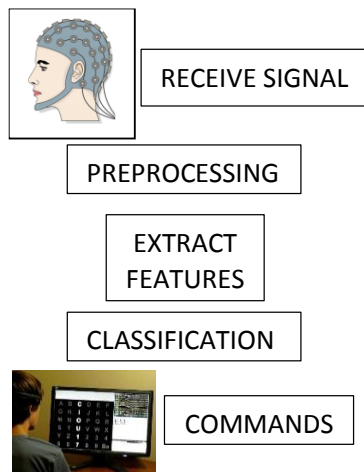


Figure 1: BCI block diagram

P300 speller was first proposed by Farwell and Donchin. In their work, they designed a matrix with 6x6 characters and commands. The rows and columns of the matrix shine and fade during the time they are determined in a random manner. In this order, the user focuses on the letter he / she wishes to express.

Randomly burning rows or columns will be stimulated at the moment the person intersects the targeted letter, and the P300 wave occurs after about 300ms after the targeted letter has shone. This signal is analyzed to determine which letter is focused[1].

II. METHODS

The following steps are required to design a P300 speller:

- The signals taken from the brain should be sampled and transferred to the computer.
- The time of each glare must be transferred to the computer when the signal is transferred to the computer.
- The signal must be separated from the next specified range by accepting the transmitted flash times. This should be done for each glare. (This is called trial.)
- Specified trials are subjected to filtering and sub-sampling.
- If desired, as previously mentioned, traits are made from trials and trained for the classifier[2]. If desired, the amplitude values of the trials used for direct sub-sampling are used as training data.
- The training data obtained from a certain number is included in the classifier algorithm determined for the training phase and the system is trained.
- Then the received signals are extracted according to the glare times as described above. According to the model of the system being trained, new signals are classified by checking whether the P300 wave is present or not. According to the result of the classification, the lettering process is done.

In this study, some classification methods on BCI Competition III data were compared both online and offline[3]. These are LDA, SVM, KNN classification methods[4]. LSSVM, a different SVM technique, was also tried[5]. After the online analyzes were done, letters were written through codes written with each class method. According to the data given in BCI Competition III Challenge 2004, the correct letter determination is 38,82% using a single channel with 85 education letters. In this study, 63 of the 100 test letters were estimated on the same datasets using 8 channels. In this study, the accuracy rates of both online and offline analyzes increased to 95%.

III. APPLICATIONS

Offline analysis is the work done with known training data on the labels after the raw data has been fully recorded. The online analysis is the kind of analysis that the person who manages the experiment pre-gave the letters only in the training set, does not have pre-test in the test set, and the letters that the system predicts as a result of the examination of the data are shown on the screen in real time. The available data was taken over 64 channels. 85 letters for the training data and 100 data for the test data. 2 different subjects recorded the signal in this way. In this study, the raw data received for 85 letters as mentioned above were drawn according to the specified 8 channels and disaggregated the data covering 667 milliseconds after each glare as a data. These electrodes are: PO8, PO7, Fz, Cz, Pz, Oz, P3, P4. Afterwards, the data obtained in size 85x180x1280 by placing the channels side by side are reduced by 10 times in a 10: 1 ratio (data size is 85x180x128). After this operation, data can be separated into 85x12x15x128 It was inserted. Each row and column are stacked in-between and the data is returned to 85x12x128 by increasing the difference between the data. Finally, the data (85x12) is converted to 1020x128 format, ie, x128. The first 75 letters (first 900x128 part) of this information obtained for 85 letters were taken as feature matrix. The feature matrix representing 75 letters was introduced into 4 different classifiers. The 120x128 dimension matrix, which represents the last 10 letters, has been tested and questioned whether the P300 wave is present in its content. Table 1 shows the accuracy, sensitivity and specificity ratios obtained from the complexity matrices extracted from 4 different types of classifiers for 10 letters.

Online analysis was followed by similar methods to offline analysis. The size of the feature matrix is 1200x128, since the so-called test data is saved for 100 letters. The 85-letter data obtained for the training data were trained for the online analysis and the 100-letter test data was tested. Accuracy, sensitivity and specificity ratios of the 100-letter data tested were determined, and each epoch was analyzed to determine which letters could be obtained from the classifiers in the code written in this study. The following are the results of the online analysis and the different letters for the A and B subjects in which letters can be printed. In addition, while obtaining the property matrix, the average of 15 trials was taken as a second method. No effect on the classification result was observed.

In the BCI Competition III Challenge 2004, a second way from 85-character training data was followed by classifying training models, and a 100-character test data obtained in BCI Competition III Challenge 2004 was put into 4 different classification processes and a performance evaluation was performed. In the second way, classifier training models are created, while the number of trials with a label value of 1 is equal to the number of trials. Table 4.6 and Table 4.7 shows the performance evaluation and the number of accurate corrected characters obtained for Subject-A and Subject-B, respectively. Figure 4.4 shows the estimated characters by using the classifier training model obtained in the second way.

A second ready-made data is the data given by Ulrich Hoffman, which has a size of 942x128x64, obtained from 64 channels[6]. In this data, 8 channel data has been taken and added side by side and 8 is sub-sampled. The most recent data

has been reduced to 942x128. Each of the 942 lines represents a trial. Of these data, 800 trial chapters were trained, and 142 were estimated to include P300 waves. Below are the accuracy, sensitivity and specificity ratios from the complexity matrices extracted from 4 different types of classifiers for 142 trials.

IV. CONCLUSION

All results obtained at the end of these applications are given in the tables below. Table information is given in the table descriptions.

Table 1: The results of offline analysis for BCI Competition III Challenge 2004 Subject-A and classifier comparisons

SUB-A	ACCUR	SENS	SPEC
LDA	90,83	65,0	96,0
KNN	82,50	10,0	97,0
SVM	88,33	35,0	99,0
LSSVM	91,67	55,0	99,0

Table 2: The results of offline analysis for BCI Competition III Challenge 2004 Subject-B and classifier comparisons

SUB-B	ACCUR	SENS	SPEC
LDA	95,83	95,0	96,0
KNN	83,33	15,0	97,0
SVM	90,83	45,0	100
LSSVM	98,33	95,0	99,0

Table 3: Model equal label data (Subject-A)

SUB-A	ACCUR	SENS	SPEC	TRUE LETTERS NUM
LDA	92,92	66,5	98,20	45
KNN	85,67	23,0	98,20	9
SVM	91,08	49,50	99,40	27
LSSVM	92,0	55,0	99,40	33

Table 4: Model equal label data (Subject-B)

SUB-B	ACCUR	SENS	SPEC	TRUE LETTERS NUM
LDA	95,08	85,0	97,1	63
KNN	90	41,5	99,7	6
SVM	84,33	17,5	97,7	23
LSSVM	98,33	95,0	99,0	63

[illegible]

Figure 2: The character result of the model with different number of label data

SUB-A	ACCUR	SENS	SPEC	TRUE LETTERS NUM
LDA	86,83	76,5	88,9	34
KNN	70	75	69	3
SVM	83,33	78	84	16
LSSVM	87,33	74,5	89,9	34

Table 6: Model without equal label data (Subject-B)

SUB-B	ACCUR	SENS	SPEC	TRUE LETTERS NUM
LDA	88,33	83,5	89,9	31
KNN	65,42	65,42	64,4	2
SVM	76,75	89	74,3	12
LSSVM	86,33	87,5	86,1	29

[illegible]

Figure 3: The character result of the model with the same number of label data

Table 7: Ready data from Hoffman were analyzed and compared by classifiers

HOFFMAN	ACCUR	SENS	SPEC
LDA	76	78	70
KNN	64	70	56
SVM	85	78	70
LSSVM	85	95	64

V. COMMENTS

The reason for the low sensitivity in the ready data is that the P300 wave is known to be 6 times less when the line and column are known to be non-existent. However, the sensitivity of LDA and LSSVM classifiers was high. While the offline analysis was performed, the first 75 letters were trained, and then 10 letters were estimated to contain P300. Accuracy, sensitivity and specificity values of the estimated 10 letters Table 1 and Table 2. An offline analysis was performed so that it was close to working online. This method provides more realistic results when letter extraction is performed.

The accuracy, precision and specificity are of course proportional to the letter obtained, but the letters obtained are

known to be printed from both rows and columns in the case of P300. P300 existed in both lines and columns. The higher the accuracy, the higher the accuracy of the P300 in both the row and the column, unless the information is available, the letter extraction can not be done correctly. In addition, in LDA and LSSVM classifier data, high letter extraction and accuracy rates were obtained in both A and B subjects.

Performance evaluations and character estimation of online analyzes using classifier training methods obtained by synchronizing data with different label values using data from BCI Competition III Challenge 2004 are shown in Tables 3 and 4. According to the data obtained by this method, LDA and LSSVM methods are seen to be more advantageous in the classification of the P300 wave. It is seen in Table 3 and 4 that the success rate of the KNN method in the classification of the P300 wave is low compared to other classification methods. The SVM method failed with LDA and LSSVM in correct character estimation. The low success of the SVM in character estimation can be attributed to the low accuracy of the accuracy and specificity ratio, but also because of the fact that the row and column information is not included when evaluating the result of the other classifier training model.

If we compare these two classifier training methods by looking at Tables 3, 4, 5 and 6, the accuracy and specificity ratios of Table 5 and 6 in Table 3 and 4 are obtained by using the method of obtaining the classifier training model where the data with different label values are not equalized. It is concluded that all classifier methods are high according to the accuracy and specificity obtained by the classifier training model where the data with different label values are equalized. However, it is seen that the sensitivity ratios obtained by the KNN and SVM methods of the classifier training model where the data with different label values are not equalized are much lower compared to the sensitivity ratios obtained with the model with different label values. In addition, it is seen that the correct character estimation numbers made with the model with different label values in Tables 3 and 4 are not equal, compared to the correct character estimation numbers made with the model with different label values in Tables 5 and 6. According to this comparison, it is concluded that the classifier training model where data with different label values are not equalized is more advantageous than the classifier training model where the data with different label values are equalized.

ACKNOWLEDGEMENT

This study was supported by The Karabuk University Scientific Researches Projects (Project No: KBUBAP-18-YL-074) research grant. We are indebted for its financial support.

REFERENCES

- [1] L. A. Farwell and E. Donchin, "Talking off the top of your head : toward a mental prosthesis utilizing event-related brain potentials," pp. 510–523, 1988.
- [2] V. Abootalebi, M. H. Moradi, and M. A. Khalilzadeh, "A new approach for EEG feature extraction in P300-based lie detection," *Comput. Methods Programs Biomed.*, vol. 94, no. 1, pp. 48–57, 2009.

- 2009.
- [3] B. Blankertz and G. Schalk, "Documentation Wadsworth BCI Dataset (P300 Evoked Potentials) Data Acquired Using BCI2000's P3 Speller Paradigm BCI Competition III Challenge 2004," pp. 1–8, 2004.
 - [4] D. J. Krusienski, E. W. Sellers, S. Bayoudh, D. J. Mcfarland, T. M. Vaughan, and J. R. Wolpaw, "A comparison of classification techniques for the P300 Speller," vol. 299.
 - [5] N. Gedik, "Least squares support vector mechanics to predict the stability number of rubble-mound breakwaters," *Water (Switzerland)*, vol. 10, no. 10, 2018.
 - [6] U. Hoffmann, G. Garcia, J. M. Vesin, K. Diserenst, and T. Ebrahimi, "A boosting approach to P300 detection with application to brain-computer interfaces," *2nd Int. IEEE EMBS Conf. Neural Eng.*, vol. 2005, pp. 97–100, 2005.

Investigating Personalised Applications in MOOCs: the challenge of achieving transparency

A. S. SUNAR¹ and S. WHITE²

¹ Bitlis Eren University, Bitlis/Turkey, assunar@beu.edu.tr

² University of Southampton, Southampton/UK, saw@ecs.soton.ac.uk

Abstract - Massive Open Online Courses (MOOCs) offer online courses, in which participants with widely diverse cultural and educational backgrounds can study. These participants may have very different goals and expectations of their courses. Unfortunately, MOOCs do not always meet the needs and expectations of participants. In response, critiquing the static design of MOOCs has triggered discussions amongst researchers considering personalising MOOC experiences. Initial solutions attempted to create adaptive MOOC environment by methods such as implementing recommender systems, personalising feedback and gamification. This study investigates the affordances of recent personalisation applications in MOOCs and discusses the results from the perspective of transparency and possible future improvements. The findings demonstrate that the most popular addition is recommender systems to suggest personalised learning resources and/or navigational paths through the course. Additionally, the outcomes demonstrate that personalising MOOCs alone for the study experience is not sufficient, since it does not glean enough information from the participants. Problematically, integrating external tools to acquire more data may violate privacy even though it may contribute more transparency to the personalisation process.

Keywords - Personalisation, MOOCs, Recommender Systems, Adaptive Learning, Distance Learning

I. INTRODUCTION

MOOCs emerged as a new practice of distance online learning in 2008. The emergence of MOOCs brings many considerations to the attention of researchers working on educational and technological issues. Not so long after the release of the first MOOCs adopting different pedagogies in 2008¹ and 2011², the discussion around the idea of personalised MOOCs has raised [1].

Researchers concerning about high dropout rates commonly observed in MOOCs proposed that a personalised MOOC experience could decrease dropout rates. Researchers investigating possible reasons for dropout highlighted some key factors: lack of motivation, poor time management skills, feeling loss in overloaded course resources, and lack of self confidence to become involved in conversations. Assorted types of personalisation applications were developed for

addressing these issues. Example responses include integrating a feedback system to maintain learners' motivation and linking external tools to help learners plan their schedule and goals.

Even though researchers may agree on the necessity of personalisation in MOOCs, it is arguable that how and to what extent personalised applications will be applied to MOOCs. Bonk et al. [2] investigated the MOOC instructors' attitude towards personalised MOOCs. The study shows the analysis of a questionnaire completed by 152 MOOC instructors from around the world. They found that one-third of the instructors put effort on meeting learners' unique needs during the course although fewer respondents were concerned about it during the course delivery. These personalisation efforts could include preparing a brief video each week to address questions mentioned in discussions or allowing the participants to choose their own assignments rather than a sophisticated personalised application. In addition, the majority of them reported interest in personalisation for their next MOOC delivery.

Our research conducted in 2015 investigated the personalisation attempts in the early years of MOOCs [1]. The study shows that MOOC researchers have extensively discussed about the needs and applications of personalised MOOCs. While the literature on either the needs or proposals for a personalised application was on sparse at that time, many completed research on personalised applications in MOOCs are available today. Subsequent growth in this area is sufficient to warrant further investigation.

In this new study, we systematically investigated published research papers in the years between 2016 and 2018 in order to answer the question how personalisation is recently handled in MOOCs. We identified six journal articles and 20 conference papers in the first 10 pages retrieved through a simple Google Scholar search with the keyword 'personalised MOOCs'. This study answers the following questions:

- What affordances are provided by MOOCs for a personalised MOOC experience?
- What are the challenges for transparent and privacy protected MOOC personalisation?
- What improvements could be applied for MOOCs?

¹ <https://sites.google.com/site/themoocguide/3-ck08—the-distributed-course>

² <http://www.robotics.stanford.edu/~ang/papers/mooc14-OriginsOfModernMOOC.pdf>

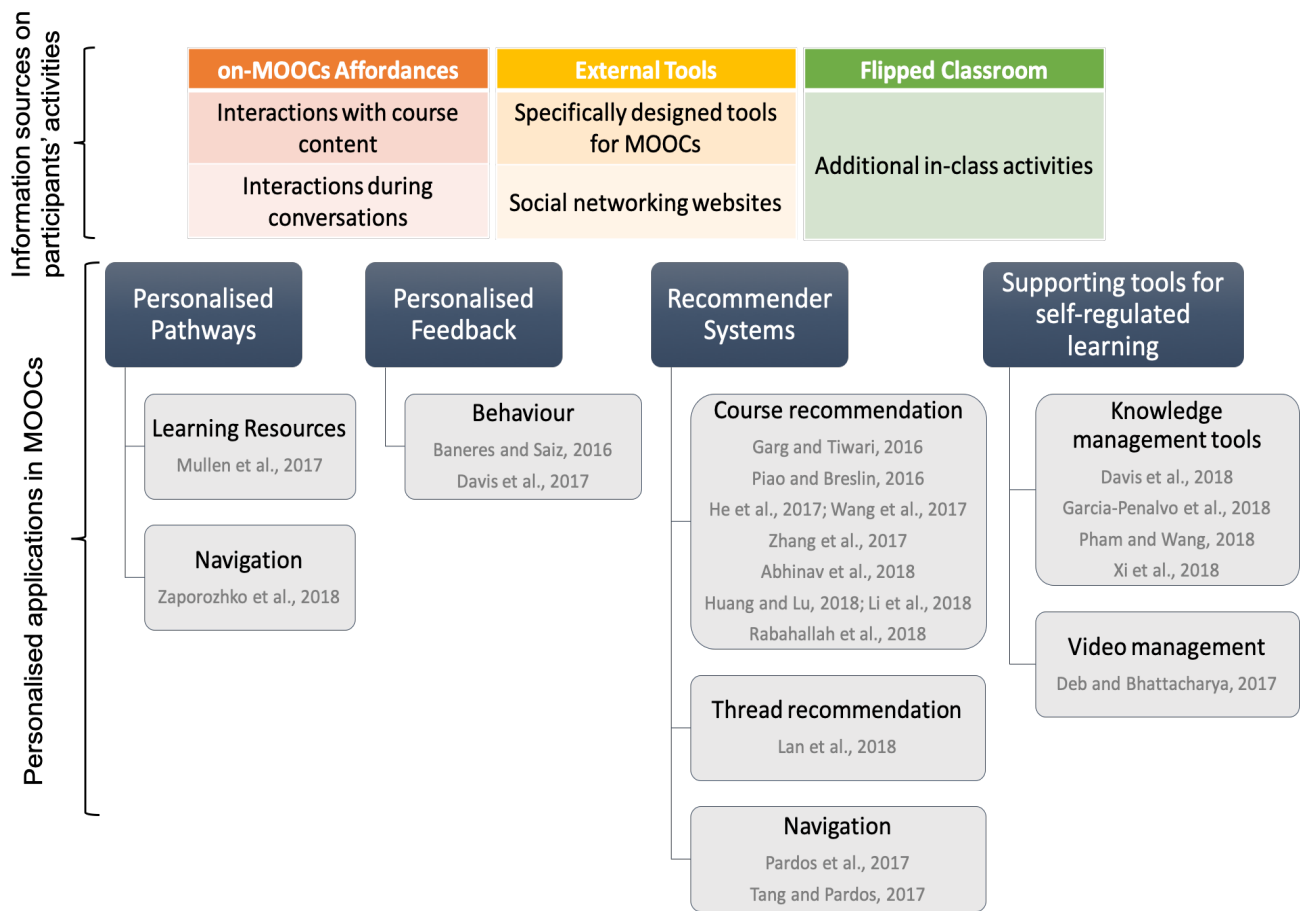


Figure 1: Categorisation of personalised MOOC applications.

II. ANALYSIS

MOOCs glean information about the participants' behaviours from their interaction with the course, for example, when they opened a page or how many assignments they completed. Other methods of tracking participants' online activities, involve following digital footprints. For example, some MOOCs promotes the use of external learning management tools or social networking websites so that they may have additional log data of the participants, activities. Additionally, some instructors use MOOCs in their on campus classroom as flipped classrooms ([3], [4]). They have direct observation of their students along with their MOOC activities.

Figure 1 shows the categories of resources for tracking the participants to understand their behaviours.

Researchers have used these data sources to develop personalised applications. The commonly applied personalised applications are also demonstrated in Figure 1.

A. Recommender System

Among the samples, recommender systems are the most applied method, falling into three types.

Course recommendation. Recommender systems are mostly developed to generate a selection of courses for those searching a suitable MOOC to study. Information on people's

preferences are extracted from different sources. For example, to achieve the most suitable course recommendation, Piao and Breslin [5] used the personal profiles on LinkedIn using their job, skills, and education information.

Garg and Tiwari [6]; Zhang et al. [7]; He et al. [8]; Li et al. [9]; Abhinav et al. [10]; Huang and Lu [11]; Wang et al. [12]; Rabahallah et al. [13] use course and user attributes collected through the MOOC platform.

Discussion thread recommendation. Thread recommendation is one of the most highly developed applications for MOOCs in years. Our previous study investigating personalisation studies in the first years of MOOCs, show that researchers have used sentiment analysis and text mining along with other attributes to present the most related conversations to participants [1]. Even though the number of applications with thread recommendation has fallen, it is still developed to meet the needs of participants who want to find the appropriate conversations for getting feedback or interact with others. For example, Lan et al. [14] develop an algorithm to rank the threads for each learner to read.

Navigation recommendation. Tang and Pardos [15]; Pardos et al. [16] offer a method for navigating the participants to the most suitable page in the course based on their previous behaviour including cognitive state and preferences.

B. Personalised Pathway

One of the solutions to address the static design of course content in MOOCs is via dynamic presentation of the previously prepared content, especially personalised display for each learner. In order to decide which content to present whom, each study defines different type of learners and type of content. Mullen et al. [17] provide personalised learning pathways according to the professional level of the students. Zaporozhko et al. [18] also determine the type of learners according to their dominant learning approach and the type of learning objectives in each learning content and then personalise the learning path for each individual.

C. Personalised Feedback

Generating personalised feedback on participants' behaviour, course progress, and assignments are one of the assistive personalised tools for helping the participants along the course. Baneres and Saiz [19] create an intelligent tutoring system which generates personal feedbacks on learning experience of the students.

Davis et al. [20] build a personalised feedback system on the social comparison theory. The system provides feedback on the student's performance compared to the successful participants' accomplishments to encourage them.

D. Supporting Tools for Self-Regulated Learning

Since available affordances on the platforms for personalising MOOC experience are very limited, many MOOC instructors provide an external platform/application where their participants manage their achievements and plans. These systems may provide feedback and recommendations to help learners in time-consuming tasks such as re-watching videos or time management of future tasks.

For example, Davis et al. [21] propose an external knowledge management system for helping learners maintain their motivation and keep them on track by allowing them plan the study and achievements. Garcia-Penalvo et al. [22] propose an adaptive hybrid MOOC where participants manage their choices and community while studying in a course. Similarly, Xi et al. [23] propose a system where a recommendation of the next course page to study is presented and allow participants to modify their learning paths.

Deb and Bhattacharya [24] propose a system for helping learners cut the necessary part of videos and take notes on them. This system again aims to speed the re-watching of a video for finding an important part of it or repeat a lecture.

Differently from the commonly developed systems, Pham and Wang [25] optimised a mobile device app for MOOCs. Their system uses clickstream data along with emotional signals such as facial expression and heart rate with the help of the cameras on the phone.

participants. Dalsgaard and Paulsen [26] describe it as a central aspect of cooperative learning which enables students to make use of each other while at the same time maintaining individual freedom. The authors point out that participants do not always share things in a cooperative learning environment but instead they may update their personal pages on social networking sites. The authors assert that personal pages/profiles on social networking sites enable individual flexibility and transparency in communication with others.

Most frequently the constructivist and connectivist MOOCs promote the use of other social platforms for knowledge exchange amongst students. For example, Maggio et al. [27] encourage their participants to create and share resources so that everyone would have multiple choices of paths through the course. The authors claim that personalisation is possible by providing multiple pathways without applying predictive analysis and guiding the participants to the personalised paths. However, providing so many opportunities without guidance may cause the participants to again feel lost in the many opportunities they may have outside of the MOOC platform.

Even if the MOOC providers would like to use the activities outside of the platform, the issue of privacy comes up to the table. The MOOCs providers express concern about the participants' privacy and collect the data anonymously on the platform. In case that the participants use other social networking websites, the MOOC platforms do not always give permissions to link participants' information in the platform with other external tools. Therefore, the MOOC platforms would not be able to provide a personalised service for the participants in this use case. He et al. [8] highlight this situation by describing MOOCs as information isolated islands as there are very limited personal information sources about participants especially if the person is not very socially active on the platform. Therefore, a paradox between transparency and privacy for personalisation occurs.

The current MOOC platforms provide simple personalised affordances such as adjusting the speed of video or subtitles and transcripts in many languages for all the courses on the platform. The additional personalised applications such as the ones presented in this paper are usually for the certain MOOCs where the instructors of the course put effort to personalise the study in their course(s). Therefore, despite the huge amount of statistically proven successful attempts for personalised MOOC applications, the use of them is not widespread on the platforms. The reason behind this could be the challenges associated privacy or the financial burden to apply such affordances on the platforms. However, if the platforms desire to provide a personalised environment by maintaining transparency and privacy, they need to invest in personalised applications and enrich the variation of tools on the platforms.

III. DISCUSSIONS

In MOOCs, participants study as individuals, which is one of the advantages of their flexible environment. Even though participants share their opinions and questions in discussion forums, this is not as a part of a collaborative study with other

IV. CONCLUSION

The study reported in this paper investigated the current personalised applications applied in MOOCs in the years between 2016 and 2018. The study especially investigated the

challenges of achieving transparency while developing personalised applications. The findings show that course recommendation systems are the most applied personalisation method. External knowledge management systems for supporting self-regulated learning in MOOCs are the second commonly developed systems. The results show that the social networking tools outside of the MOOC platforms provide transparency at the same time as potentially undermining the personal privacy of MOOC participants.

REFERENCES

- [1] A. S. Sunar, N. A. Abdullah, S. White, and H. C. Davis. "Personalisation of MOOCs: The state of the art." In *Proceedings of the 7th International Conference on Computer Supported Education-Volume 1*, pages 88–97. SCITEPRESS-Science and Technology Publications, Lda. 2015.
- [2] C. J. Bonk, M. Zhu, M. Kim, S. Xu, N. Sabir, and A. R. Sari. "Pushing toward a more personalized MOOC: Exploring instructor selected activities, resources, and technologies for MOOC design and implementation." *International Review of Research in Open and Distributed Learning* 19, no. 4, 2018.
- [3] S. M. J. Azizul and R. Din. "Teaching and Learning Mathematics on Geometry using Geogebra Software via MOOC." *Journal of Personalized Learning* 2, no. 1, 40-51, 2018.
- [4] L. Pingping, W. Chao, and L. Bailin. "Research on Resource Sharing Mechanism of MOOC Based on Big Data." In *2017 International Conference on Social science, Education and Humanities Research (ICSEHR 2017)*. Atlantis Press, 2017.
- [5] G. Piao, and J. G. Breslin. "Analyzing MOOC entries of professionals on LinkedIn for user modeling and personalized MOOC recommendations." In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, pp. 291-292. ACM, 2016.
- [6] V. Garg, and R. Tiwari. "Hybrid Massive Open Online Course (MOOC) recommendation system using machine learning." 11-5, 2016.
- [7] H. Zhang, H. Yang, T. Huang, and G. Zhan. "DBNCF: Personalized courses recommendation system based on DBN in MOOC environment." In *2017 International Symposium on Educational Technology (ISET)*, pp. 106-108. IEEE, 2017.
- [8] X. He, P. Liu, and W. Zhang. "Design and Implementation of a Unified Mooc Recommendation System for Social Work Major: Experiences and Lessons." In *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, vol. 1, pp. 219-223. IEEE, 2017.
- [9] X. Li, T. Wang, H. Wang, and J. Tang. "Understanding User Interests Acquisition in Personalized Online Course Recommendation." In *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*, pp. 230-242. Springer, Cham, 2018.
- [10] K. Abhinav, V. Subramanian, A. Dubey, P. Bhat, and A. D. Venkat. "LeCoRe: A Framework for Modeling Learner's preference." In *Educational Data Mining Conference*, 2018.
- [11] R. Huang, and R. Lu. "Research on Content-based MOOC Recommender Model." In *2018 5th International Conference on Systems and Informatics (ICSAI)*, pp. 676-681. IEEE, 2018.
- [12] Y. Wang, B. Liang, W. Ji, S. Wang, and Y. Chen. "An improved algorithm for personalized recommendation on MOOCs." *International Journal of Crowd Science* 1, no. 3, 186-196, 2017.
- [13] K. Rabahallah, L. Mahdaoui, and F. Azouaou. "MOOCs Recommender System using Ontology and Memory-based Collaborative Filtering." In *ICEIS (I)*, pp. 635-641. 2018.
- [14] A. S. Lan, J. C. Spencer, Z. Chen, C. G. Brinton, and M. Chiang. "Personalized thread recommendation for MOOC discussion forums." In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 725-740. Springer, Cham, 2018.
- [15] S. Tang, and Z. A. Pardos. "Personalized Behavior Recommendation: A case study of applicability to 13 courses on edX." In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, pp. 165-170. ACM, 2017.
- [16] Z. A. Pardos, S. Tang, D. Davis, and C. V. Le. "Enabling real-time adaptivity in MOOCs with a personalized next-step recommendation framework." In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*, pp. 23-32. ACM, 2017.
- [17] J. Mullen, C. Byun, V. Gadepally, S. Samsi, A. t Reuther, and J. Kepner. "Learning by doing, High Performance Computing education in the MOOC era." *Journal of Parallel and Distributed Computing*, 105-115, 2107.
- [18] V. Zaporozhko, I. P. Bolodurina, and D. I. Parfenov. "A genetic-algorithm approach for forming individual educational trajectories for listeners of online courses." In *Proceedings of Russian Federation & Europe Multidisciplinary Symposium on Computer Science and ICT*. 2018.
- [19] D. Baneres, J. Saiz. "Intelligent tutoring system for learning digital systems on MOOC environments." In *2016 10th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)*, pp. 95-102. IEEE, 2016.
- [20] D. Davis, J. Jivet, R. F. Kizilcec, G. Chen, C. Hauff, and G. J. Houben. "Follow the successful crowd: raising MOOC completion rates through social comparison at scale." In *Proceedings of the seventh international learning analytics & knowledge conference*, pp. 454-463. ACM, 2017.
- [21] D. Davis, V. Triglianios, C. Hauff, and G. J. Houben. "SRLx: A Personalized Learner Interface for MOOCs." In *European Conference on Technology Enhanced Learning*, pp. 122-135. Springer, Cham, 2018.
- [22] F. J. Garcia-Peñalvo, A. I. Fidalgo-Blanco, and M. L. Sein-Echaluce. "An adaptive hybrid MOOC model: Disrupting the MOOC concept in higher education." *Telematics and Informatics* 35, no. 4, 1018-1030, 2018.
- [23] J. Xi, Y. Chen, and G. Wang. "Design of a Personalized Massive Open Online Course Platform." *International Journal of Emerging Technologies in Learning (iJET)* 13, no. 04, 58-70, 2018.
- [24] S. Deb, P. Bhattacharya. "Enhancing Personalized Learning with Interactive Note Taking on Video Lectures—An Analysis of Effective HCI Design on Learning Outcome." In *International Conference on Next Generation Computing Technologies*, pp. 558-567. Springer, Singapore, 2017.
- [25] P. Pham, and J. Wang. "Adaptive Review for Mobile MOOC Learning via Multimodal Physiological Signal Sensing-A Longitudinal Study." In *Proceedings of the 2018 on International Conference on Multimodal Interaction*, pp. 63-72, ACM, 2018.
- [26] C. Dalsgaard, and M. Flate Paulsen. "Transparency in cooperative online education." *The International Review of Research in Open and Distributed Learning* 10, no. 3, 2009.
- [27] L. Maggio, A. Saltarelli, and K. Stranack. "Crowdsourcing the curriculum: A MOOC for personalized, connected learning." *EDUCAUSE Review*, 2016.

Unification of IT Process Models into a Simpler Framework

B.AYGUN^{1,2}

¹Middle East Technical University, Ankara/Turkey, betularac@gmail.com

²Huawei Turkey R&D Center, Istanbul/Turkey, betul.aygun@huawei.com

Abstract - Information technology usage has become compulsory for all organizations whether government or private organizations to achieve visibility, compete rivals and execute their missions better. To get desired result from usage of information technology, IT of organization has to be managed well. Up till now, various frameworks are developed to manage it well. Best examples for this kind of frameworks are COBIT and ITIL, containing all processes which can be handled in IT management and becoming widespread through the world. COBIT and ITIL are complementary frameworks rather than competitors. Due to this reason, organizations must implement both of them instead of choosing one of them. In addition to these, ISO/IEC 27001 which focuses on information security management process is a quite famous IT standard in terms of security. This study provides organizations to meet requirements of these which are process based frameworks and standards complementary to each other, with a unique implementation by taking unification of processes in a more simple and understandable way. Consequently, it provides reduction in duplicate work and prevents inconsistencies that may occur. Also, if an organization applies one of them, then by looking at the mapping between them, the deficiencies of the other governance tools can be easily monitored. In addition, including CMMI level two requirements motivate the organization to implement higher maturity level of CMMI. Moreover, this study provides organizations to implement ISO 27001 management structure which establish a foundation for extension to technical structure of it. Besides these, this study provides an alignment of the proposed model with COBIT and ITIL which helps organization to trace ITIL and COBIT simultaneously.

Keywords - IT governance, IT service management, IT Frameworks/Standards, IT processes.

I. INTRODUCTION

STRUCTURING and managing IT activities has become one of the main focus of the organizations due to the fact that information technologies strategically differentiate organizations [1]. To be successful in business, IT infrastructure should be effectively managed in the organization because of the fact that “IT is now at the core of most organizations” to execute strategy [2]. The practice of IT governance has emerged as a discipline to enable organizations better managing IT assets and so investments. In addition to this, to enable effective IT management, a great

number of IT governance standards/frameworks are emerged, including ITIL (Information Technology Infrastructure Library), COBIT (Control Objectives for Information and Related Technology), ASL (Application Services Library), Six Sigma, CMM/CMMI, IT Service CMM, SAS70, ISO 17799, SOX, SysTrust and PRINCE2 [3]. These frameworks cover different aspects of IT and not a single one is adequate to provide effective IT Governance.

Organizations applying IT governance discipline earn at least 20% higher return on assets than those with weak IT Governance practice [4]. Moreover, based on the experience on five Australian organizations, implementing ITIL means the transforming IT service management to gather vital benefits such as more control on testing and changes, more predictable infrastructure, improved consultation within organization, reduced server failures, documented and consistent IT service management processes and compatible recording of incidents [5]. Thus, IT governance is essential for organizations independent of the size. However, effectively governing IT is not as easy as considered. To manage IT as a hole efficiently and effectively, different models provided should be applied together [2]. There is high increasing not only in the usage of standards/frameworks individually but also in implementing several framework/standards simultaneously [6]. However, implementing different models into an organization is considerable difficult and requires more expertise. Since, many of the existing frameworks in IT Governance are supplementary and each framework has focused on different areas, a mix-and-match approach is often taken by organizations that aim is to govern IT effectively [2]. But mix-and-match approach for organizations is more difficult than considered. Hayden stated that implementing multiple frameworks separately needs more cost and effort in the organization’s compliance initiatives due to redundancy duplication derived from similarities between various frameworks [7].

Eventually, organizations realized that IT governance frameworks are necessary for their success, but now another new problem which is that which frameworks should be considered as essential arises. According to Hardy, unification of IT governance frameworks should include COBIT, ITIL and ISO 27001 which have become widely popular around the world [8]. Besides, Solms underlined the benefits of integrated application of COBIT, which provides wider reference and

integrated platform, and ISO 17799, which provides more detailed guideline about information security [9]. In addition to these ideas, CMMI, having very high level mature and well known framework in Turkey, is also complementary to COBIT and ITIL [10]. Due to their popularity around the world and their affirmative consequences for the organization, COBIT, ITIL, CMMI and ISO 27001 frameworks are covered in this study.

The rest of this paper is organized as follows; Section 2 describes the literature review about the subject. In Section 3, the unification of the main standard/frameworks is explained. After all, only two main processes which are Risk Management and Information Security management are explained in detail and the alignment of the processes defined with used standard/frameworks are given. Section 4 includes the justification of the proposed solution in a very well-known organization in Turkey. Section 5 explains the conclusion of this study and includes future work for it.

II. LITERATURE REVIEW

Although, academic society and IT organizations realized that most of frameworks especially COBIT and ITIL should be implemented together, there is no clear study on that subject producing a new framework containing both of their requirements totally. Firstly, ISACA produces a report based on aligning COBIT and ITIL but does not produce new processes. It just only shows the relationship between the ITIL sections and COBIT control objectives. But this document is quite important document referenced for designing processes during this study. There is an another study which produce a new framework for only one process "Information security management" by unifying ITIL Security management and COBIT DS5 (Ensure System Security) Control objective [18]. Most of the organizations especially the consulting firms and software organizations such as IBM Tivoli, HP ITSM and Alcyone Consulting Firm, produce their own processes based on ITIL and COBIT, but not totally depends on these standards and they are not for academic study. For this reason, they do not need to be explained further in this chapter.

In the study of Bounaguiet al., they propose a new framework for Cloud Computing (CC) governance that combines endorsed IT models; ITIL, COBIT and ISO 27002 [19]. In this study, governance for cloud computing services instead of IT services is handled. Requested activities are defined and aligned with COBIT, ITIL and ISO 27001. The proposed governance framework includes only the small part of the processes of Service Design of ITIL. On the contrary, in our study, all processes necessary for IT governance are unified.

Moreover, the study of Mourad et al. also combines the two popular IT governance standards, COBIT and ITIL [20]. First of all, they just align the control objectives in COBIT to the five pillars of ITIL. Also, they did not consider CMMI and ISO 27001. In the study of Nurlistianiet al. also combines the COBIT with Service Design and Service Operation processes

of ITIL for the audit of information system [21]. In the study of Paredes-Gualtor et al., they compare the Enterprise Architecture (EA) with COBIT and ITIL by using some most important processes of them [22]. However like in other studies, they do not handle all the contents of these standards, they only take some small part. Although, they all believe that applying one them is not enough for better governance, there is not study unifying these approaches completely.

III. METHODOLOGY

Implementing, managing and supporting IT processes and services effectively results in more success, less downtime hours, less failures, excessive revenue, scarce cost, good communication and realization of business objectives [23]. According to managers, IT service management and IT governance frameworks are not separated, they should be combined to provide powerful IT governance and best practice for service management [6], [17], [24]. Recently, organizations adopt multiple frameworks COBIT and ITIL at the same time [6]. However, implementation of all these frameworks results in important issues which are interrelationship and process overlaps; for example, configuration management is handled by both CMMI and ITIL. Thus, overall plan should be adopted rather than separate plan for each framework process adoption [6]. While defining the processes during this study, ITIL is based and COBIT processes are added over them. Because, ITIL is strong in processes and explains the process flows (i.e. the way of doing job) although COBIT is composed of processes, it does not contain steps and tasks to realize processes [25]. So, COBIT focuses on what the organization should do but do not explain how it should be done. On the contrary, ITIL defines how processes are achieved by giving flow charts [26]. As a result, in this study, processes are based on ITIL since implementation of ITIL is easier than the COBIT and ITIL mentions "how" instead of what. In addition to this, the processes in this study are enriched by COBIT's control objectives since COBIT contains wide set of resources including all the information that organizations need while adopting IT governance and control framework [27]. Since security has become the important process that has to be handled, ISO/IEC 27001 management document is covered to define information security management process flow with the ITIL and COBIT. Management structure of ISO 27001 are Plan (establishing), Do (Implementing and operating), Check (Monitoring and Reviewing) and Act (Maintaining and improving). These four structures are taken into consideration in information security management process as in ITIL and ISO 27001. While forming processes, ITIL V3 and COBIT are considered basically by taking unification of their processes. During taking unification of ITIL processes with COBIT processes, the document "Mapping of ITIL V3 with COBIT" produced by ISACA is based especially to detect the

overlapping of processes and intersection of them [28]. This document includes the alignment of COBIT- ITIL and ITIL-COBIT bidirectional. Not all processes in this study are only emerged from COBIT and ITIL. There is an exception case in Knowledge Management process. Knowledge management process is divided into three part; Knowledge Management Strategy, Knowledge Transfer; and Data and Information Management [29]. To minimize the number of processes and functions and to prevent complexity, although data and information management is considered separated process in COBIT and a separated function in ITIL; in the study's scope, data and information management is handled in Knowledge Management process. The detail of the knowledge management process and the other processes that cannot be included in this study are explained in the thesis [30].

The steps and structure of defining processes prepared in the scope of this study is suitable with ITIL processes. Each process contains definition, purpose, scope, business value, activities & tasks (if applicable, most of them are figured out as flow charts), inputs/outputs, roles and metrics.

There exist twenty five processes which are explained within three main parts;

- Principles
- General Information
- Flow Chart

All processes defined in the scope of this study are shown in the following figure which states that which processes are based on which framework/standards. Each color shows the different combination of frameworks¹.

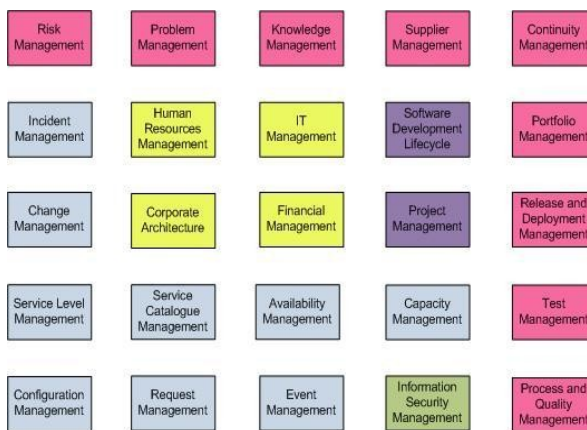


Figure 1: Processes based on which frameworks/standards.

Since, the processes in the scope of this study is prepared based on the process definition in ITIL books, some set of functions like service desk or application management is not stated as processes like ITIL. The characteristic of processes stated in ITIL Service Operation book should be measurable, specific results, customers and responds to specific events. Since there is no chance to count results of service desk like “how many Service Desks were completed?” service desk cannot be thought as process on the contrary to COBIT [31],

[32]. For this reason, four functions which are Service desk, application management, technical management and operational management are defined during this study.

Due to the fact that explaining all of the processes and functions in this paper is not feasible, only two popular processes, risk management and information security management, are detailed in this paper as an illustration.

A. Risk Management Process

1) Principles

1. Risk management framework contains acceptable level for risks, documenting residual risks and pull residual risks to an acceptable level by defining mitigation strategies.

2. Risks are recorded and maintained.

3. Probability of occurring risks is assessed by quantitative and qualitative methods.

4. Risk Mitigation Methods;

- Getting addition UPS
- Set up a Fault tolerant systems
- Holding alternative supplies
- Comprehensive backup system

5. Risk strategies are defined as the following;

- Avoidance
- Reduction
- Sharing
- Acceptance

6. Cost, benefit and responsible of risk remediation plans are defined.

7. After implementing risk remediation plans, any variance is reported to executive manager.

2) General Information

Table 1: The General Information about Risk Management Process

Process ID	P.6
Process Name	Risk Management
Sub Processes Name	Two phases: Risk Analyze and Risk Management
Purpose	Risks are evaluated and potential benefits are analyzed. For each action, risks and opportunities are defined and suitable reactions are identified.
Scope	Ensuring that risk management is integrated into management process and risk management is implemented continuously and consistently Making risk evaluation Proposing and delivering risk remediation plans
Value to Business	Analyzing and announcing IT potential risks effect into the business processes and goals
Inputs	Security threats and gaps Supplier risks Project risk management plan IT strategic and tactical plans History risk trends Contingency Test Results

¹Blue: ITIL based , Green: COBIT, ITIL and ISO 27001 based, Purple: COBIT and CMMI based
Red: ITIL and COBIT based , Yellow: COBIT based

Outputs	Risk Assessment Risk Reporting IT Risk Management Guideline IT Risk Remediation Plan
Roles	Risk Manager
KPI	Percent of IT conditions assessed Percent of IT risks defined newly Number of incidents results from risk defined management process Percent of critic IT risks developed remediation plans against

The flow chart of risk management is given in Figure 2. The inputs, outputs and the process are detailed in this chart. The left part of the chart is mainly focuses on the risk management. On the right part, the steps necessary for the risk analysis are given. The outputs of the risk analysis are the crucial inputs for the risk management part.

3) Mapping Framework with ITIL

In the Table 2, alignment with the ITIL sections and COBIT control objectives of the Risk management process is detailed.

Table 2: Alignment with ITIL and COBIT

ITIL	COBIT
SS 9.5 Risks	PO9 Assess and Manage IT Risks
SD 8.3 Risks to the services and processes	AII.2 Risk Analysis Report
CSI 5.6.3 IT service continuity management	DS4.9 Offsite Backup Storage
	DS11.6 Security Requirements for DataManagement

B. Information Security Management

1) Principles

1. Information security policy (ISP) and Information security management systems (ISMS) should be defined in Security Frameworks.

2. ISP should support senior IT management activities and business management activities.

3. Security policies should be known by customers.

4. Security policies should be reviewed annually.

5. Information security should be compatible with business security and business needs.

6. All IT service providers should have comprehensive information security management policy and necessary security controls.

Security Framework should contain;

- ISP
- ISMS
- Security Strategy
- Security Organization structure
- A set of security controls
- Security risks management

- Communication strategy and plan about security
 - Training and awareness of this strategy and plan
8. ISP should contain;
- An overall ISP
 - Use and misuse of IT assets policy
 - An account and access control policy
 - Users should enter with identity verify mechanism.
 - Access rights of users should be defined by adapting with business needs, documented and kept in central pool.
 - Access rights are requested by users, approved by system owner and implemented by person responsible for security.
 - Users identities and access rights are recorded in central pool.
 - Approval mechanism should be developed for opening and closing user account.
 - All user accounts and privileges should be reviewed regularly.
 - A password control policy
 - An e-mail policy
 - An internet policy
 - An anti-virus policy (For protecting technology and information systems from malicious software)
 - An information classification policy
 - Make security related technology resistant to tampering and do not disclose security documentation unnecessarily
 - Data security policy
 - A document classification policy
 - A remote access policy
 - A policy with regard to supplier access of IT service, information and components
 - An asset disposal policy
 - Network security policy (Use security techniques to authorize access and control information flows from and to networks like firewall)
 - Protecting technology related to security against fire, water
 - Organizing the generation, change, revocation, destruction, distribution, certification, storage, entry, use and archiving of cryptographic to ensure the protection of keys against modification and unauthorized disclosure
 - Defining a security way for transaction of sensitive data on trusted path

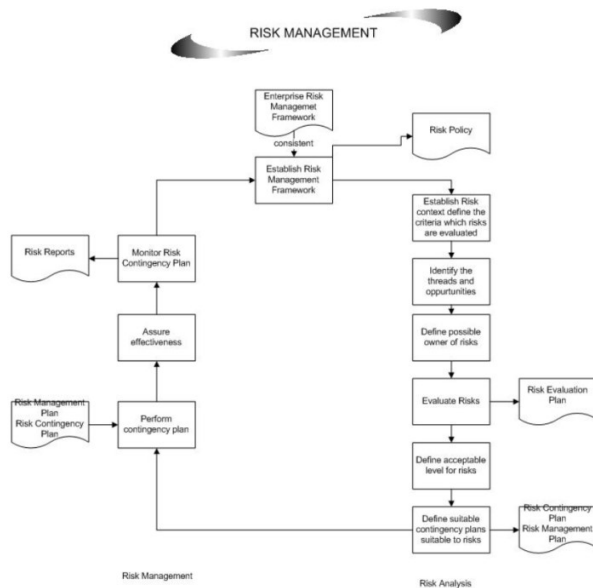


Figure 2: Flow Chart of Risk management process

2) General Information

Table 3: The General Information about Security Management Process

Process ID	P.15
Process Name	Information Security Management
Purpose	Information should be available when required (availability). Information is on served only those who have right to know(confidentiality). Information is complete, accurate and protected (integrity). Information exchanges between enterprises or with partners canbe trusted. (Authenticity and non-reputation).
Scope	Production, maintenance and distribution of ISP Understanding the current and future security requirements ofbusiness Implementation of security controls Documentation of all security controls Management of suppliers and contracts in terms of access to thesystem in conjunction with Supplier management Management of all security breaches and incidents related to allsystems and services. The proactive improvement of security controls, security riskmanagement and the reduction of security risks. Integration of security aspects within all IT SM processes. Management must establish and maintain an InformationSecurity Management System (ISMS) to guide the developmentand management of a comprehensive information security programme. Testing security regularly
Value to Business	ISM maintains and enforced the ISP aligned with the businesssecurity policy and the requirements of

	corporate governance. ISM raises awareness of the need for the security within all ITservices and assets throughout the organization.
Inputs	Business security plans and Risk analysis IT Strategic Plan Service information including details of partner and suppliers Risk assessment Details of security events and breaches from incidentmanagement and problem management Change information Data classification Application security control conditions
Outputs	Information Security Management Policy (ISMS) A set of Security Controls Security threats and gaps Security audits and audit reports Security test schedules and plans Policies and processes and procedures for managing partners and suppliers and their access to services and information Security Training request
Roles	Security Manager Top Executive Manager in business and IT (Security Policy plan should be authorized.)
KPI	Percentage decrease in security breaches reported to Service Desk Percentage decrease in the impact of security breaches and incidents Percentage increase in SLA conformance to security clauses Increase in acceptance and conformance of security procedures Increased support and commitment of senior management The number of suggested improvements to security procedures and controls Decrease in the number of security non-conformance detected during audits and security testing Increased awareness of the security policy and its contents throughout the organization

The flow chart of information security management is given in Figure 3. The inputs, outputs and the process are detailed in this chart. Four main structure; Plan, Do, Act and Check are detailed in these charts. There are four main processes started in each flow, and the output of one flow is the input of the following flow.

3) Mapping Framework with ITIL

Table 4: Alignment with ITIL and COBIT

ITIL	COBIT
SD 4.6 Information security management	AI1.2 Risk Analysis Report
SO 4.5 Access management	DS5 Ensure Systems Security
SO 5.4 Server management and support	DS11.6 Security Requirements for Data Management
CSI 5.6.3 IT service continuity management	

IV. IMPLEMENTATION

In order to prove effectiveness and enforceability, processes are implemented in a very well-known and successful organization in Ankara. To keep the confidence of the organization since it is military based, during this chapter; organization name is stated as Organization AB. AB has an internal IT service named as Information Management Directorship (IMD). All IT services are delivered by IMD to their staffs. IMD consists of approximately seventy five employees. It has own service desk application developed by themselves which are expanded during this study named as Service Management System.

While implementing processes in organization, some small processes are combined and one of them has not been implemented yet. The underlying reason is that each organization has its own culture, approach and needs. Also, IMD has not have enough personnel to implement each process separated. Another important reason is the time restriction of this study. This section provides firstly implemented procedures by comparing the process generated in the scope of this study. Afterwards, the results of implementation of processes are explained.

Any organization who has business services supported by IT service needs IT processes to manage their IT services efficiently and effectively so their business services.

Organization AB has an internal IT supporting business services. In this IT, fifteen processes are defined; IT Management, Service Management, IT Infrastructure Management, Risk and Information Security Management, Process Quality Management, Corporate Architecture Management, Supplier Management, Human Resources, Project Management, Incident Management, Request Management, Change Management, Software Lifecycle and Test Management for IMD. All processes are documented named as department guideline instruction. All department guidelines have consistent structure; Goal, Scope, Definitions and Abbreviation, Related Documents, Role and Responsibilities, Inputs and Outputs, Flow Diagram, Method, Performance Metrics and Appendixes.

Table 5: Processes implemented in IMDB and defined in the proposed model

Processes defined in this study	Processes implemented in IMD of AB Organization
IT, Portfolio and Financial Management	IT Management
Corporate Architecture	Corporate Architecture
Risk Management	Risk and Information Security Management
Software Development Lifecycle	Software Development Lifecycle
Service Level Management	Service Management

Service Catalogue Management	Service Management
Configuration Management	IT Infrastructure Management
Capacity, Continuity and Availability Management	IT Infrastructure Management
Event Management	IT Infrastructure Management
InformationSecurity Management	Risk and Information Security Management
Supplier Management	Supplier Management
Human Resources Management	Human Resources Management
Project Management	Project Management
Change Management	Change Management
Release and Deployment Management	Release and Deployment Management
Test Management	Test Management
Knowledge Management	-
Request Management	Request Management
Incident and Problem Management	Incident Management
Quality Management	Quality Management

As seen in the Table 5, except knowledge management process which is implemented slightly, all processes are implemented in IMD of AB Organization; however, some of them are combined.

The reason why knowledge management process has not been implemented yet is the need of high maturity level of it. The organization in low maturity level does not understand Knowledge Management (KM) subject however organizations with high maturity level are apt to introduce KM practices [33]. The organization which processes are implemented is not in high maturity level to apply knowledge management process in a restricted and short time.

Since, AB organization is quite small and few staff, some processes are merged while implementing. Just because, some processes related to each other are owned and operated by same staffs like configuration management, capacity management and continuity management are all owned and operated by same group in IMD.

All processes are implemented and have been in final reviews by AB organization. To finish all the processes implementation, it has been waited to finish service management system (extension of service desk application) which is being developed by IMD software team.

V. CONCLUSION AND FUTURE WORK

Organizations have to manage their IT services if any to reduce failures generated by IT. Because of this need, IT

Governance has started to become more important and wide spread around the world since 1991. Many standards or frameworks arise since then. However, these standards are complementary rather than competitors. So, most organizations implement these standards together.

The main focus of this study is to provide IT processes which are completely based on ITIL processes if any and COBIT control objectives where processes defined in ITIL do not supply. By implementing defined new processes, organization automatically implement COBIT and ITIL requirements for IT. These processes can also include CMMI up to level 2 by adding some more requirements. In addition to these, ISO 27001's management is also added to Information Security process. Another aim is to reduce ITIL's sophistication and diffusiveness and COBIT's shallow for organizations. This study helps organizations to implement IT processes easily. By giving the relation between ITIL, COBIT and defined processes help the organization to follow the content of COBIT and ITIL also.

For future work, all processes' activities can be extended technically by giving more information about how to use the configuration management tool or service desk tool detail, new processes added or existing processes are extended to meet engineering level of CMMI i.e. CMMI level 3., existing processes can also be aligned with very well-known frameworks/standards such as PMI, ISO 20000 or PRINCE. Also, to measure organization compliance with each process, there should be a set of questions which organization can easily understood the level of the organization in terms of standards and to realize the gap between their own processes and the best practice processes can be handled as future work.

REFERENCES

- [1] Agarwal, R. & Sambamurthy, V. *Principles and Models for Organizing the IT Function*. MIS Quarterly Executive, (1:1), pp. 1- 16, 2002
- [2] Symons, C., Cecere, M., Young, G. & Lamberd, N., *IT Governnace Framework, Structures, Processes, And Communication*. Cambridge, MA: Forrester Research, Inc. 36563, 2005
- [3] Larsen, M. H., Pedersen, M. K. & Andersen, K.V. *IT Governance: Reviewing 17 IT Governance Tools and Analysing the Case of Novozymes A/S*. In HICSS 2006: Proceedings of the 39th Annual Hawaii International Conference on System Sciences. Washington, DC, USA: IEEE Computer Society.
- [4] Ross, J. & Weill, P., *Recipes for Good Governance*, CIO: Australia's Magazine for Information Executives, 17:4, p. 1., 2004
- [5] Cater-Steel, A., Toleman, M. & Tan, W-G., *Transforming IT Service Management – the ITIL Impact*. Proceedings of the 17th Australasian Conference on Information Systems- ACIS. Adelaide, Australia, 6-8 December, 2006
- [6] Cater-Steel, A., Tan, W-G. & Toleman, M. (2006). *Challenge of Adopting Multiple Process Improvement Frameworks*. In: 14th European Conference on Information Systems (ECIS 2006), 12-14 June 2006, Goteborg, Sweden.
- [7] Hayden, L. (2009). *Designing Common Control Frameworks: A Model for Evaluating Information Technology Governance, Risk, and Compliance Control Rationalization Strategies*. Information Security Journal: A Global Perspective, Vol. 18, No. 6, 297-305.
- [8] Hardy, G. (2006). *Guidance on Aligning COBIT, ITIL and ISO 17799*. Information Systems Control Journal, Vol. 1, p. 32-33.
- [9] vonSolms, B., *Information Security governance: COBIT or ISO 17799 or both?* Computers & Security, Vol. 24, pp. 99-104, 2005
- [10] Curtis, B., *Integrating CMMI with COBIT and ITIL*. Borland, 2005
- [11] Brown, A. E. & Grant, G. G. (2005). Framing the Frameworks: A Review pf IT Governance Research. Communication of the Assoc. Inf. Syst. 15, 696-712.
- [12] IT Governance Institute (2006). IT Governance Global status Report - 2006. ITGI and PricewaterhouseCoopers. Available from www.itgi.org
- [13] Weill, P. & Woodham, R. (2002). *Don't Just Lead, Govern: Implementing Effective IT Governance* (CISR Working Paper No. 326). Center for Information Systems Research, MIT Sloan School of Management. Retrieved June 10, 2010, from Massachusetts Institute of Technology Web site: <http://dspace.mit.edu/bitstream/1721.1/1846/2/4237-02.pdf>
- [14] Galup, S., Dattero, R., Quan, J. & Conger, S. (2009). *An Overview of Information Technology Service Management*. Communications of the ACM, Vol. 52, No. 5, pp. 124-127.
- [15] Choi, W., & Yoo, D. (2009). *Assessment of IT Governance of COBIT Framework*. Communications in Computer and Information Science, Vol. 62, pp. 82-89.
- [16] IT Infrastructure Library - Service Strategy. (2007). United Kingdom: The Stationary Office.
- [17] Sallé, M. (2004). *IT Service Management and IT Governance: review, comparative analysis and their impact on utility computing*. Hewlett-Packard Company.
- [18] Da Cruz, E. & Labuschagne, L. (2006). *A New Framework For Bridging The Gap Between It Service Management And It Governance From A Security Perspective*. Academy of Information Technology at the University of Johannesburg. Retrieved June 10, 2010 from http://icsa.cs.up.ac.za/issa/2005/Proceedings/Full/072_Article.pdf
- [19] Bounaguia, Y., Mezriouia, A. & Hafiddia, H. *Toward a unified framework for Cloud Computing governance: An approach for evaluating and integrating IT management and governance models*, Computer Standards & Interfaces, pp. 98-118, 2019
- [20] Mourad, E.B., Malik, M., Anong, A. C. & Mustappa, B. *Combination between Cobit 5 and ITIL V3*. International Journal of Advanced Engineering Research and Science, p. 41-47, 2011
- [21] RiniNur, L. & RZ Abdul Aziz. *Audit of Information System Using Cobit 5.0 And Itil V3 For Information System Of Academic*. Prosiding International conference on Information Technology and Business (ICITB). 2018.
- [22] Paredes-Gualtor, J, Moscoso-Zea, O. & Luján-Mora, S. *The Role of Enterprise Architecture as a Management Tool*. 2018 International Conference on Information Systems and Computer Science (INCISCOS). IEEE, 2018.
- [23] IT Infrastructure Library - Service Design. (2007). United Kingdom: The Stationary Office.
- [24] Mingay, S. & Bittinger, S. (2002). Combine Cobit and ITIL for Powerful Governance (No. TG-16-1849). Gartner Inc.
- [25] Hoekstra, A. & Conradie, N. (2002). Cobit, ITIL and ISO 17799 How to use them in conjunction. Retrieved June 10, 2010 from http://www.cccure.org/Documents/COBIT/COBIT_ITIL_and_BS7799.pdf
- [26] Hill, P. & Turbitt, K. (2006). *Combine ITIL and COBIT to Meet Business Challenges*. BMC Software. Retrieved July 23, 2010 from http://www.smcgltd.com/files/documents/bmc_bpwp_itil_cobit_06.pdf
- [27] COBIT Quickstart 2. Edition. (2007). United States of America: IT Governance Institute.
- [28] Mapping of ITIL V3 with COBIT 4.1. (2008). United States of America, Rolling Meadows: ISACA & ITGI
- [29] Lo, R., & Richards, B. (2009). *ITIL® v3 Knowledge Management Why this new process is more than just your knowledge base*. ThirdSky, pp. 1-32.
- [30] Aygun, B. (2010). *Unification of IT process Models into a simple framework supplemented by Turkish web based application*. Turki: Department of Information Systems.
- [31] IT Infrastructure Library - Service Transition. (2007). United Kingdom: The Stationary Office.
- [32] Objectives for Information Technology. (2007). United States of America: IT Governance Institute (ITGI).
- [33] Sepulveda, J. M., & Tapia, G. G. A Framework for Evaluation and Prioritization of Knowledge Management Initiatives by an Analytical Network Process. Proceedings of 19th International Conference on Production Research, pp. 1-8. Santiago-Chile.

An IOT based Mobile Radar System

Ezgi AYDOĞAN, Erkan BOSTANCI, Mehmet Serdar GÜZEL, Ayhan AYDIN

SAAT Laboratory, Computer Engineering Department, Ankara University, Ankara/Turkey,

ezgiaydogan96@gmail.com

ebostanci@ankara.edu.tr

mguzel@ankara.edu.tr

ayaydin@ankara.edu.tr

Abstract - Internet is one of the biggest innovations in the 20th and 21st centuries. This innovation allows anyone in any part of the world to reach infinite possibilities easily. By the day-to-day development and expansion of the internet, technology called IoT which makes everyday life much easier, is used in many fields and makes objects smarter is getting more important. Today; IoT plays a major role in daily life, as most of the daily work and activities require the use of the internet, the collection or processing of data. Because; IoT's goal is to form automated systems that are able to make human life more efficient and easier. This paper presents a useful IoT device example which acts as a 180° radar system and store its data in a Raspberry which is a central server. After discussions about IoT technology is, the paper explains the process of developing a mobile radar system and suggests potential future improvements.

Keywords – IoT, Arduino, Radar System, Sensor, Raspberry Pi, Wireless Sensor Networks

I. INTRODUCTION

Internet is a global computer network system in which computers, servers and smart devices communicate with each other and data exchange as well. This system uses the standard internet protocol called TCP/IP to serve millions of users around the world. Thanks to its all-purpose nature, TCP/IP communication can be integrated into different kind of devices so that the devices can get in intercommunicate easily. Internet of Things called IoT is a giant network that can transfer and receive data by making devices smarter. The word *smart* means to making devices communicate with each other without human intervention. Moreover, IoT device is the device that communicates over IP traffic by automating the operations on this network to make life easier and more efficient. [1] When IoT devices communicate over IP without any wire, this IoT communication is also known as Wireless Sensor Network (WSN). WSN refers to a group of spatially dispersed and dedicated sensors for monitoring and recording the physical conditions of the environment and organizing the collected data

at a central location. In addition to providing user interaction, it is simply based on the communication of objects. Objects that can connect to the Internet and collect data with sensors transmit or receive the data they collect by aid of a node. Thus, IoT devices are equipped with wireless network cards to connect wirelessly to the backbone of a network and to have internet access.

Sensors are the most important key parts of IoT devices. Sensors are capable of detecting and measuring physical phenomena such as temperature, pressure and obstruction. [2] One of the most important and frequently used feature of a sensor is distance measurement to object. Parking systems, automatic doors, liquid level measurements and defense industry are areas where this type of sensor is commonly used.

Collecting and transmitting data accurately and completely is vital for data processing and real-time monitoring. Communication between server and sensor to which the data is to be collected and transmitted is performed by IP traffic. Developing radar systems used in distance measurements requires developing hardware and software, moreover, put them together in such a way that they work together in harmony.

This paper is based on Mobile Radar System using sensors with these techniques. The aim of this study is to store distance measurement while collecting data around 180° at the same time by data exchange between Arduino and Raspberry Pi. Such a system can be used in maritime vehicles with image-processing or other defense industries to detect and warn in case of danger. Evenmore; if more sensitive sensor is used and this system is improved with sound, it could be also used as an alarm system. Since it is “deploy and forget” system, once it is developed, it does not need any intervention easily.

The paper is structured as follows: Section II is about the recent works on IoT world, Section III represents methods used in this study, Section IV mentions the study, Section V shows the results and conclusions obtained by this study are drawn in Section VI.

II. RELATED WORKS

Since the birth of Internet in 1989, the connection of objects to internet has become widespread. The Trojan Room coffee machine may be one of the first examples. Afterwards, in 1999, Kevin Ashton revealed the concept of IoT by inferences and sensor definitions from other smart devices developed. [3] Evenmore, at that time, RFID based identification system which has an important role in IoT world has been invented. In 2000s, IoT devices started to be used on commercial purposes. Then, by the time innovative companies such as Cisco, IBM, LG and Ericson announced some IoT devices, this concept gradually entered social life.

And recently, IoT has gained more effectiveness and become much more widespread used. Because of its heterogeneity, capacity of interoperability with other devices, flexibility and mobility; IoT does not only mean smart devices that connect to Internet. It became a paradigm which suggests solutions to daily-life activities in agriculture, security, medicine and so on.

Such situations that people want to take precautions or are not able to interfere but could be a problem make daily-life easier thanks to IoT. Damages can now be minimized by many IoT devices aimed at taking measures against natural disasters. The floodlights that broadcast the current water level and the location of the sudden change through the internet, devices that calculate the possibility of lightning or the alarms that predict the landslide.[4] are the best proofs how this concept is helpful.

By continuous increase in the world population which means increasing demand for food, the IoT has begun to show itself in agriculture, and today, smart agriculture has become one of the fastest growing field in the IoT. Detection of soil moisture and nutrients, controlling water us for plant growth, determining the required fertilizer and tractors without driver are the places where these devices are used constantly. [5]

Since it is directly human interactive, one of the most important and frequent branches in which IoT is used is medicine. Along with the devices developed in this field, rapid progress has been made in health sector. Myo which is used in orthopedics for patients who need to be exercised after fracture is good enough as an example of those devices, Thanks to Myo, while patients are watching their progress, doctors can measure and check the angle of that movement. One of the examples in medical field is Telemedicine technology. This technology enables face-to-face interviews with computers, networks and other

technologies and equipment in this area, although experts, doctors and patients are in different locations. Additionally; it provides time saving and efficiency for both doctors and patients. [6] Smart lenses and smart devices that measure the amount of glucose in the blood are other examples. [7]

In addition to being a facilitator for human life, IoT technology is also widely used for protection purposes. The most obvious example is the IoT systems used in mines. These systems, by means of sensors, measure the gas concentration in the work area of the workers, the amount of hazardous gas that each mine can emit, and transmit them to a data processing unit. [8]This unit may prevent accidents likely to occur due to some transmitted data being too high.

The increasing popularity of IoT in the fields of economy, academia and the state gives ideas about future applications. [9]Implementation of transport laws, monitoring air quality, discovering emergency routes and lightning cities more efficiently are examples of future application estimates.

III. METHODS

A. Arduino

Arduino is an open source electronic platform whose hardware and software are designed for easy-to-use. The hardware-based Arduino cards are basically the process of reading and processing an input value and produce an output from that value as shown in Figure 1. This input value may be a finger on a reader or an object approaching to the sensor, while the output may be turning on a led or making an engine start. Therefore, Arduino cards are often called microcontrollers. Arduino consists of Arduino IDE for development, Arduino libraries, a microcontroller program called AVRdude and a compiler called AVR-GCC.

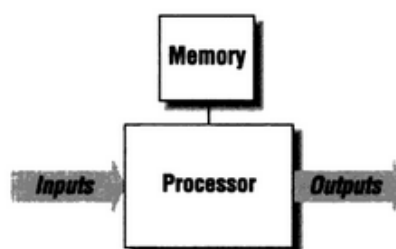


Figure 1: Generic Arduino card process

Arduino software includes integrated development environment(IDE) and libraries. This development environment, called Arduino IDE, is written in the Java and provides great convenience to program Arduino, but also contains many libraries and examples of these libraries. The libraries were written in C and C ++, compiled with AVR-GCC.

The codes transferred to the Arduino are compiled with the AVR-GCC and codes compiled with the AVRDUDE component are programmed.

The hardware features of Arduino cards may vary for different models; the basic procedure in all models is to generate a data using input pins located on the card, process it, and produce outputs to the outside such as sound, light, movement.

B. Packet Transmission and Network

Network is a structure where computers and other devices communicate with each other through communication channels and this structure is based on OSI model. [10]

- 1) Physical Layer: It enables the transmission of signals to devices and determines the physical properties such as voltage, current, speed, data type. A wired or wireless connection can be preferred for the transmission of signals. Wired connection is more reliable but more demanding as it carries the transport of digital signals between devices over a cable. On the other hand, wireless connection uses radio waves to provide signal flow. Since it is not possible to carry digital signals without cable, it returns them into analog signals and thus, two devices can communicate with each other without any sort of cable and independent from distance between them. A wireless network called Wi-Fi or IEEE 802.11 is designed as a wireless LAN (WLAN) technology and is a protocol used in all wireless network cards.
- 2) Data-Link Layer: This layer decides whether transmitted data will be verified or not. Data-link layer is divided into two sublayers. In the first layer, which is called MAC layer, the physical addresses of the devices where the data comes or goes are determined and in the second layer, LLC, error is detected.
- 3) Network Layer: This layer is responsible for the logical addressing of the packages and delivering packages to the correct address with the help of the routing table. Another important one is the Internet Protocol (IP), which transmits the packets of interconnected devices from the source IP address to the IP address of destination. This transmission occurs thanks to routing algorithms. As a communication type between devices, wireless network is more suitable for this study in terms of its portability.
- 4) Transport Layer: It ensures safe or unreliable transport of data to the destination address. Divides incoming packets into smaller packets and reconfigures outgoing packets. One of the protocols running on this layer is TCP, a reliable protocol that

performs these transport operations. Therefore; it is used to send sensitive information between the two devices and cares about whether the sent or received data is lost. UDP is another protocol that runs on the transport layer and is responsible for sending packets. But; unlike TCP, UDP does not control data flow, does not care about loss, and does not care if data is transmitted. Thus, it is faster but not reliable. For real-time monitoring it is important to be fast while transmitting data to the server. So; transmission over UDP is used in the study.

- 5) Session Layer: It is responsible for establishing a connection between the two devices, managing and terminating this connection.
- 6) Presentation Layer: Checks whether the outgoing data on the application layer or the incoming data on the session layer are appropriate.
- 7) Application Layer: This layer manages user interactions with the operating system or application. Created or received data could be HTTP requests, answers, TELNET command, file transfers and so on. To monitor data, Android application runs on it.

C. Raspberry Pi

Raspberry Pi which can be defined as a small and affordable computer was first developed for people who try to understand how computer works and how to code by Raspberry Pi Foundation. And today, there are so many scenarios such as weather station, virtual assistant, media center, smart home appliance, robot brain, drone assistant. Even it is slower than any modern laptop or desktop computer, it is still a well-equipped Linux computer and can perform all the tasks expected in low power consumption. Although Raspberry Pi is often thought to be the same as Arduino, it is more convenient to create and implement more complex circuits. Its Wi-Fi module in many models can communicate with different devices via wireless network.

D. C and C++

The C programming language is a high-level programming language developed in the mid-1970s. Compared to the languages used before C, it is more readable, flexible and efficient in terms of memory usage. C is a procedural language which means it can be easily divided into parts, easier detection can be made and it provides convenience. Although the purpose of the exit was to write Unix programs, it is now used to write applications in many fields. In particular, many microcontrollers have C compiler. And this situation makes C programming language vital for embedded systems.

C++ programming language is an object oriented language improved on C language. Therefore; this language is very similar to C language in terms of syntax, but it is different in terms of structure. Programming with C++ provides to create objects, while in C programming is done with procedures not objects. Since object-oriented languages allow creating objects within the program, it reduces code repeats and makes code management easier. Despite of these advantages, it is not preferred over embedded systems with limited resources and efficiency because C++ language requires very high memory and processing power.

In this study, with the aim of programming Arduino card, Arduino programming language; which includes C and C++ and can be divided into 3 parts as structure, values and functions; is used.

E. Python

Python is an open-source, dynamic structure, object-oriented, interpretable, high level programming language. Python's simple and easy-to-write format makes it very popular in recent years, as well as the fact that it does not require a compiler, and this programming language is preferred. Python programming language and libraries were used as an aid in ensuring communication between Arduino and Raspberry Pi.

IV. MOBILE RADAR SYSTEM

The structure of the study consists of 3 stages as seen in Figure 2 and those are summarized as measuring distance by ultrasonic sensor, making Arduino card connect to a network providing IP and communication between Arduino and Raspberry Pi over UDP.

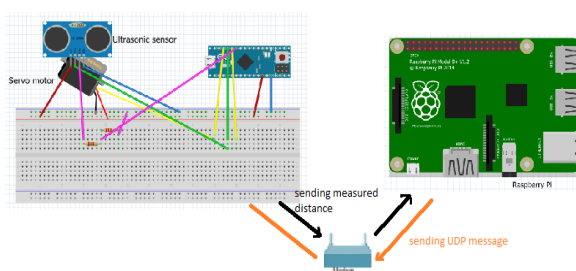


Figure 2: System architecture

Modem is used for converting digital signals in a computer into analog signals. It enables computers to connect to a defined public network and access other computers in remote locations that are connected to this public network. Although there are many sorts of modems such as Ethernet, USB, PCI, the most widely used one is wireless modems. Wireless modems both publish a network and allow computers equipped with wireless network card in the same area to access the internet and each other.

During the study, wireless modem type is preferred to eliminate wiring difficulties and ensure the portability.

Necessary connections must be provided for ultrasonic sensor to communicate with Arduino. For this purpose, sensor's "trig" and "echo" pin must be connected to two digital pins of Arduino. Where those connections start and finish is also coded in the program which will be uploaded to Arduino card.

The ultrasonic sensor emits an airborne ultrasound. This ultrasound returns to the output source when it encounters an object or an obstacle in its spreading area.

The distance of the objects can be calculated by considering the time that the ultrasound spreads and the speed of sound. The speed of the sound is constant and is 340 m/s or 0.034 cm/μs. So; if the formula below is considered:

$$\text{Distance} = \text{Speed} \times \text{Time}$$

distance can be calculated easily. But the ultrasound goes out from "trig" and comes to "echo" pin travels through two times of distance at which the object is. Therefore; half of the value must be considered while calculating.

Measured distances were monitored by the serial monitor provided by the Arduino IDE. By changing the value of function delay() used before printing the distance, different number of values were calculated at the same time interval and compared according to the graph as shown in Figure 3.

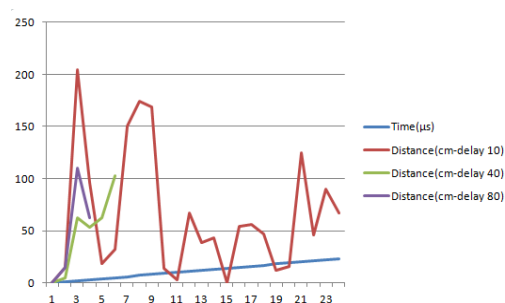


Figure 3: Measured values with different delay time

The way how to connect the Arduino card to a network depends on the features of Arduino and the network type such as wireless or ethernet. One way is using an Arduino card without wireless network card and external Wi-Fi card module to make Arduino connected. In such a system; ultrasonic sensor gets the distance value, write it to the file which is in external SD card module to store files in Arduino and sends the file via external Wi-Fi card module over FTP. But; transferring the data read on the Arduino card back to the Wi-Fi module and

communicating with Raspberry Pi will not be efficient. Moreover; to transfer the files over FTP, Arduino card must be coded as an FTP client but as it is mentioned before Arduino has limited hardware facilities and slower than UDP. Thus; FTP is not appropriate for the aim of real-time monitoring.

Instead, Arduino card with wireless network card is preferred and thanks to functions in Arduino library, card can be connected to a network with a few functions. With UDP, the Arduino UDP server is used to transfer distances from Arduino to Raspberry Pi and the Raspberry Pi is programmed as UDP client. Thus; how often the distances are collected on the central server is determined by the frequency of requests made by Raspberry Pi.

As a final step for Arduino, with 180° servo motor is used to scan area. 3rd pin of motor is connected to Arduino like one for power, another one for ground and the other one for making Arduino know there is a servo motor attached on itself.

Raspberry Pi is programmed with the Python code to act as a UDP client and wait for a response from the server, Arduino. Other programming languages such as C could be used for this. But; Python has become a more effective option with its easy typing, speed and ready-made libraries.

In code uploaded to Raspberry Pi is structured as: UDP port is started to listen. UDP message is sent as a client. Arduino notices the package on the UDP port that it listens to. Then, Arduino sends back the current measured distance value to the Arduino according to the IP and port number known thanks to received UDP packet. The incoming package has UDP and IP headers. In UDP title as shown in Figure 4; the port number of the source device that sends the packet, the summary to control the data and the port number of the destination device to which the packet will go. Arduino gets all these information with functions in UDP library.

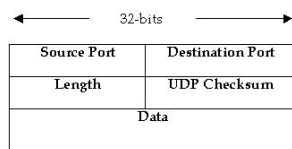


Figure 4: UDP packet header

In the IP header; there are information about communication between two devices such as protocol, source and destination addresses, summary of data. Thanks to this information; information about where the package comes from and where it will go and whether there is any

corruption in the data can be learned by both the server and the client.

After the Arduino code is loaded and the Python code is executed; they can communicate and transmit data via Raspberry Pi and Arduino UDP sockets. Scheduled tasks are also available in Raspbian Pi, as in all GNU / Linux distributions, to automate the Raspberry Pi script without human intervention.

In GNU/Linux systems, there are commonly used tasks to schedule other tasks. They are called cronjobs. Tasks scheduled with *cronjobs* are regularly executed. Therefore; writing datas receiving over UDP to a file is scheduled with cronjobs in Raspberry Pi. So; they are able to be reached any time from the file.

V. CONCLUSION AND FUTURE WORKS

As a result, the obtained data can be transmitted, not to a device such a device that it has limited resources, to Raspberry Pi which can make easier tracing and has more powerful performance via UDP. Two devices can communicate with each other accurately. In the final step, instead of viewing datas in an Android application, Processing which is a software to code within the context of the visual arts is used to monitor as shown in Figure 5. Processing software takes the data from a port which is defined and draws lines according to that data. Green lines mean there is nothing in range while red lines refer an object.

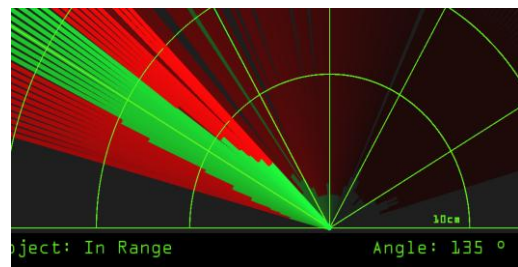


Figure 5: Monitoring data

This study will be improved by Android. Data stored in a file in Raspberry Pi as a central server will be monitored in an Android application real-time. More sensors could be added and their different locations could be tracked in the application.

REFERENCES

- [1] Wireless Sensor Networks, Shuang-Hua Yang Principles Design and Applications Shuang-Hua Yang, Springer, London
- [2] Bob Familiar and Jeff Barnes 2017 B. Familiar and J. Barnes, Business in Real-Time Using Azure

- IoT and Cortana Intelligence Suite, DOI 10.1007/978-1-4842-2650-6_
- [3] A state of the art review on the Internet of Things(IoT) History, Technology and fields of deployment, 2014 International Conference on Science Engineering and Management Research (ICSEMR)
- [4] Internet of Things for Disaster Management: State-of-the-Art and Prospects, IEEE Access (Volume: 5), 2017
- [5] Internet of Things(IOT) and Cloud Computing for Agriculture: An Overview V.C. Patil¹, K.A. Al-Gaadi², D.P. Biradar³ and M. Rangaswamy¹, Agriculture 2012 (AIPA 2012)
- [6] The application of IOT in medical system Dongxin Lu , Tao Liu, 2011 IEEE International Symposium on IT in Medicine and Education, 9-11 Dec. 2011
- [7] Medical Internet of Things and Big Data in Healthcare Dimitar V. Dimitrov, MD, PhD Diavita Ltd., Varna, Healthcare Information Reserach, 2016, Bulgaria.
- [8] Discussion on Application of IOT Technology in Coal Mine Safety Supervision ZhangYinghua, FuGuanghua, ZhaoZhigang, HuangZhian, LiHongchen, YangJixing, International Symposium on Safety Science and Engineering in China, 2012 (ISSSE-2012)
- [9] Future Internet: The Internet of Things Architecture, Possible Applications and Key Challenges Rafiullah Khan , Sarmad Ullah Khan , Rifaqat Zaheer , Shahid Khan, Frontiers of Information Technology (FIT), 2012 10th International Conference on
- [10] The OSI reference model, J.D. Day ; H. Zimmermann, Proceedings of the IEEE (Volume: 71 , Issue: 12 , Dec. 1983)

Web and Mobile Based Online Joint Working Platform Development for University Students

G. GÜMÜL¹, E. ÖZTÜRK¹ and N. AYDIN ATASOY¹

¹Karabuk University, Karabuk/Turkey, gumulgokhan@gmail.com

¹Karabuk University, Karabuk/Turkey, ezgiozturk96@gmail.com

¹Karabuk University, Karabuk/Turkey, nesrinaydin@karabuk.edu.tr

Abstract – Web based and mobile based applications are part of our life because they are easily accessible. These kinds of applications are appearing in areas such as education, business life, health, social life etc. It was both costly and long process to perform any work before mentioned above web and mobile solutions for us. Nowadays, these applications provide facilities to us in our education and business life, easy transportation, collaborative work, interactive work and speed, etc. In this study, we develop a web and mobile based application that can be used for education platform. Students are able to create a joint working platform with this application. At the same time this platform brings the student and his / her teacher together in a common structure.

Keywords – Mobile and web application, online joint working, web service, reacts native.

I. INTRODUCTION

WEB application is the name given to programs accessed via internet, network connection. Web applications can be used without any program installation process. It is not required any updated When the server is updated, all computers that are using the application is updated. Web applications provides a wide using facility [1]. Web based applications can be created with various programming languages. Although each programming languages have different function and features, when developing software, it can be developed an application with the same function using many programming languages. All software that is used via internet browser and provides communication with http protocol is example of web based application [2]. Nowadays, it is possible to access web based applications via browsers which can be installed on phone and tablets through common mobile application [3]. Mobile devices have been getting adopted by big masses people around the world since 2000's. It is shown that mobile devices and mobile applications are widely used worldwide and, in our country, according to the statistical results. There are different options for developing mobile applications. Some examples of these are Native, Hybrid, Cross and Responsive platforms. [4]. Using of paper and pencil is decreasing as technology is constantly evolving web and mobile applications increase. There are

various note taking, note sharing applications are available. Some of those; Notes.io, Evernote, Google Keep, OneNote, Simplenote, Any.Do, Todoist: To-Do List, Wunderlist: To-Do List& Tasks [5]. In this study, it is aimed to combine students in a common area. It is aimed at a platform that Students' online lecture notes during the day, information about the courses, activities at the school and club activities, sharing his / her notes.

II. METHOD

Many different technologies are used in the study. Ms-SQL is preferred as a database. SQL Server 2014 Management Studio is used in the management of the generated database [6]. Asp.Net Mvc is used for web application developing. Mvc; is consists of the initials of View, Controller, and each word represents a layer [7]. It is based on Web Api Rest architecture. REST architecture supports methods that can be implemented with http requests such as Get, Post, Put, Delete etc. [8]. Xml, Csv, Html, and the Json data type in this study can also be returned to the client in the Rest architecture. Visual Studio is one of the applications used as a code development environment, in this study. Visual Studio is an Ide is produced by Microsoft that is used to create graphical user interfaces, Windows forms, web services or web applications. There are several programming languages supported by Visual Studio. C # is preferred as the programming language on the web side of this study [9]. Postman is an application that used to test web services, view these test result and to document [10]. The written web service is tested with Postman. The purpose of using is the user interface is easy and understandable. React Native [11] is used in mobile application development. It is a development environment that provides to write native applications based on the ReactJS framework.

III. APPLICATION

This study consists of three parts: Web site, Web service and Mobile application. The operating mechanism of the application is shown in Figure 1.

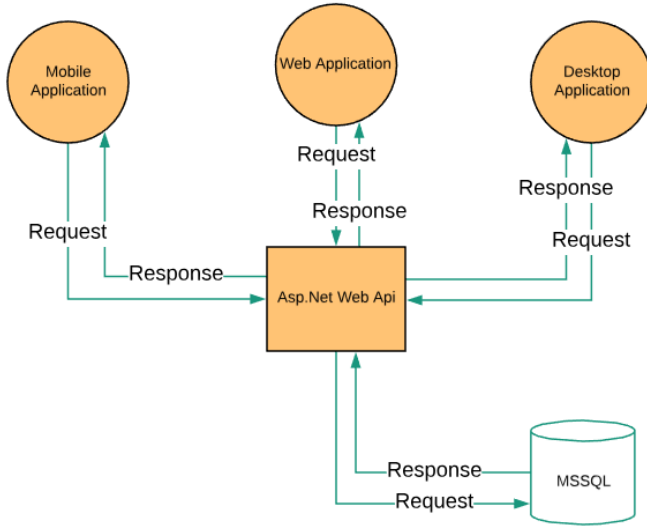


Figure 1: Application operating mechanism.

A. Web Site and Web Service

Web sites can be used on user's internet environment for sharing information, sometimes at any sale of product, sometimes as a means of promotion, or it create a structure where many different users interact with each other. There are many languages, environments and different design pattern structures that we can use when designing websites. Asp.Net Mvc Framework technology is used in this study Asp.Net Mvc is a .Net Framework technology for creating web applications. Asp.Net MVC design pattern consist structure that called Model, View and Control. The model is the structure that containing data relevant to the application. This structure separates the data layer from the application, so there is no need to know where the data layer is in other layers, so the View layer is the layer where the user interacts with the application. In this layer, Html, Css, JavaScript etc. technologies are used by familiar with web technologies Controller layer acts as a bridge in this pattern. It processes the request from the user and takes necessary actions accordingly

Web services are platform-independent technology that responds to data types such as Json, Xml. Its independence from the platform comes from the fact that the structure is based on http. Asp.Net Web Api, a framework used to create services that can communicate through the Http protocol, is used when creating a web service. Asp.Net Web is based on Api Rest architecture. Rest architecture is a structure based on communication between server and client. Rest services are

called Restful services. Web service operating mechanism is shown in Figure 2 and Web site user screens are shown in Figure 3.

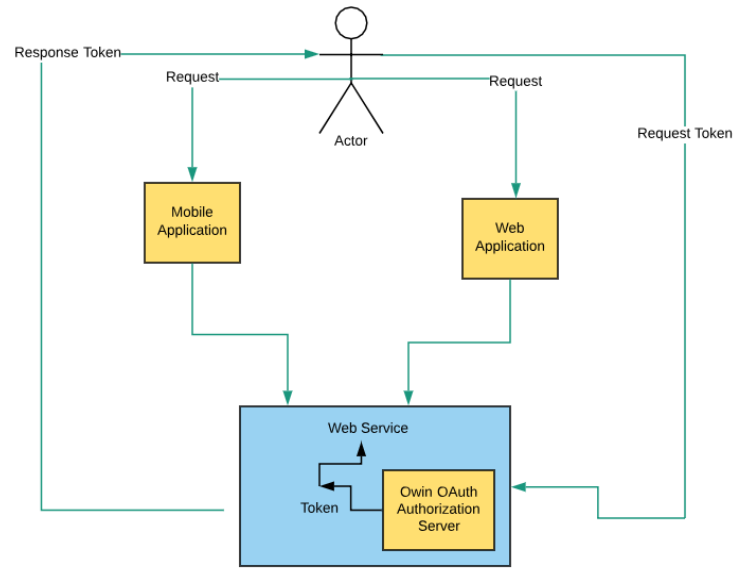


Figure 2: Web service operating mechanism.



Figure 3: (a) Website application home screen.

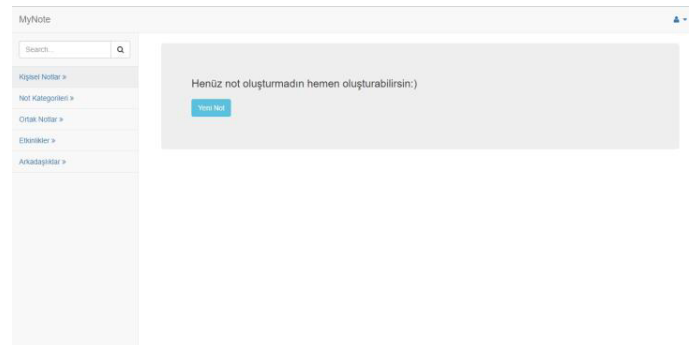


Figure 3: (b) User home screen.

Figure 3: (c) User new note creating screen.

Figure 3: (d) User notes screen.

Figure 3: (e) Picture gallery.

Figure 3: (f) Friend insert screen.

Figure 3: (g) User file screen.

Figure 3: (h) User activity screen.

B. Mobile Application

The mobile application is software designed and coded specifically for mobile devices such as smartphones and tablet devices. Increasing the usage rate of users from desktop devices to mobile devices is made mobile applications more valuable and important for users [12]. Mobile applications are made using IOS and Android operating system. It is possible to develop mobile applications for each project. The point to be considered is that the application will be developed taking into account the operating systems. Mobile applications are faster and easier to use than website, this encourages users to use mobile applications. For this reason, the time spent by users with their mobile devices increases day by day [13].

Hybrid, one of the mobile application development platforms that enables the development of mobile applications for different platforms via basic web technologies. This approach, which is shaped by the HTML5 and Angular architecture, enables the user to be able to work in close proximity to each other in different platforms with practical and easy methods. The application can be achieved with a single development work on both Android and IOS platforms [14].

One of the biggest advantages of Native mobile application development is the ability to access the capabilities and information of the device. Native applications are applications written in the same language as the system. Native application provides possibility of users to develop android and IOS application simultaneously with a single panel. Users can access device management and permissions of their mobile applications and edit them as they wish. [15].

The mobile part of this study is created using React Native. React Native ReactJS is a development environment that allows mobile applications to be written through Native interfaces based on the framework. [16]. User login screen of the study is shown in Figure 4.

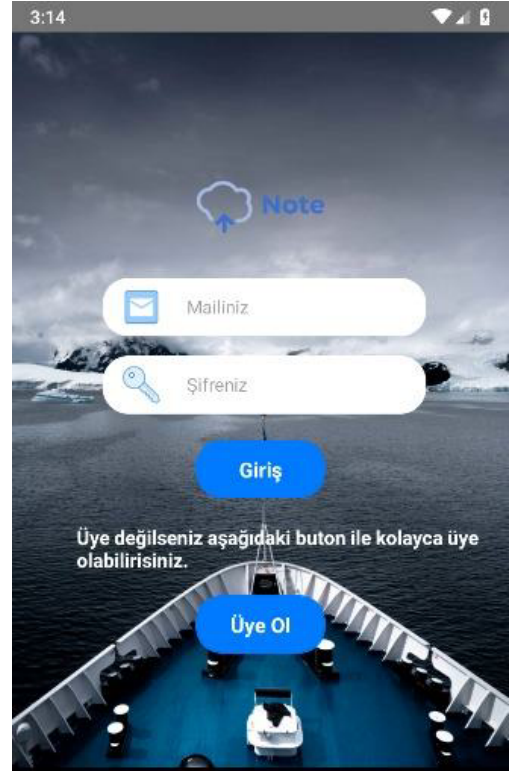


Figure 4: Mobile application user login screen.

The user must be registered in advance to be able to access the application. The user registration screen is shown in Figure 5.

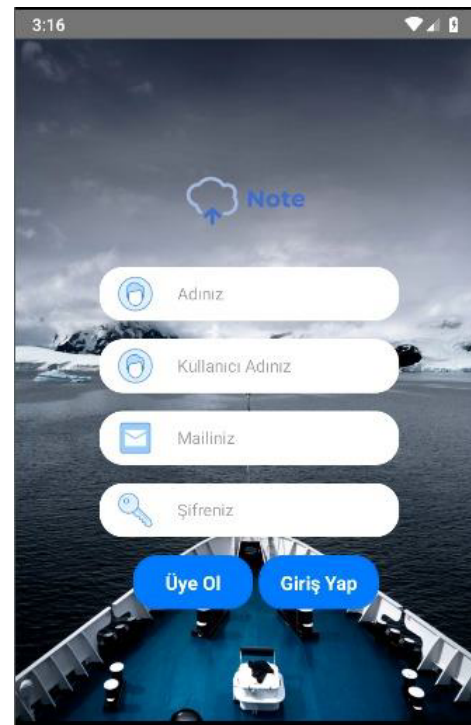


Figure 5: Mobile application user registration screen.

The sidebar menu screen will appear after user login the application. The sidebar menu screen is shown in Figure 6.

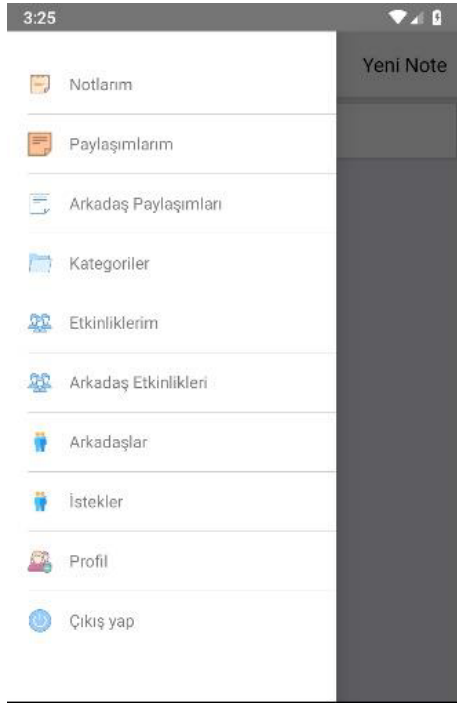


Figure 6: Sidebar menu.

My notes section includes the user notes, my sharing section includes the notes that other users share, friends sharing section includes other users that the user is friend, categories section includes which category belongs to the notes, my activities section includes the activity that is created by other users, friends activity includes other users' activity that the user is a friend, friends section includes other users that the user is a friend, request section includes friendship request that is sent by other users, profile section includes the user's own profile info (name, user name, e-mail).

IV.CONCLUSION

When technology is getting to develop, paper and pen are placed instead of tablets, phones and computers [17]. We have developed an interrelated website, web service and mobile application in this study to suit the needs of today. First of all, web service has created that can be called the roof of the application, the website and mobile application are united under this roof. The students were brought together in a common area, a platform was created in which students could share their online course notes, information about courses, activities at the school, club activities, notes and also notes about the activities that the person did during the day. In the next version of the application, it is aimed to bring the study on a common platform not only for universities and students, but also for employees and employers in business life.

REFERENCES

- [1] Internet: <https://www.mediatick.com.tr/blog/web-application-nedir>
 - [2] Internet: <https://wmaraci.com/nedir/web-tabanlı-uygulama>
 - [3] Internet: <https://www.mediatick.com.tr/blog/web-application-nedir>
 - [4] Nilgün Özdamar Keskin, Hakan Kılınc "Mobil Öğrenme Uygulamalarına Yönelik Geliştirme Platformlarının Karşılaştırılması ve Örnek Uygulamalar" Açıköğretim Uygulamaları ve Araştırmaları Dergisi, vol. 1(3),pp.68-90, June 2015.
 - [5] Internet: <https://webrazzi.com/2017/02/10/not-tutmak-icin-en-iyi-7-mobil-uygulama/>
 - [6] Internet: [http://www.dahiweb.com/microsoft-sql-server-nedir-ne-ise-yarar/\(2003\)](http://www.dahiweb.com/microsoft-sql-server-nedir-ne-ise-yarar/(2003))
 - [7] Internet: [http://www.borakasmer.com/asp-net-mvc-nedir-ne-ise-yarar/\(2012\)](http://www.borakasmer.com/asp-net-mvc-nedir-ne-ise-yarar/(2012))
 - [8] Internet: [https://coderwall.com/p/cpdzuw/asp-net-web-api\(2007\)](https://coderwall.com/p/cpdzuw/asp-net-web-api(2007))
 - [9] Internet: [https://wmaraci.com/nedir/visual-studio\(2014\)](https://wmaraci.com/nedir/visual-studio(2014))
 - [10] Internet: <https://www.getpostman.com>
 - [11] Internet: <https://facebook.github.io/react-native/>
 - [12] Bora Aslan, Füsun Yavuzer Aslan, "Mobil Programlamanın Önemi ve Bir Müfredat Önerisi", in 3.Uluslararası Meslek Yüksek Okulları Sempozyumu, pp: 81-88,
 - [13] Internet: <https://www.moradam.com/20180211202905/mobil-uygulama-nedir-mobil-uygulama-yapan-firmalar-hakkinda-bilmeniz-gerekenler/>
 - [14] Internet: <https://medium.com/ios-development-turkey/hybrid-ve-native-uygulama-geli%C5%9Firme-220a0eda947c>
 - [15] Internet: <https://destek.mobiluygulamamerkezi.com/native-uygulama-gelistirme/>
 - [16] Internet: <http://devnot.com/2017/react-native-ile-android-ve-ios-programlama/>
- Görkem Tılıç, "Yabancı Dil Öğreniminde Kullanılan Etkileşimli Mobil Uygulamalar: Duolingo Örneği", Eğitim ve Öğretim Araştırmaları Dergisi (Journal of Research in Education and Teaching), vol. 5, pp. 303-313, June 2016.

Human-Computer Interaction with One-Shot-Learning

E. ŞEKER AYDIN¹ and O. FINDIK²

¹ Karabuk University, Karabuk/Turkey, esmaseker@karabuk.edu.tr

² Karabuk University, Karabuk/Turkey, oguzfindik@karabuk.edu.tr

Abstract – Gesture recognition is one of the growing research areas. It has a wide area of application including human-machine interaction, sign language etc. Various approaches have been proposed to recognize gestures. Some of these approaches require prior knowledge of gestures. Others need a very large dataset. However, the one-shot learning approach does not require prior knowledge of gestures and a very large set of data. Hand gestures play an important role in people's lives. Thus, the hand gesture recognition system provides an innovative and natural way for people to interact with the computer. In this paper, a variety of methods for recognizing hand gestures are proposed with a one-shot learning approach. We aim to obtain a single image from each gesture video. Two different datasets are used. These are ChaLearn dataset, which is a very rich dataset, and the other one is a novel dataset designed by us. It is difficult to train the system with only one example. However, both datasets offer many new possibilities for the system. The second dataset transfers video to the file as a direct frame via designed interface. Three basic methods were used in the study. These are STD, MHI, 2D FFT. Three different distance measurements are used to match the gestures which are Levenshtein, Mahalanobis, and Frobenius. In the results, it is seen that the secondary dataset designed by us present a higher success rate. In addition, it is observed that Levenshtein distance measurement supplies a more accurate classification in both datasets.

Keywords – human computer interaction, one-shot learning, std, mhi, 2d fft.

I. INTRODUCTION

Recognizing human gestures is an important issue for human computer interaction. Therefore, gesture-based devices are prominent in recent years. Such devices use direct body gestures without being connected to any apparatus. Medicine, robotics, sports analysis, film, games, biomechanics etc. areas are some of the various application areas used to recognize human gesture [1]. However, recognition of human gestures is a very challenging research area in the framework of human machine interaction.

Various methods and approaches have been proposed and applied to define human gestures. Some of these approaches remained in theory, and the other part did not achieve the desired success in practice [2,3,4,5,6]. Some of these approaches can be summarized as follows. HMM (Hidden Markov Model) has been widely used to model human gestures. HMM is a stochastic model and suitable for handling

stochastic properties of gesture recognition. Instead of using model geometric properties, gestures are converted into sequential symbols. This model is used to represent gestures and parameters which are learned from training data [7,8,9]. HMM is used in the star skeleton study [8]. The star skeleton is a rapid skeletal technique that connects the center of the target object to its ends. An action is made up of a series of star skeleton over time and transformed into a feature vector sequence with HMM. Then, a code structure was created to know the star skeleton in each type of action and to measure the similarity between feature vectors [8]. In that study, the HMM has been observed to be having deficiencies and the model has been developed over time. In the following years, some researchers have proposed a global HMM to recognize gesture sequences using a one-shot learning database [10]. However, this model remains complicated due to the large number of reasons.

DBN (Dynamic Bayesian Networks) is a more general probability model that combines HMM and Kalman filters [11]. It is also a generalized version of Bayesian networks. However, such approaches require a large set of data to cover all classes of gestures [11,12,13,14]. The CRF model was first introduced as a framework for segmenting and labeling array data and creating probabilistic models [15]. Later, some researchers applied it to recognize the dynamic hand gestures. 11 different human gestures were classified using CRF. This model is not a prominent method for gesture trajectory recognition. However, it performs well to remove the most representative sequences. The number of samples and complex calculations are needed to recognize gestures with the CRF model [10,15,16]. HOG (Histogram of Oriented Gradients) and HOF (Histogram of Optical Flow) are some of the common features used in one-shot based approaches [17,18]. HOG can well analyze the distribution of local density gradient directions of the appearance and shape of an object [17,19,20]. This feature works in the local grid unit of the image, thus providing good invariance on geometric and optical deformation [17,19]. The HOF tries to predict the motion between the two images in each case. The HOF equation is expressed as a single equation with two variables. In all HOF methods, the additional conditions are provided to estimate the flow [17,21,22].

These approaches have many common disadvantages. Some of these are expressed as: i) They need a large number of

datasets. ii) They must have prior knowledge about the gestures. iii) They have trouble in recognizing new gestures.

In our study, one-shot learning approach was used to recognize hand gestures. Four basic processes were applied when using this approach. These are 1) STD (Standard Deviation), 2) MHI (Motion History Image), 3) 2D FFT (2D Fast Fourier Transform), 4) Distance measurements used to compare two images. These distance measurements are Levenshtein, Frobenius, and Mahalanobis [23,24]. We employ two different datasets in the study. The first one is ChaLearn, which is a well-known dataset. The second dataset is a novel one which was designed by us. The study does not require prior knowledge about gestures. It also easily recognizes new gestures added to the dataset.

II. METHODS

A. Gesture Recognition Datasets

Until now, many datasets have been created to develop, educate and evaluate algorithms. In addition, these datasets have been made open to public [25]. However, gesture recognition data requires their own special datasets, so that there are many gesture recognition datasets commonly available today. Some of the most common datasets are NTU RGB+D (Nanyang Technological University's Red Blue Green and Depth Information), MSRGesture3D, and MSRC-12 (Microsoft Research Cambridge-12) [25,26]. NTU RGB+D dataset consists of 4 different models which are RGB videos, depth map sequences, 3D skeleton data, and infrared videos. This dataset has 56.880 gestures samples [25,27]. MSRGesture3D dataset was created by the Kinect device and contains American Sign Language gestures performed 2 or 3 times by 10 people. The dataset has only images of the segmented hand gestures [25,28]. MSRC-12 is a dataset that uses skeleton data to recognize human gestures. It contains images collected from 30 people performing 12 gestures and has 6.244 gesture examples in total [25].

Two different datasets were used in the present study. These were ChaLearn Gesture dataset and the novel dataset produced by us. The Kinect v1 camera is used within the ChaLearn dataset and only the upper part of the human body is recorded. The dataset includes both RGB and depth videos consisting of 20 files [29]. There are 10 training videos and 37 test videos in each file. Figure 1 shows some frames from the training set in the dataset. The ChaLearn dataset consists of 9 categories. These are: 1) activity (writing, answering the phone), 2) pantomime (movement applied to imitation actions), 3) dance, 4) body language, 5) gestulations, 6) illustrators, 7) emblems (symbolic or political movements), 8) sign language, 9) signals (such as diving signals) [25,29]. Only depth videos of the dataset were used in the study. We used a Kinect v2 camera to build our dataset in which only the upper gestures of the human body were recorded. An interface for the camera was also designed. With this interface, the gesture video taken from the user was recorded as the frames in the dataset. Figure 2

shows some frames from the dataset. In the study conducted with the ChaLearn dataset, video was processed directly. Therefore, some unnecessary frames emerge causing a misclassification. However, due to the use of the direct frame in the dataset we have prepared, the unnecessary frames that were formed together do not appear in the study. This provides a more accurate classification. In addition, a gesture was repeated four times in the training set of the created dataset. Thus, a better training is provided to prevent misclassification resulting from camera shift.

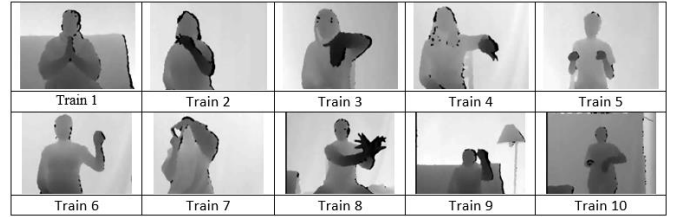


Figure 1: Some frames from the training dataset.

B. Summary of the Proposed Approach

The proposed approach focuses on four main processes. These are i. STD (standard deviation), ii. MHI (motion history image), iii. 2D FFT (2D fast fourier transform), iv. Distance measurements used to compare two images. These distance measurements are Levenshtein, Frobenius, and Mahalanobis.

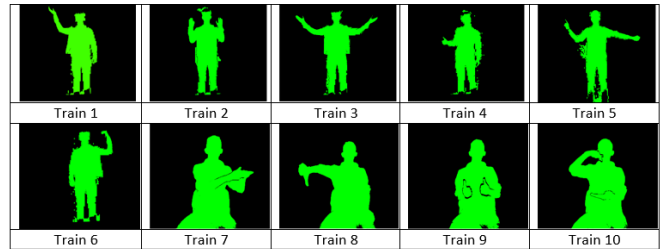


Figure 2: Some frames from the our dataset.

In the study, depth videos within the ChaLearn dataset were used. The usage of depth data gives two important advantages. Firstly, it gives the pixel density information in proportion to the distance of the object to the camera using grayscale. Secondly, it uses the grayscale threshold on each pixel of the frame to separate the foreground from the background in the frame [30]. In order to determine the threshold level, Otsu method is used to convert a grayscale image to a binary image [30,31,32]. As shown in Figure 3, using this method, it is assumed that the frame consists of two color classes as foreground and background. Then, the intra-class variance value of these color classes is calculated.

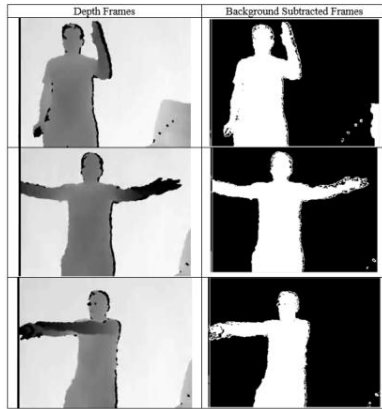


Figure 3: Otsu threshold used to depth frames.

The STD method is used to collect all the frames in the video in a single image. The method consists of 4 steps. These steps are: 1) All frames in a video are collected and divided by the total number of frames. Thus, an average frame is obtained. 2) Each frame in the video is subtracted from the average frame. 3) The squares of the obtained frames are calculated and collected. 4) The result is divided by the total number of frames in the video [33]. Thus, the gesture within a video can be found by the STD method. In Figure 4, some videos from the training set within the ChaLearn dataset are shown with this method.

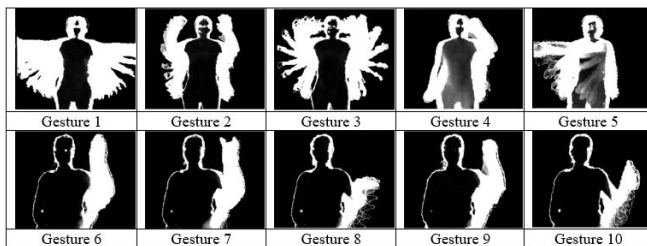


Figure 4: STD applied videos in the ChaLearn dataset.

Since the STD method does not indicate the direction of flow of the gesture, the history information of the gesture is lost. Therefore, MHI method is also used in the study. The MHI method protects history information of the frame [34,35]. This method presents two important advantages. Firstly, the method is insensitive to noises. Thus, the dominant gesture information is protected. Secondly, this method can also be applied in low light areas and low power CPUs. Figure 5 shows some videos from the training set within the ChaLearn dataset with this method.

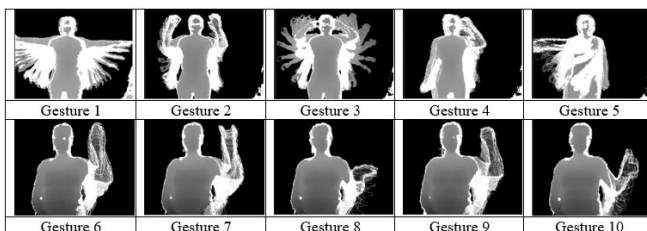


Figure 5: MHI applied videos in the ChaLearn dataset.

All the gestures in the ChaLearn dataset always return to their original positions. However, the camera is free to move for test samples, while it remains stationary during training samples. This causes incorrect matching. 2D FFT method is used to prevent misclassification [36]. 2D FFT is applied before STD and after MHI method. Thus, the effect of the camera is reduced and the matching between the training and the test data becomes easier.

In the proposed approach, 3 different distance measurements are used for the classification process. These are Levenshtein, Mahalanobis, and Frobenius. Levenshtein distance measurement performs addition, deletion and relocation between two arrays, two words or two sentences [37,38,39]. In this study, this distance measurement is used to compare the similarity between the two images.

The ChaLearn dataset is divided into training dataset and test dataset. There is only one gesture within each video in the training dataset. However, there may be more than one gesture in videos within the test dataset. For this reason, it is necessary to distinguish between multiple gestures in videos in the test set. The action boundary method is used to separate multiple gestures from one video [40,41].

III. EXPERIMENTAL RESULTS

In this study, firstly STD method was used. However, the direction information of the gesture is not preserved by this method as explained previously. Therefore, MHI method is also used to supply gesture history. In addition, the 2D FFT method has been used with both methods to prevent errors that may occur in camera shifts.

Two different datasets were used in the study. Each of the frames in the ChaLearn dataset consists of 240x320 matrices. Otsu threshold method is used for this dataset. Using the method, the frame consists of two classes. Then the intra-class variance value is calculated. When the variance value is minimum in the class, the variance value between the classes becomes maximum yielding faster and more accurate results. We did not use the Otsu threshold method in the study for our novel dataset. The ChaLearn dataset was created using the Kinect v1 camera, while Kinect v2 camera was used in the second dataset. An interface is designed in the second dataset in which the gesture video taken from the user is transferred directly to the file as an image with using this interface.

It is known that the improper separation of gestures causes a misclassification. Fortunately, all gestures in the ChaLearn dataset are returned to their initial position. This makes it easier to distinguish the different gestures within the video. Each of the training videos has a single gesture, while there are multiple gestures within the test videos. Therefore, firstly, the different gestures in the test videos need to be separated from each other. In the study, we used the action boundary method to separate multiple gestures in a video in the test dataset. In this method, the first frame of the video is compared with all other frames, and when the first frame matches the frame, all the frames up to that frame are accepted as a gesture. Figure 6

shows a graph of a video in the test dataset. There are 4 different gestures in this video. The action boundary method was not needed because the video was transformed directly into frames through an interface in our dataset.

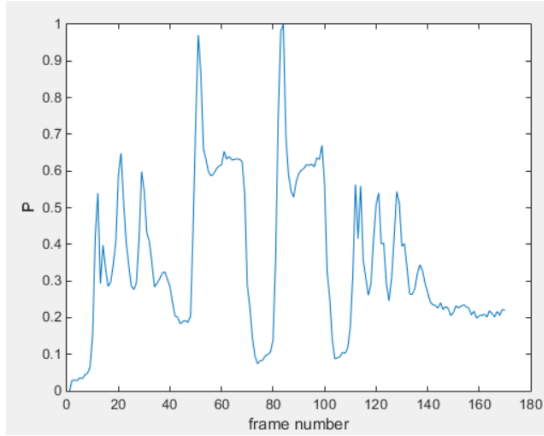


Figure 6: Action boundary method applied a video in the dataset.

Three different distance measurements were used to make the classification process. These distance measurements are Levenshtein, Mahalanobis, and Frobenius. Levenshtein distance measurement, which is still used in Google's infrastructure, presents a more accurate classification compared to other distance measurements. Levenshtein distance measurement is normally used for string expressions. However, we use it to compare two images. If the images in the training and test datasets are V and Y , respectively, the Levenshtein distance measurement is $L(V, Y)$. If $L(V, Y)$ is the minimum, then the images are similar. But if $L(V, Y)$ has the maximum value, these two images are regarded as different.

In the study, after the different gestures in the videos in the first dataset were separated from each other, 3 basic operations were used in 9 different ways. It was observed that the 7th method gave the most accurate classification result from these 9 methods. In all methods, firstly, Otsu threshold method was used. These methods are 1) After different gestures in the test dataset were separated, the training and the test datasets were

compared with each other by STD method. 2) The background was subtracted from the frames. Thus, the focusing on the gesture was provided. Finally, the training and the test datasets were compared with the MHI method only. 3) The camera shift causes a misclassification. For this reason, 2D FFT method was applied before comparing the training dataset and test dataset with STD method. 4) After the MHI method, 2D FFT method was applied to prevent misclassification. 5) Levenshtein distance measurement, which is one of the distance measurements, was used with the 3rd method. 6) Levenshtein distance measurement was used to match gestures along with the 4th method. 7) STD, MHI, 2D FFT methods, which are known as 3 basic operations, were used with Levenshtein distance measurement to make a more accurate classification. In the 8th and 9th methods, 3 basic operations were used together with Mahalanobis and Frobenius distance measurements respectively. The methods used in the study and their success percentages are shown in Table 1. All methods in the study was examined and it was seen that the 7th method had the highest success rate compared to the other methods. As a result, the most accurate classification was made by the 7th method. The second highest success rate was the 6th method because it retained gesture direction information.

In the study, the dataset that we prepared as the second dataset was also used. With this dataset, 3 basic operations were also classified in 9 different ways. Table 2 shows the methods and success rates obtained with the second dataset. In the study with the second dataset, the 7th method was found to be more successful. In addition, the success rates of the second dataset were higher than the first dataset. In the study conducted with the ChaLearn dataset, video was processed directly. Therefore, it was included in the study of unnecessary images. This has led to misclassification. However, due to the use of the direct image in our dataset, the unnecessary images that were formed together did not take place in the study. This provided a more accurate classification. In addition, a gesture was repeated four times in the training set of the created dataset. In this way, the wrong classification was prevented by providing better education.

Table 1: Summary of the proposed approach with the ChaLearn dataset.

Methods	1	2	3	4	5	6	7	8	9
Depth data	√	√	√	√	√	√	√	√	√
Background subtraction	√	√	√	√	√	√	√	√	√
STD	√								
MHI		√							
2D FFT before using STD			√		√		√	√	√
2D FFT after using MHI				√		√	√	√	√
Intelligent classification	√	√	√	√	√	√	√	√	√
Frobenius distance measurement								√	
Mahalanobis distance measurement									√
Levenshtein distance measurement					√	√	√		
Percentages of gesture recognition	50.22	53.69	51.68	56.64	55.53	63.55	76.15	41.70	52.20

Table 2: Summary of the proposed approach with our dataset.

Methods	1	2	3	4	5	6	7	8	9
STD	√								
MHI		√							
2D FFT before using STD			√		√		√	√	√
2D FFT after using MHI				√		√	√	√	√
Frobenius distance measurement								√	
Mahalanobis distance measurement									√
Levenshtein distance measurement					√	√	√		
Percentages of gesture recognition	53.92	58.22	55.45	63.33	60.10	72.47	84.12	49.65	56.17

I. CONCLUSION

The study aims to recognize and classify hand gestures with one-shot learning approach. Two different datasets were used in the study. These are the ChaLearn Gesture Dataset (CGD2011) and our novel dataset produced by us. The first dataset consists of videos. In this dataset, there are different gestures in each file. Therefore, different results were obtained from each file. The success rate in some files was lower than other files due to difficult finger gestures and unexpected gestures of the person. For this reason, different gestures in a video were separated from each other. Second, the background was extracted from the frames using the Otsu threshold method. Finally, 2D FFT, STD, and MHI methods, known as 3 basic operations, were used with various combinations. Three different distance measurements were used to classify gestures. These were Levenshtein, Frobenius, and Mahalanobis. Levenshtein distance measurement was observed to make a more accurate classification than other distance measurements. In the study, we used the dataset that we prepared as the second dataset. Kinect v2 camera was used for construction of this dataset. An interface was created for the dataset. With this interface, the gesture video taken from the user was transferred to the file as an image. In the study conducted with the ChaLearn dataset, video was directly processed. Therefore, it was included in the study of unnecessary frames. This situation led to misclassifications. However, due to the use of the direct frame in the dataset we have prepared, the unnecessary frames that were formed together did not take place in the study. This provides a more accurate classification. In addition, a gesture was repeated four times in the training set of the created dataset. In this way, the wrong classification was prevented by providing better education. In this study, 9 methods with 3 basic operations were applied to this dataset. However, due to the absence of unnecessary frames in dataset and 4 repetitions of a gesture in the training set, more successful results were obtained. Moreover, there was no need for action boundary method with the interface used in our dataset. We aim to recognize and classify finger gestures and all body gestures in future studies.

REFERENCES

- [1] N. C. Kılıboz, U. Gündükbay, "A hand gesture recognition technique for human-computer interaction", *Journal of Visual Communication and Image Representation*, vol. 28, pp. 97-104, April 2015.
- [2] J. Tang, H. Cheng, Y. Zhao, H. Guo, "Structured dynamic time warping for continuous hand trajectory gesture recognition", *Pattern Recognition*, vol. 80, pp. 21-31, August 2018.
- [3] M. A. R. Ahad, J. K. Tan, H. S. Kim, S. Ishikawa, "Human activity recognition: various paradigms", in *Internat. Conf. on Automation and Systems (ICCAS) 2008*, pp. 1896-1901.
- [4] L. Baraldi, F. Paci, G. Serra, L. Benini, R. Cucchiara, "Gesture recognition using wearable vision sensors to enhance visitors' museum", *IEEE Sensors Journal*, vol. 15, pp. 2705-2714, March 2015.
- [5] C. Wang, Z. Liu, M. Zhu, J. Zhao, S.-C. Chan, "A hand gesture recognition system based on canonical superpixel-graph", *Signal Processing: Image Communication*, vol. 58, pp. 87-98, October 2017.
- [6] P. Hong, M. Turk, T. Huang, "Gesture modeling and recognition using finite state machines", in *IEEE Conference on Automatic Face and Gesture Recognition 2000*, pp. 410-415.
- [7] J. Beh, D. K. Han, R. Durasiwami, H. Ko, "Hidden markov model on a unit hypersphere space for gesture trajectory recognition", *Pattern Recognition Letters*, vol. 36, pp. 144-153, January 2014.
- [8] X. Gong, L. Han, J. Wang, M. Ran, "Recognition and simulation of parachute action based on continuous hidden markov model", *Chinese Automation Congress (CAC) 2017*.
- [9] P. Premaratne, S. Yang, P. Vial, Z. Iftikhar, "Centroid tracking based dynamic hand gesture recognition using discrete Hidden Markov Models", *Neurocomputing*, vol. 228, pp. 79-83, March 2017.
- [10] S. Belgacem, C. Chatelain, T. Paquet, "Gesture sequence recognition with one shot learned CRF/HMM", *Image and Vision Computing*, vol. 61, pp. 12-21, May 2017.
- [11] H.-II Suk, B.-Kee Sin, S.-Whan Lee, "Hand gesture recognition based on dynamic Bayesian network framework", *Pattern Recognition*, vol. 43, pp. 3059-3072, September 2010.
- [12] H. Cheng, Z. Dai, Z. Liu, Y. Zhao, "An image-to-class dynamic time warping approach for both 3D static and trajectory hand gesture recognition", *Pattern Recognition*, vol. 55, pp. 137-147, July 2016.
- [13] A. H.-Vela, M. Á. Bautista, X. P.-Sala, V. P.-López, S. Escalera, X. Baró, O. Pujol, C. Angulo, "Probability-based dynamic time warping and bag-of-visual-and-depth-words for human gesture recognition in RGB-D", *Pattern Recognition Letters*, vol. 50, pp. 112-121, December 2014.
- [14] J. Wan, Q. Ruan, W. Li, "One-shot learning gesture recognition from RGB-D data using bag of features", *Journal of Machine Learning Research*, vol. 14, pp. 2549-2582, 2013.
- [15] J. L. Raheja, S. Subramaniyam, A. Chaudhary, "Real-time hand gesture recognition in FPGA", *Optik*, vol. 127, pp. 9719-9726, October 2016.
- [16] R. Krishnan, S. Sarkar, "Conditional distance based matching for one-shot gesture recognition", *Pattern Recognition*, vol. 48, pp. 1302-1314, April 2015.
- [17] J. Konecny, M. Hagara, "One-shot-learning gesture recognition using HOG-HOF features", *Journal of Machine Learning Research*, vol. 15, pp. 2513-2532, 2014.

- [18] A. F. Bobick, J. W. Davis. (2001, March). The recognition of human movement using temporal templates. *IEEE Trans. Pattern Anal. Machine Intell.* (23)3. pp.257-267.
- [19] C. Yang, D. K. Han, H. Ko, "Continuous hand gesture recognition based on trajectory shape information", *Pattern Recognition Letters*, vol. 99, pp. 39-47, November 2017.
- [20] M. A. R. Ahad, (2011). *Computer vision and action recognition: a guide for image processing and computer vision community for action understanding*. Atlantis Ambient and Pervasive Intelligence. Atlantis Press.
- [21] H. Zhou, D. Tao, Y. Yuan, X. Li, "Object trajectory clustering via tensor analysis", in *16th IEEE Internat. Conf. on Image Processing (ICIP) 2009*, pp. 1925-1928.
- [22] Z. Jiang, Z. Lin, L. S. Davis. (2012, March). Recognizing human actions by learning and matching shape-motion prototype trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol. 34.
- [23] U. Mahbub, H. Imtiaz, T. Roy, M. S. Rahman, M. A. R. Ahad, "A template matching approach of one-shot-learning gesture recognition", *Pattern Recognition Letters*, vol. 34, pp. 1780-1788, November 2013.
- [24] A. Bobick, J. Davis, "An appearance-based representation of action", in *Proc. 1996 International Conference on Pattern Recognition (ICPR'96)*, vol. 7270, pp. 307-312.
- [25] S. Ruffieux, D. Lalanne, E. Mugellini, O. A. Khaled. (2014). *A survey of datasets for human gesture recognition*. From book: Human-Computer Interaction Advanced Interaction Modalities and Techniques. vol. 8511, pp. 337-348.
- [26] G. Modanwal, K. Sarawadekar, "Towards hand gesture based writing support system for blinds", *Pattern Recognition*, vol. 57, pp. 50-60, September 2016.
- [27] Y. Song, D. Demirdjian, R. Davis, "Tracking body and hands for gesture recognition: NATOPS aircraft handling signals database", in *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG2011)*. 2011.
- [28] V. Athitsos, C. Neidle, S. Sclaroff, J. Nash, "The American Sign Language Lexicon Video Dataset", in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2008.
- [29] J. Wan, S. Z. Li, Y. Zhao, I. Guyon, S. Escalera, "ChaLearn looking at people RGB-D isolated and continuous datasets for gesture recognition", in *IEEE Conference Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2016.
- [30] P. P. Vijay, N. C. Patil, "Gray scale image segmentation using Otsu thresholding optimal approach", *Journal for Research*, vol. 02, July 2016.
- [31] C. Yu, C. Dian-ren, L. Yang, C. Lei, "Otsu's thresholding method based on gray level-gradient two-dimensional histogram", *2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR)*, April 2010.
- [32] V. I. Levenshtein (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, vol. 163, pp. 845-848.
- [33] S. S. Rautaray, A. Agrawal, "Vision based hand gesture recognition for human computer interaction: survey", *Artificial Intelligence Review*, vol. 43, pp. 1-54, January 2015.
- [34] M. A. R. Ahad, J. K. Tan, H. Kim, S. Ishikawa, "Motion history image its variants and applications", *Machine Vision and Applications*, vol. 23, pp. 255-281, March 2012.
- [35] U. Mahbub, T. Roy, M. S. Rahman, H. Imtiaz, S. Serikawa, M. A. R. Ahad, "One-shot-learning gesture recognition using motion history based gesture silhouettes", in *International Conference on Industrial Application Engineering (ICIAE)*, 2012.
- [36] D. Wu, F. Zhu, L. Shao, "One shot learning gesture recognition from RGBD images", in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2012.
- [37] R. C. Gonzalez, R. E. Woods. (2001). *Digital Image Processing*. Addison-Wesley Longman Publishing Co. Inc.
- [38] Z. Feng, B. Yang, N. Lv, T. Xu, J. Yin, X. Zhao, S. Feng, "Motion-towards-each-other-based hand gesture initialization", *Pattern Recognition*, vol. 48, pp. 4049-4056, December 2015.
- [39] S. Konstantinidis, "Computing the edit distance of a regular language", *Information and Computation*, vol. 205, pp.1307-1316, September 2007.
- [40] R. P. Sharma, G. K. Verma, "Human computer interaction using hand gesture", *Procedia Computer Science*, vol. 54, pp. 721-727, 2015.
- [41] J. Wan, G. Guo, S. Z. Li, "Explore efficient local features from RGB-D data for one-shot learning gesture recognition", in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, Aug. 2016.

Local statistical features for multilingual artificial text detection from video images

A. JAMIL¹, J. RASHEED¹, B. BAYRAM²

¹ Istanbul Sabahattin Zaim University, Istanbul/Turkey, akhtar.jamil@izu.edu.tr

¹Istanbul Sabahattin Zaim University, Istanbul/Turkey, jawadrasheed@gmail.com

² Yildiz Technical University, Istanbul, Turkey, bayram@yildiz.edu.tr

Abstract— Textual information present in videos or images offers very compact and accurate form of information that can be exploited for semantic analysis of multimedia content. This paper exploited statistical features for detection of multilingual horizontally aligned artificial text from video images. First, local entropy of the image in a local neighborhood was calculated. The local entropy is relatively higher in text regions as compare to the non-text regions. However, some complex backgrounds exhibiting similar features as that of text were also included as candidate text regions. To overcome this scenario, a local gradient difference operation is performed to the input image and then averaged with entropy information. Some morphological operations were then performed followed by geometrical constraints and heuristics to suppress noisy element which were not potential candidates for text. Moreover, to improve the overall precision of the proposed systems, the text lines were detected using the horizontal projection profiles. Finally, adaptive thresholding was applied to extract the text components. Experiments were conducted with various video images to evaluate the proposed method and it showed promising results in terms of recall, precision, and F-measure.

Keywords— text detection, local entropy, gradient difference, projection profiles

I. INTRODUCTION

As the amount of devices connected to internet continues to grow (e.g., smart phone, hand held devices, digital cameras etc.), the amount of multimedia data produced and made available online has also increased exponentially. In addition to the audio visual content, the text present in the video images provides an important clue that can be exploited for developing video indexing and retrieval systems. A number of methods have been proposed in the past for text detection and localization from multimedia data. However, to cope with the challenges raised by diverse amount of data requires efficient techniques for automatic textual content detection, localization and extraction. We propose local statistical measured based on entropy and gradient difference for horizontally aligned artificial English and Urdu text detection from video images.

Generally, text present in videos falls into one of the two categories: (a) artificial text – which is manually added to the video by a user to give more semantics to it, and (b) scene text - which is captured naturally as part of the video. Artificial text usually appears in either horizontal or vertical direction with high contrast to make it more readable by users. On the other hand, since scene text is captured naturally as part of a video so its nature is unpredictable. Moreover, it has low resolution, low contrast, arbitrary

orientations, non-uniform illumination, and suffers from perspective distortion [1].

Text detection, localization and extraction from video images are not a new problem for semantic analysis of multimedia content as several methods exists for text detection. Some of the work done for text detection from multimedia content is presented below. A detailed survey on state of the art approaches for text detection, localization and recognition can be found in [2]. A new approach is proposed in [3] based on wavelets for text detection from video images followed by multi-frame verification for text localization. Edge-based features is used in [4] for detecting artificial Urdu text from video images. The system calculated vertical gradients and then averages the gradient magnitude in a fixed neighborhood of each pixel. The resulting image is binarized and the horizontal run length smoothing algorithm (RLSA) is applied to merge possible text regions. Finally, to eliminate the noise, edge density filter and geometrical constraints were applied to remove false positives. Local thresholding and hysteresis edge recovery is also used in [5] for text detection. Moreover, adaptive thresholding, dam point labeling, and inward filling is used for text localization and extraction. Likewise [6] also applied a similar approach for Farsi/Arabic text detection from video images. In the final step, they employ Local Binary Pattern (LBP) based features and support vector machine for text classification.

A novel approach is proposed in [7] based on optical flow of text for detecting dynamic cured text appearing in video images. These characteristics include constant velocity, uniform magnitude distribution and unique angle distribution along with k-means clustering algorithm to identify candidate text regions. In their other work, [8] they employee gradient vector flow (GVF) and neighbor component grouping for multi oriented text detection from video images.

A generic method for multilingual text detection in [9] comprises of two phases. To extract text information from background, Laplacian of Gaussian is applied on gray level of image to detect edges and noise. The resultant is then transformed to obtain possible text candidate by filling holes between two lines structure. Finally, the full connected component analysis is used by taking the difference of edge information and hole filled region to find complete region of text. Like other text edge detection techniques, in [10] a distributed canny edge detector is used along with edge preserving Maximally Stable Extremal Regions (eMSER) to enhance the accuracy of text detection by parallel processing of partitioned video frame segments.

An end-to-end system in [11] takes input image, convert it into gray scale and binarized it before eliminating non text line components. Furthermore, on vertical profile one-dimensional difference of Gaussian filter (DOG) is applied to obtain the text lines. It removes the false lines in successive word recognition step, whose average eccentricity of connected components are below defined threshold, to get the final text.

A local image operator proposed in [12] is based on Fast Stroke Width Transform (FSWT) for multilingual text detection from video images. First, for a given input image, the stroke width is determined then proceeded to determine connected components based on stroke width information. Finally, the connected components are classified as either text components or background. In [1] gradient directional and spatio-temporal information is exploited for multi-oriented and multi-scale text detection and tracking from videos. [13] proposed a unified framework for text detection, localization and tracking from videos. For text detection stroke map information is used while strong edge information facilitates localization of individual text lines. For text tracking in consecutive video frames, SURF features have been exploited.

Recently, the text detection research has also entered in era. Beside these conventional approached such as edge detection or stroke width transformation, new convolutional and deep neural network based algorithms are adopted that directly target text detection as in [14, 15, 16, 17, 18].

In [14], a trained network targets the text detection prediction by single light weighted neural network using deep Full Convolutional Neural Network (FCN) based pipeline model and detector named as Efficient and Accuracy Scene Text (EAST) detection pipeline. Similarly, in [15], convolutional neural network is combined with discrete cosine transform (DCT) for text detection. The image is divided into fix number of blocks in order to apply DCT and with the help of zig-zag scheme, Huffman encoding and other decoding mechanism the text is detected.

Another novel approach is presented in [18] to detect candidate text regions by corner response feature map, which is then divided into candidate lines in order to apply proposed fuzzy c-means clustering-based separation algorithm to get clear text layer, so that optical character recognition can recognize it easily.

This paper proposes a new method based on local entropy and gradient difference for horizontally aligned artificial text from video images. The rest of the paper is organized as follows. In section II this research's proposed work is stated. Section III describes the experimental results and finally the concluding remarks highlighting future research directions.

II. PROPOSED METHODOLOGY

A. This proposed methodology focuses on exploiting entropy and gradient difference for text detection from video images. These two outputs are then averaged and some Image Analysis

As the textual areas contain a number of edge content, this edge information should be linked to form a text line. Therefore, we employee morphological processing (dilation) to combine these edges into a text line. However, the dilation

not online combines the textual information, but some noisy elements are also added thereby producing undesirable results. To overcome these issues some geometrical constrains and heuristics are introduced.

morphological operations are then performed followed by applying some geometrical constraints and heuristics to suppress the unwanted background noise. The text lines are separated by applying horizontal projections. Finally, adaptive thresholding is performed to extract the text objects. The schematic diagram of the proposed methodology is shown in Fig. 1.

B. Local Entropy

Entropy of an image provides statistical measure of randomness of pixel values in a neighborhood. Following Shannon's formula is used to calculate the entropy of the image with L gray levels and P_k as probability of the k^{th} gray level.

$$\text{Entropy} = -\sum_{k=0}^{L-1} P_k \log_2(P_k) \quad (1)$$

Where $P_k = \frac{n_k}{M \times N}$ and n_k is the number of pixels in the neighborhood $M \times N$. In our case, we considered 7×7 neighborhoods for text detection which was empirically calculated.

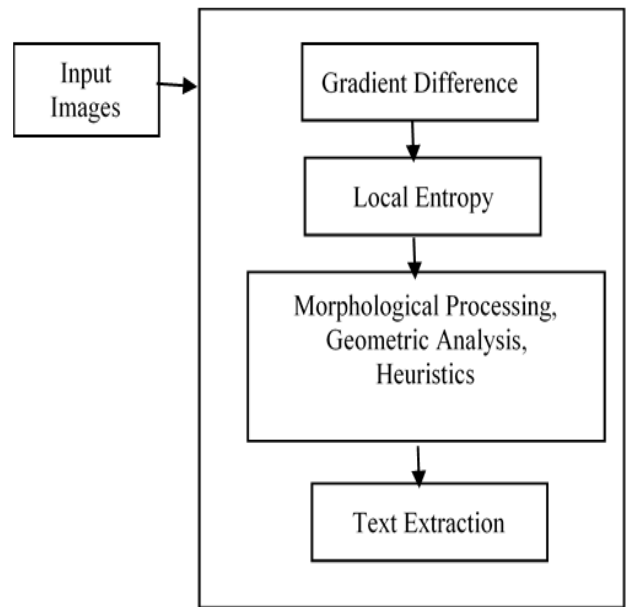


Fig. 1. Schematic diagram of the proposed method

C. Gradient Difference

To compensate for the information missed by local entropy information, we calculated the local gradient difference in a fixed neighborhood and then averaged with local entropy to highlight textual areas. This work is motivated by [19], which employ gradient difference operator in a local neighborhood for detection of text candidate regions. The gradient operator finds the difference between the maximum and minimum intensity values in the local neighborhood. The output is then averaged with the output of the entropy filter. The output image contains

brighter text regions while background is suppressed. Fig. 2 shows the results obtained for above steps.



(a)



(b)



(c)



(d)

Fig. 2. a) Input image, b) Local entropy, c) Gradient difference, d) Average of (b) and (c).

Since, we are interested in horizontally aligned text; therefore, we assume that the width of the bounding box should be greater than the height of the box to make sure that the bounding box contains at least two words. Heuristics were determined empirically such as minimum height and width

of the bounding box. Bounding boxes not satisfying both geometric and heuristics were discarded considering them as non-text part.

D. Text Line Detection

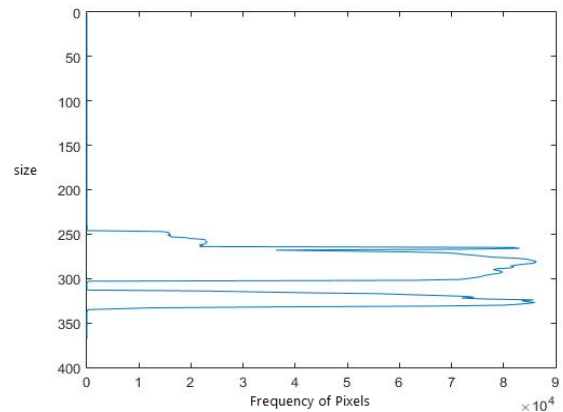
In some situation when there were more than one text lines in a video frame, the morphological operations caused multiple lines to overlap with each other and they become part of one bounding box. Empirically, we evaluated that this can happen in two situations: either text lines are spatially very close or a background element with high contrast connects the text line during morphological operation. This affects the overall precision of the system. Therefore, to remove these false alarms we applied horizontal projections to separate text lines and draw bounding boxes separately around each text line. Fig. 3 shows text localization and the horizontal project of the text lines.

E. Local Thresholding

The text extraction is an important part of text detection process. Selection of a good threshold value plays a crucial role. A very large value of threshold will result in loss of some important information, while a very small value of threshold may result in some unwanted information as part of the text. In this paper we take Sauvola's method [20] for thresholding in a local neighborhood. It is computed using the mean (μ) and standard deviation (σ) for pixel at (x,y) in a window centered around the pixel at (x, y) as:



(a)



(b)

Fig. 3. a) Text localization and line separation, b) Horizontal text lines projection profile.

$$T(x, y) = \mu(x, y) * \left[1 + k \left(\frac{\sigma(x, y)}{R} - 1 \right) \right] \quad (2)$$

Where R is the maximum value of the standard deviation and k is a parameter which takes positive values in the range [0.2, 0.5].

III. EXPERIMENTAL RESULTS

In this survey we aim to identify the mediating effect of learning orientation on the relationship between leadership style and firm performance. To test the propositions, a field survey using questionnaires was conducted.

A benchmark data set containing horizontally aligned artificial text for both Urdu and English languages is used for evaluation [21]. These videos are captured from variety of sources mainly news channels. Some video images were also taken from sports channels, stock exchange and cartoon related TV channels containing textual information. For evaluation, we selected 600 images. The English dataset consisted of 300 images while for Urdu we used 300 images containing horizontally aligned artificial text. Fig. 4 shows a sample output obtained from the proposed method.

Precision, recall and f-measure are used as the standards for evaluating the proposed method. In our method, we also utilized these measures for evaluating the proposed system. Our method for text detection achieved an overall recall of 84%, precision 85% and f-measure of 87% as summarized in Table 1 for each language dataset.

TABLE I. PERFORMANCE OF THE PROPOSED METHOD FOR ENGLISH AND URDU DATASETS

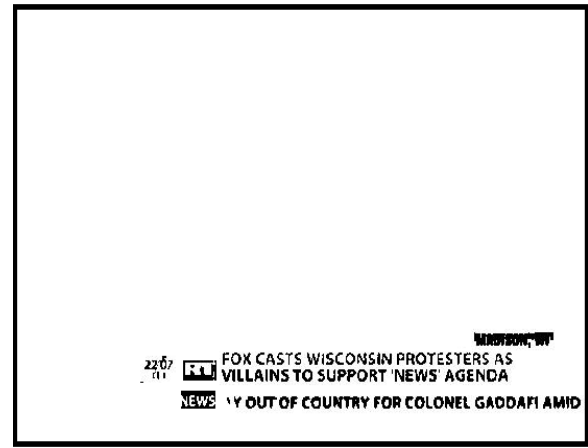
Language	Performance Results			
	Dataset	Precision	Recall	F-measure
English	300	0.86	0.91	0.88
Urdu	300	0.83	0.88	0.85

A. Experiment on Processing Time

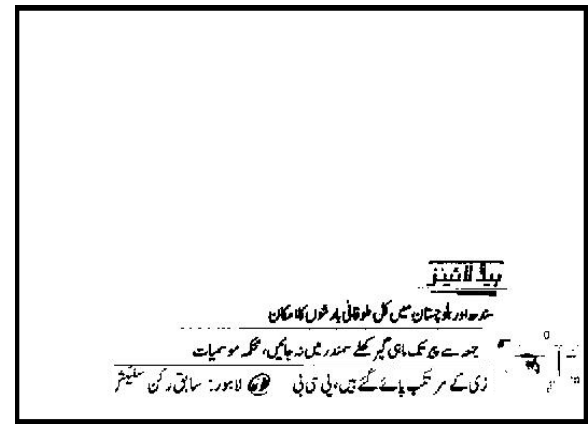
It was observed that there is a slight difference between precision for English and Urdu text, which is due to non-uniform spacing between text elements in case of Urdu text and number of textual elements contained in the video images. Sometimes, multiple bounding boxes were drawn in the same text line thereby affecting the precision of the system, while in English, as there is uniform spacing between words in a text line, so the precision is more.

IV. CONCLUSION

This paper, proposed a new technique using statistical measures based on local entropy and Laplacian of Gaussian for horizontally aligned artificial text detection from videos and images. A good detection rate was obtained on datasets for both English and Urdu. The method begins by calculating local entropy and LoG then averaging them to get rid of noisy elements. Further, application heuristics and geometrical constraints followed by some morphological processing helped suppressing noise. Finally, adaptive thresholding was employed for textual content extraction from video images.



(a)



(b)

Fig. 4. Text extraction a) English text, b) Urdu text.

In future we would like to make it more robust by integrating a machine learning approach like SVM or ANN. Moreover, this method can also be extended for other languages by carrying experiments on datasets of different languages, which is our aim in future.

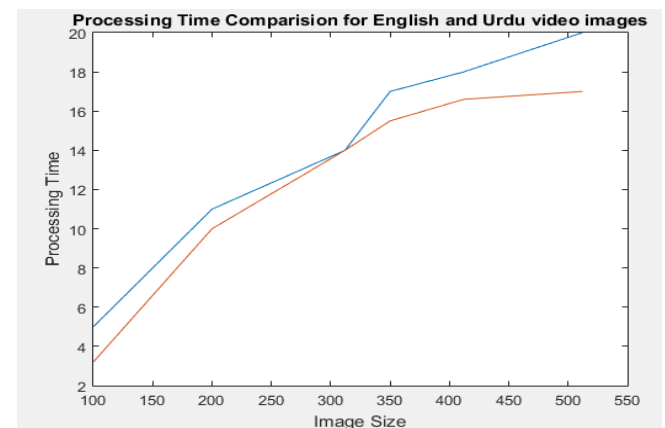


Fig. 5. Average processing time for English and Urdu video

REFERENCES

- [1] W. He, X. Y. Zhang, F. Yin, and C. L. Liu, "Multi-Oriented and Multi-Lingual Scene Text Detection with Direct Regression," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5406–5419, 2018.

- [2] Q. Ye and D. Doermann, "Text Detection and Recognition in Imagery: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 7, pp. 1480-1500, 2015.
- [3] X. Huang, "Automatic Video Text Detection and Localization Based on Coarseness Texture", *Fifth International Conference on Intelligent Computation Technology and Automation*, 2012.
- [4] A. Jamil, I. Siddiqi, F. Arif and A. Raza, "Edge-Based Features for Localization of Artificial Urdu Text in Video Images", *International Conference on Document Analysis and Recognition*, 2011.
- [5] M. Lyu, Jiqiang Song and Min Cai, "A comprehensive method for multilingual video text detection, localization, and extraction", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 2, 2005, pp. 243–255.
- [6] M. Moradi and S. Mozaffari, "Hybrid approach for Farsi/Arabic text detection and localisation in video frames", *IET Image Processing*, vol. 7, no. 2, 2013, pp. 154-164.
- [7] P. Shivakumara, M. Lubani, K. Wong and T. Lu, "Optical flow based dynamic curved video text detection", *2014 IEEE International Conference on Image Processing (ICIP)*, 2014.
- [8] P. Shivakumara, T. Phan, S. Lu and C. Tan, "Gradient Vector Flow and Grouping-Based Method for Arbitrarily Oriented Scene Text Detection in Video Images", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, 2013, pp. 1729-1739.
- [9] H. T. Basavaraju, V. N. M. Aradhya, and D. S. Guru, "A novel arbitrary-oriented multilingual text detection in images/video," in *Information and Decision Sciences, Advances in Intelligent Systems and Computing*, vol. 701, S. C. Satapathy, J. M. R. S. Tavares, V. Bhateja, J. R. Mohanty, Eds. Springer, 2018, pp. 519–529.
- [10] S. A. Pote, and M. A. Mehta, "An improved technique to detect text from scene videos," *IEEE International Conference on Communication and Signal Processing (ICCSP)*, 2017, pp. 1190–1194.
- [11] W. Ohyama, S. Iwata, T. Wakabayashi, and F. Kimura, "Detection and recognition of arabic text in video frames," *IEEE 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 07, 2017, pp. 20–24.
- [12] Jeong, Munho, and Kang-Hyun Jo. "Multi language text detection using fast stroke width transform." *Frontiers of Computer Vision (FCV)*, 21st Korea-Japan Joint Workshop on. IEEE, 2015
- [13] T. Yusufu, Y. Wang and X. Fang, "A Video Text Detection and Tracking System", *IEEE International Symposium on Multimedia*, 2013.
- [14] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, "EAST: An efficient and accurate scene text detector," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2642–2651.
- [15] P. M. Osina, Y. A. Bolotova, and V. G. Spitsyn, "Text detection algorithm on real scenes images and videos on the bases of discrete cosine transform and convolutional neural network," *IEEE International Siberian Conference on Control and Communication (SIBCON)*, 2017, pp. 1–4.
- [16] W. Huang, Y. Qiao, and X. Tang, "Robust scene text detection with convolution neural network induced msr trees," in *Proceedings of Computer Vision ECCV*, 2014.
- [17] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Reading text in the wild with convolutional neural networks," *International Journal of Computer Vision*, vol. 116, 2016.
- [18] W. Lu, H. Sun, J. Chu, X. Huang, and J. Yu, "A novel approach for video text detection and recognition based on a corner response feature map and transferred deep convolutional neural network," *IEEE Access*, vol. 6, 2018, pp. 40198–40211.
- [19] P. Shivakumara, T. Phan and C. Tan, "A Gradient Difference Based Technique for Video Text Detection", *10th International Conference on Document Analysis and Recognition*, 2009.
- [20] J. Sauvola, T. Seppanen, S. Haapakoski, and M. Pietikainen, "Adaptive document binarization," in *Proceedings of the Fourth International Conference on Document Analysis and Recognition*, vol. 1, 1997, pp. 147–152.
- [21] I. Siddiqi and A. Raza, "A Database of Artificial Urdu Text in Video Images with Semi-Automatic Text Line Labeling Scheme", *Thinkmind.org*, 2012.

Real Time Distributed Controller For Delta Robots

Ali Sharida¹, Iyad Hashlamon

Mechanical-Engineering department, Palestine Polytechnic University, Hebron, Palestine. Emails:
156049@ppu.edu.ps, iyad@ppu.edu

Abstract— This paper investigates a real time distributed controller for a 3 DOF delta robot. This robot is equipped with 3 servo motors, each motor has an attached encoder. The approach aims to use parallel computing to implement control methods. Hence, the computation is distributed among 4 microcontrollers (MCU's), these MCU's are connected to each other using CAN bus protocol. Each MCU is connected to one actuator and its attached encoder to form an Intelligent sensor-actuator system (ISAS). The fourth MCU is used to compute the control law. At each sample time, the ISAS broadcast a message using the CAN bus to the MCU containing the information about the motor position. Then according to the control law, ISAS receives the corresponding controller value that has to be applied to the motor. All required periodic, aperiodic and sporadic tasks, were implemented and will be handled by these MCUs. Using this design, the computation time of control law can be minimized. More, this method increases the flexibility of the system. The results show the applicability of the proposed distributed controller.

Keywords—Real Time Control; Distributed Control; Delta Robot

1. INTRODUCTION

Delta robots are widely used in applications that require very fast motion and accuracy, such as picking and placing [1, 2]. The main advantage of these robots is the ability to produce high acceleration at the end effector. Furthermore, as the mass of the overall system is relatively low, these robots can achieve a high load capacity. Therefore, this robot attracted many researchers to develop dynamic models and controllers [3-7].

However, the modeling of this robot is complex, which makes the design of the controller more complex and time consuming. The comparative study in [8] illustrates the processing time of the fundamental three approaches of modeling Delta robot (Principle of Virtual Work, the Newton-Euler Formulation, and the Lagrangian Formulation). Although the results were Principle of Virtual Work 0.73 sec, Newton-Euler Formulation 1 sec, and Lagrangian Formulation 0.37 sec, these processing times are high for real time systems. Even more, it becomes more difficult when the processing time of reading the sensors in parallel, computing the control law of three motors in parallel, and so forth are considered. Further, it is a real challenge when microcontrollers such as PIC-microcontroller are considered to be used for education purposes with adaptive control approaches.

To overcome the computational cost, distributed control approaches were used. In these approaches, several microcontrollers (MCU's) are used, each controller is assigned

its own job(s). The microcontrollers communicate with each other using a communication protocol to form an overall real time Network Control System NCS.

Many protocols can be used to establish a real time network, such as I2C [9], SPI [10] and CAN protocol [11]. Among them, CAN protocol has many advantages including very simple physical construction, it supports auto retransmission of lost messages and it supports different error detection capabilities. Therefore, it is considered the most suitable communication method for real time applications [4].

A distributed controller is reported in [12], where the controller is designed to control a slave robot from a master arm using SPI protocol. In [13], the researcher implemented an embedded controller for 5 DOF manipulator using SPI protocol using a simple PID controller. Although SPI is a very simple protocol and it depends on the principle of master and slave communication, any fault in the master MCU will lead to shut down all network. Furthermore, SPI communication requires more signal lines than other communication protocols, which increase the complexity of the network.

The principle of server and client communication is used in [14], where the server is an operating system running controller, and the client is a computer with WINDOWS operating system. This structure is very efficient for single controlled client, however, in robotics more than 1 client are required usually. So, this method is very costly for applications that require multiple clients or nodes.

This paper proposes a real time control for the 3 DoF delta robot. It uses 4 MCU's each one consists of a microcontroller and CAN bus receiver-transmitter. One MCU is used to compute the control law. The other three MCU's are connected to the actuated joints through an electronic interfacing modules, each one of the three MCU's along with the actuator and sensory system forms an intelligent sensor-actuator-system ISAS. Each ISAS is connected to one actuator and one sensor i.e. one DoF. Further, it can communicate with the other ISAS's and the controller MCU through CAN bus communication protocol. The ISAS reads the actuator position through an encoder, forms the necessary signal processing and prepare the ready measured data in message and broadcast it to the CAN bus. This message will be received by the beneficiary MCU message, and in the same way from all ISAS's. The controller MCU computes the required control law and broadcast it on the CAN bus. Each ISAS will receive its own message and skip the others, and then each ISAS analyses the message and applies the required signal on the actuator. This approach enhances flexibility to the system for changing the control approach and adding other jobs in addition to

distributing the computational load among 4 MCU's. The work is carried on using MATLAB with the TrueTime toolbox [15]. The overall block diagram of this approach is shown in Figure 1.

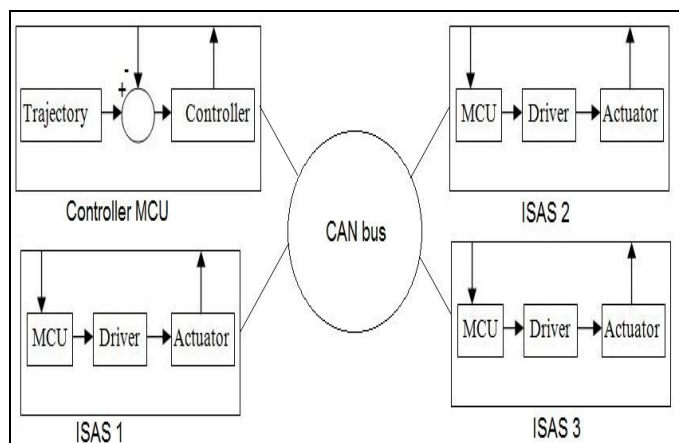


Figure 1. Overall block diagram

The rest of the paper is organized as follows: Section 2 The real time design. The results and the paper is concluded in section 3 and 4 respectively.

2. REAL TIME SYSTEM DESIGN

The real time NCS design starts by assigning the tasks with their specifications. For this system there are three motors to be actuated and controlled. In order to perform the control law, the following data is required: the reference signal or trajectory that the end effector must track, the forward kinematics model to transform the joint space variables to the task space variables, inverse kinematic model to transform the task space variables to joint space variables, the robot dynamic and inverse dynamic model to compute the control signal, a controller, and some computations in addition to sending and receiving data through the CAN bus. The control is divided into 12 different tasks as follows:

Task 1: it is responsible to compute the angular position and velocity of each actuator from the signal measured from an encoder attached to the actuator. This task is sporadic as the position should be correctly computed within its time limits, any delay in this task will cause drifting in position computation, which will lead to instability of the entire system.

Task 2: it is responsible to read the desired inputs from the user. This task is aperiodic task, as the user will not apply these inputs periodically. Furthermore, there is no matter if these inputs were used to compute control law in the next cycle.

Task 3: it is a periodic task, it has the jobs which perform the forward kinematics. This task depends on the results from Task 1, such that, it takes the computed angles as inputs, then it computes the related position of the end effector. This task should be executed on each program cycle, so it is a periodic task (Task 4 to 9 are periodic for the same reason).

Task 4: In this task, the inverse kinematics is computed, which will be used later to generate the required feedback variables for the controller.

Task 5: Jacobian matrix performs a transformation of the velocities from joint space to work space. This task is responsible to compute this matrix in order to compute the velocity of the end effector.

Task 6: In this task, the inverse of Jacobian will be computed, in order to get the required angular velocities of the actuators that required to move the end effector in a specific manner.

Tasks 7 & 8: These tasks are responsible of computing the required actuators torques to control the motion of the end effector.

Task 9: The aim of this task is to provide a suitable reference input signal, which will ensure that the end effector will move as the desired manner with soft start and stop.

Tasks 10 & 11: These task are responsible of sharing data among the MCU's. They are an on change based tasks, which will be enabled when the current position or the computed torque is changed. They should be executed directly when their flags are enabled to ensure that the control law will be computed correctly on any change of links kinematics.

Task 12: This is an aperiodic task, as it will be executed when the computed torque is changed. Therefore, it will apply torque signal to the actuator by computing the related voltage.

2.1. TASKS SCHEDULING AND SCHEDULABILITY TEST

The scheduling of the tasks is based on their type and timing. The periodic tasks were scheduled using the fixed priority algorithm Rate monotonic (RM). Whereas the aperiodic and sporadic tasks were scheduled by the principle of servers. In real time systems multiple servers can be used to handle the non-periodic tasks such as Bandwidth-preserving servers and sporadic [16]. In this project the non-periodic tasks were classified as the following:

1. Event based tasks

In general, an event based task enables a flag when it is released. This flag can be assumed to be a global flag that can be noticed in any part of the algorithm and can directly preempt the current executing task and jumps to a service routine. This type of tasks will be executed directly when its flag is on, and the microcontroller will interrupt any executing task in this case. The task was assumed to be non-pre-emptable and will continue executing until it is finished.

2. On change based tasks:

This type also contains a flag that will be turned on when the task is released. However this flag should be tested periodically to determine if the task is released or not. Depending on the result of the check, the algorithm decides if an interrupt is required or not. These tasks have a higher priority than the periodic tasks but less than the event based

tasks. This type was handled using Periodic Server algorithm, which creates a periodic server and schedules it using RM with the highest priority.

2.2. TASK EXECUTION TIME

The execution time of each task was computed experimentally using oscilloscope, each task was implemented individually, a hardware flag was turned on at the beginning of the execution and turned off at its end, then the on (digital 1) interval was captured. A summary of the tasks is in Table 1.

2.3. SCHEDULABILITY TEST

The worst case execution time (WCET) is considered as the time required to complete all the tasks if they released at the same instant without violating their constraints. Using the information listed in TABLE 1, WCET is 6.7 ms. To avoid processor over loading, the period time was adjusted to 8.5 ms. Each periodic task is released at the beginning of the period, these tasks should be executed before the deadline which presents the end of current period.

This paper is a course project, the sampling time was selected to be 8.5 ms. As the controller is distributed, all tasks will be executed before the deadline, as the maximum time occurs at the computation of control law which requires 6.5 ms, which is schedulable based on RM utilization factor test, as the utilization factor is about 76%.

TABLE I. TASKS SPECIFICATIONS

Tasks	Specifications		
	Name	Type	Execution Time (ms)
1	Compute system states	Sporadic	0.1
2	Read reference signal	Aperiodic	0.3
3	Forward Kinematics	Periodic	0.9
4	Inverse Kinematics	Periodic	0.4
5	Jacobian	Periodic	0.6
6	Inverse Jacobian	Periodic	0.9
7	Dynamics	Periodic	0.8
8	Inverse Daynamics	Periodic	1.1
9	Trajectory	Periodic	0.2
10	Data transmission (CAN bus)	Sporadic	0.6
11	Data reception (CAN bus)	Sporadic	0.7
12	Apply outputs	Aperiodic	0.1

2.4. DATA HANDLING AND FRAMES CREATION

The previous tasks were distributed on 4 MCUs, where each MCU represents a single node on the network as shown in Figure 2. The periodic tasks are required to compute control law were implemented on controller MCU, therefore, this node will not deal with the outputs or inputs to the actuators.

At each program cycle, it will create a CAN message frame at the end of this cycle, this frame contains 6 bytes of data, each element of torque vector is presented using 2 Bytes, the 1st Byte represents the direction of the torque, the other one represents the absolute value of the torque element as shown in Figure.3. For the other nodes, each one is connected to a single actuator and its digital sensor. It is well known that the digital encoder consists of 2 digital channels, these channels will continuously change their values as the actuator is running, therefore, to handle these changes, each MCU should enable the external interrupt. As the flag of this interrupt is enabled, the MCU should compute the equivalent actuator's joint position and velocity. Whenever one of the states is changed, the MCU will directly create a CAN message with data length of 4 Bytes as shown in Figure.4.

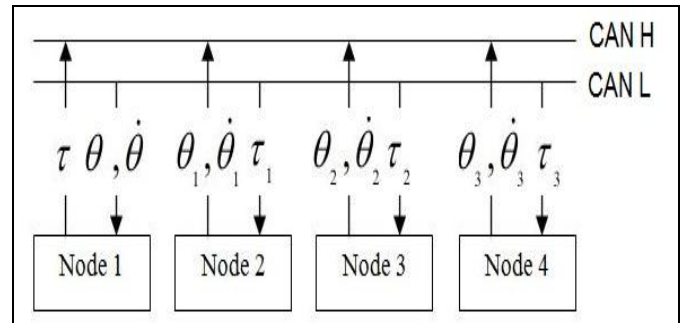


Figure 2. System network.

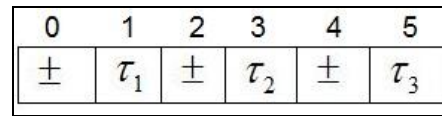


Figure 3. Control law data frame.

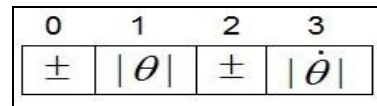


Figure 4. States data frame

3. Results

The system was implemented and simulated using MATLAB, the design of the controller MCU and ISAS 1 are shown in Figures 5, and 6 respectively.

The controller MCU receives angles frame from other MCU's, each frame contains data that represent the amplitude and the direction of the angular position along with the ID of the transmitter MCU, also it receives the desired position as a vector of the desired coordinates [X ; Y ; Z], this signal is called Reference signal. After that, it computes the required actuator's torque and creates a frame of torque data then it broadcasts this frame through CAN BUS.

The simulation is carried on using the Truetime toolbox. All kinematics and dynamics parameters are represented in Figure.7 and described in Table 2.

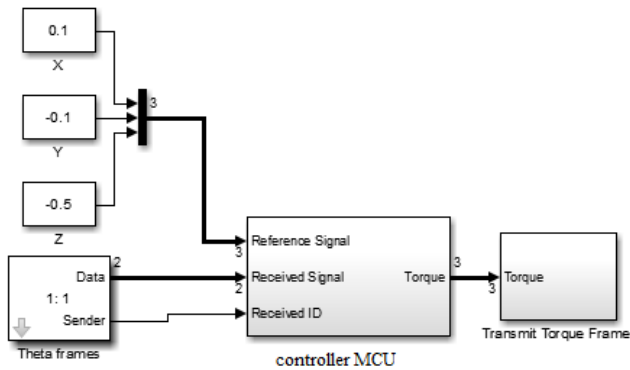


Figure 5. controller MCU design.

The other 3 MCU's are symmetrical, each one is physically connected to an actuator and a position incremental encoder. The encoder is used to measure the angular position (q) and thus to measure angular velocity(w) of the actuated joint. Both q and w which will be transmitted to the controller MCU by a frame that contains joint's information and the ID's of the receiver and transmitter (1:2 means that the MCU with ID 2 will transmit frame to MCU with ID 1).

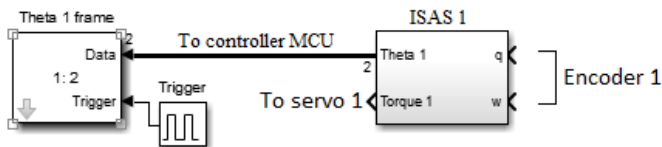


Figure 6. ISAS 1 design.

The obtained results prove the efficiency of the distributed control, Figure.8 shows the response of the robot when applying a reference signal [0.1; -0.1; 0.5] meter, the trajectory was designed such that the robot can finish its task in exactly 0.2 second. The results show that the response of the robot is stable, robust and can achieve its tasks in the desired manner using the specified trajectory time. Figure.9 shows the required torque to move the end effector from its initial to final position [0.1; -0.1; 0.5] as was planned.

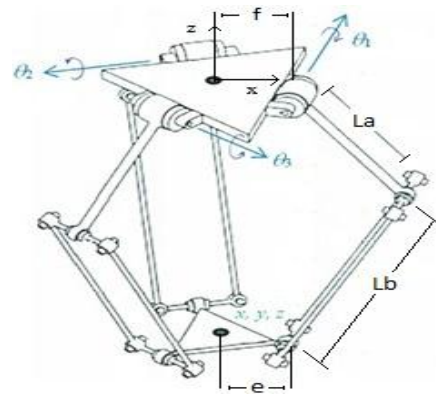


Figure 7. Delta robot with its dynamic parameters.

TABLE 2: Kinematics and Dynamics parameters

#	Parameters		
	description	symbol	value
1	Length of the upper arm	L_a	0.18 m
2	Length of the lower arm	L_b	0.435 m
3	Radius of the base platform	f	0.1 m
4	Radius of the moving platform	e	0.055 m
5	Mass of the upper arm	m_a	0.19 kg
6	Mass of the lower arm	m_b	0.055 kg
7	Mass of the moving platform	m_e	0.196 kg

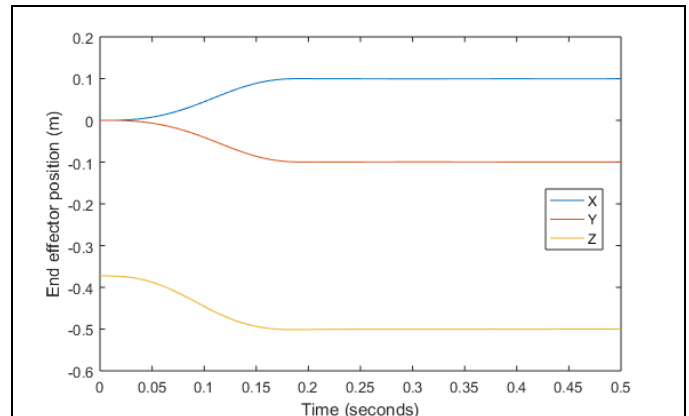


Figure 8. Robot response for (0.1, -0.1, 0.5) reference signal.

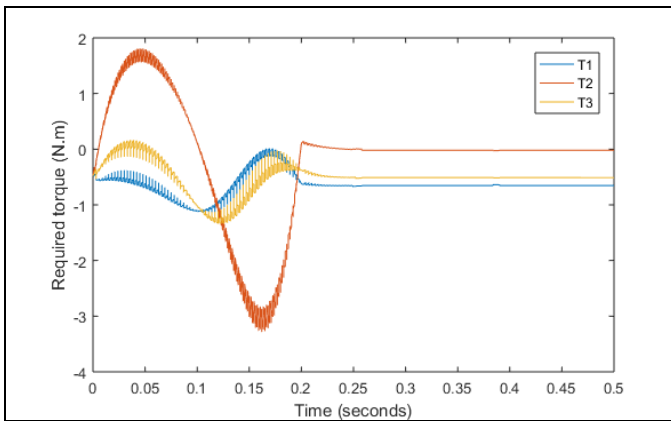


Figure 9. Required torque in joint space.

4. Conclusions

In this project, a real time distributed controller was implemented for delta robots, this design can be also applied or any type of manipulators. This design was able to handle and execute all periodic, aperiodic and sporadic tasks efficiently without any missing in their deadlines. The response of the robot was stable and rapid. Using this approach, the computation time was minimized as it was distributed on 4 MCU's. Furthermore, this approach allows sampling time to be reduced to 6.5 ms, however sampling time was adjusted to 8.5 ms to avoid processor overloading.

5. Acknowledgment

This work is a graduate course project in real time systems course at Palestine Polytechnic University.

8. References

- [1] T. Su, H. Zhang, Y. Wang, S. Wu, J. Zheng, J. Chang, et al., "Sorting system algorithms based on dynamic picking for delta robot," in 2017 Seventh International Conference on Information Science and Technology (ICIST), 2017, pp. 450-457.
- [2] S. Wang, H. Lin, R. Gai, and Y. Sun, "An application of vision technology on intelligent sorting system by delta robot," in *IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 2017 2017, pp. 1-6.
- [3] L. Yan, D. Liu, and Z. Jiao, "Novel design and kinematics modeling for delta robot with improved end effector," in *Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE*, 2016, pp. 741-746.
- [4] P. Liu and Q.-M. Li, "Modeling inverse solution of mechanism configuration for delta industrial robot," in *Design, Manufacturing and Mechatronics: Proceedings of the 2015 International Conference on Design, Manufacturing and Mechatronics (ICDMM2015)*, 2016, pp. 385-391.
- [5] Y.-L. Kuo and P.-Y. Huang, "Experimental and simulation studies of motion control of a Delta robot using a model-based approach," *International Journal of Advanced Robotic Systems*, vol. 14, p. 1729881417738738, 2017.
- [6] J. Fabian, C. Monterrey, and R. Canahuire, "Trajectory tracking control of a 3 DOF delta robot: a PD and LQR comparison," in *Electronics, Electrical Engineering and Computing (INTERCON)*, 2016 IEEE XXIII International Congress on, 2016, pp. 1-5.
- [7] J. Brinker, B. Corves, and M. Wahle, "A comparative study of inverse dynamics based on clavel's delta robot," in *Proceedings of the 14th World Congress in Mechanism and Machine Science*. Taipei, Taiwan, 2015, pp. 25-30.
- [8] Sastry, J. K. R., J. Viswanadh Ganesh, and J. Sasi Bhanu. "I 2 C based Networking for Implementing Heterogeneous Microcontroller based Distributed Embedded Systems." *Indian Journal of Science and Technology* 8.15 (2015).
- [9] Flores-Abad, Angel, and Adrian Arpidez. "Embedded control system for a 5-DOF manipulator by means of SPI bus." *Electronics, Robotics and Automotive Mechanics Conference*, 2009. CERMA'09.. IEEE, 2009.
- [10] Ran, Li, et al. "Design method of CAN BUS network communication structure for electric vehicle." *Strategic Technology (IFOST)*, 2010 International Forum on. IEEE, 2010.
- [11] Van Osch, Michiel, and Scott A. Smolka. "Finite-state analysis of the CAN bus protocol." *High Assurance Systems Engineering*, 2001. Sixth IEEE International Symposium on. IEEE, 2001.Ttt
- [12] Li, Yangmin, and Qingsong Xu. "Dynamic analysis of a modified DELTA parallel robot for cardiopulmonary resuscitation." *Intelligent Robots and Systems*, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on. IEEE, 2005.555
- [13] Guglielmetti, Philippe, and Roland Longchamp. "A closed form inverse dynamics model of the delta parallel robot." *IFAC Proceedings Volumes* 27.14 (1994): 51-56.
- [14] Olsson, André. "Modeling and control of a Delta-3 robot." *MSc Theses* (2009).
- [15] D. H. Anton Cervin, Bo Lincoln, Johan Eker, Karl-Erik Årzen, "How Does Control Timing Affect Performance? Analysis and Simulation of Timing Using Jitterbug and TrueTime," *IEEE Control Systems Magazine*, vol. 33, pp. 16-30, June 2003
- [16] Liu, Fan, Ajit Narayanan, and Quan Bai. "Real-time systems." (2000).

An Overview of Captcha Systems and Bypassing Math Captcha

Y.SANTUR

Firat University, Elazig/Turkey, ysantur@firat.edu.tr

Abstract - In web applications, communication between the user and the server is realized through web forms. These forms can be filled by real people or software called bot. A typical bot thousand and many more garbage records can be created in a second. This situation can cause performance loss and security bottlenecks in web application.

One of the methods that can be used to increase the security of web forms is the human validation application Captcha. With Captcha, the form field is placed in a question text and input that cannot be solved by computers but people can easily solve it. Form submission is not approved unless the entry is correctly filled. In this study, the current Captcha methods were compared with the advantages and disadvantages of each other and an application was applied to bypass the Math Captcha technique used as a WordPress plugin.

Keywords – Math Captcha, Regex, Web Application Security, WordPress

I. INTRODUCTION

In today's world, many web applications such as social networks, news, trips, blogs, are widely used in almost every field. Thanks to these applications, people often communicate with each other or owner the web application [1].

In web applications, communication between client and server is done via web forms including radio button, input, select box, check box, text area. However, these forms can be filled not only by humans, but also automated programs called bot. Moreover, these bots can add hundreds or thousands of records with meaningless information to the server database in seconds. In this case, it causes both unwanted recordings in the server database and performance bottlenecks. One of the methods used to secure web forms is Captcha systems.

Captcha (Completely Automated Public Turing test to tell Computers and Humans Apart) was coined in 2000 by Luis von Ahn, Manuel Blum, Nicholas Hopper and John Langford of Carnegie Mellon University [2]. A typically Captcha is a program that protects websites against bots by generating and grading tests that humans can pass but current computer programs cannot. For example, humans can read distorted text in figure, but current computer programs can't.

Captcha systems can be classified in terms of several different criteria [3]. One of them is the difficulty of the algorithm used. For example, very easy selection of the

algorithm causes bots to easily bypass the captcha system, while difficult algorithms cause people to experience bad experiences on the web page. People may be spend unnecessary time on webpage. Moreover, advertisers may be suffer financial loss.

Nowadays there are widely used captcha system which will be examined in detail in the following sections such as text, audio, picture, 3D picture, video, question-answer, natural language processing, artificial intelligence, game, math and puzzle based [4, 5]. Captcha systems are not the only used for preventing the use of bots and spam. They are also used for the development of Artificial Intelligence algorithms (AI). An example of text based captcha system is shown in Figure 1. It is sometimes used as a dataset for text processing because many people read handwriting text in web forms [6]. Then, these data will be save for AI. Similarly, the visual-based captcha system prompts describe people to live or objects in the picture. Then these data sets use for automated AI based image captioning [7].



Figure 1: An example of Captcha [2]

WordPress is one of the most used online services to create web pages [8]. WordPress is an online system used by many users across the globe as a hosting, domain name and template provider. In this system, users can quickly create web pages without needing technical information.

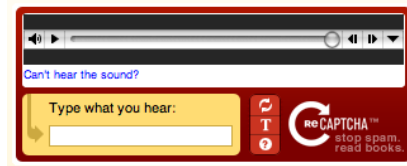
Another feature of the WordPress system is that users can easily add compatible plugins to their pages. One of these plugins is the den Math Captcha plugin which is used on more than 30,000 WordPress pages [8]. With this plugin webpage form, users can perform human validation by asking simple mathematical operations. This application is a rule based lightweight bot application which uses the regex library which bypasses the WordPress Math Captcha plugin and has been successfully tested.

II. CAPTCHA OVERVIEW

Nowadays there are text, audio, picture, 3D picture, video, math, natural language processing (NLP), AI, game and puzzle based captcha systems. These are detailed in this section.

- **Text Based:** It is very effective and simple. In this method, mostly one or two words are composed of handwritten texts. The biggest disadvantage is that the very difficult handwritten texts are time consuming [2, 9].
- **Audio Based:** This is only option available to visually impaired users. Most text-based systems also have audio support [10].
- **Picture Based:** These systems are tests in which the users have to guess those images that have some similarity. In this way, mostly users can select the pictures they asked by clicking with the mouse, so it takes a shorter time than text-based [11].
- **3D Picture Based:** Moving and/or 3D synthetic text can be solved more difficult by the computer. Therefore, this method is more advantageous than picture based methods [5].
- **Video Based:** It is a less commonly seen Captcha system. In this way, two or three words tags are provided to the user which describes a video. The biggest disadvantage of the method is the exploitation of bandwidth due to the size of video.
- **Math Based:** In this method, users are asked for various mathematical operations, from basic four process questions to complex derivative integral questions. Complex mathematical operations are not widely used because they cannot be solved by non-professionals [12].
- **NLP Based:** They are similar to text-based methods, but users are not asked read for handwriting text. Instead, a sentence is given to users and a question is asked about this sentence. This method is one of the most difficult methods that can be bypassed by bot programs, but it can also confuse people [13].
- **AI Based:** This is the methods based on NLP or image recognition methods in the previous section. This method is one of the most difficult methods that can be bypassed by bot programs [13].
- **Game and Puzzle Based:** In this method, a game scenario is presented to the users, or they are asked to combine a puzzle piece. It can take more time than other methods, or some people may find it fun [14].

Figure 2 gives an example of each of these methods. A text-based captcha system is shown in Figure 1. The Audio, Picture, 3D motion picture, Math, NLP and Game based Captcha systems with the sequence are shown in Figure 2. Picture based system to select wine pictures, NLP for entering red colored text, a game based system was used to select a basketball player.



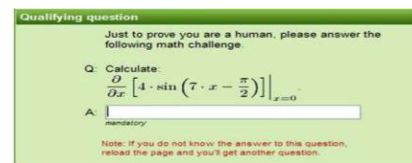
(a) Audio based Captcha [10]



(b) Picture based Captcha [11]



(c) 3D Captcha [5]



(d) Math Captcha [12]



(e) NLP based Captcha [13]



(f) Game based Captcha [14]

Figure 2: Different Captcha systems

In Captcha systems, it is the security of web by providing tests that target people can pass through but computers cannot pass. But nearly all the problems that can be solved by people today can be solved by computers through AI. For example, handwriting can be read by people or can be estimated by machine learning by computer. Even in some cases texts that a person cannot read, but can be read by the computer. The captcha systems examined in Table 1 are compared with advantages and disadvantages.

Table 1: Comparing Captcha Systems

Captcha	Major Advantages	Major Disadvantages
Text	Simple, easy, faster	<ul style="list-style-type: none"> • can be bypass with machine learning • some complex handwriting may not be readable
Audio	Simple, easy	<ul style="list-style-type: none"> • Vo-IP based systems can bring additional bandwidth load • Can be bypass with audio text conversion
Picture	Faster	<ul style="list-style-type: none"> • Can be bypass with deep learning
3D	Moving and 3d text can be more complex for computer	Can bring additional bandwidth load because of size
Math	Easy to apply	Difficult for people but easier for computers
NLP	<ul style="list-style-type: none"> • Safer for critical applications • Faster 	NLP and AI algorithms need to be developed.
AI		
Game	It may sound fun	Can bring additional bandwidth load because of size
Puzzle		

III. WORDPRESS MATH CAPTCHA PLUGIN

This plugin described as 100% effective and easy to use by owner [8]. It is hiding Captcha for logged in users and select some mathematical operation to use for others, can displaying Captcha as numbers and/or words. Figure 3 is an example of the WordPress Math captcha system. It is a simple three-variable mathematical process which is given “1” that involves simple operations such as addition, subtraction, multiplication, division. Any two of the three variables are given in the equation and the one of these is desired. Similar methods are used not only in the WordPress system but in many online services.

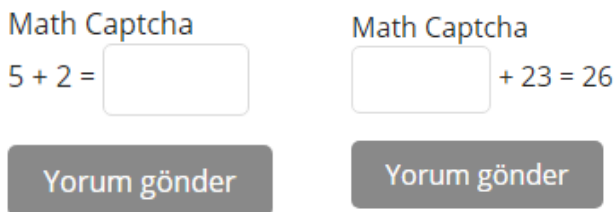


Figure 3: WordPress Math Captcha plugin

$$x \theta y = z \quad (1)$$

IV. PROPOSED METHOD

In this study, a simple jump technique that can be used for WordPress Math Captcha and similar add-ons is presented. The algorithm has 3 steps as given in Figure 4.

- 1) Bot program downloaded the web page content, Math Captcha code in the source code is obtained.
- 2) In this source code consisting of a few lines, two variables used in the simple equation, mathematical operation and the input element to enter the result of the operation are determined using Regular Expression (Regex) [17].
- 3) By performing the mathematical operation in the last step, Captcha is populated and sent the form.

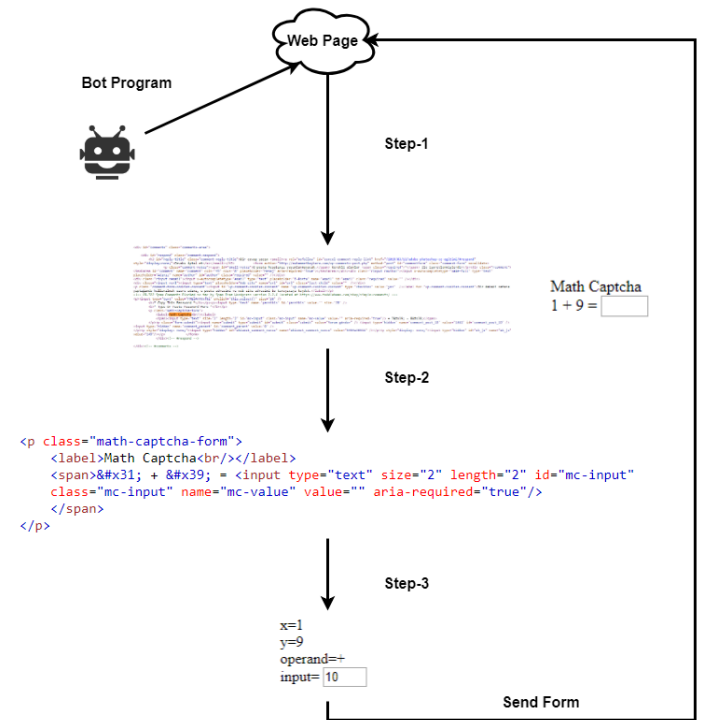


Figure 4: Proposed method for bypassing Math Captcha

The sample given above has been successfully tested in the localhost environment. The sample screenshot is shown in Figure 5. The study was performed using python programming language and Regex, urllib and request libraries [16]. All operations on the i5 processor computer lasts an under of .5 seconds.

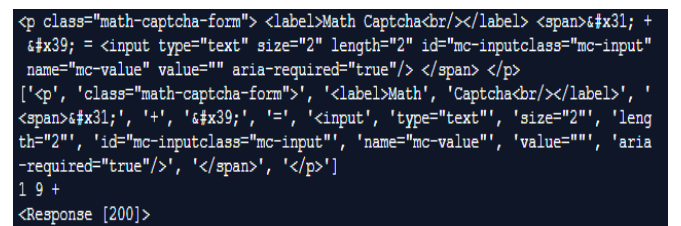


Figure 5: Bypassing Math Captcha in console environment.

V. CONCLUSION

In today's world, many web applications such as social networks, news, trips, blogs, are widely used in almost every field. Thanks to these applications, people often communicate with each other or owner the web application through web forms. But these forms not only can be filled by real people but also software called bot. This leads to numerous spam and loss of performance in web applications. One of the method can be used to secure web forms is using Captcha system. Captcha is a software that protects websites against bots by generating and grading tests that humans can pass but current computer programs cannot. But today's, thanks to AI, computers easily can bypassing many Captcha systems. In this study, the current major nine Captcha methods were reviewed and compared with the advantages and disadvantages of each other. Besides, a console application was developed and successfully tested to bypass the WordPress Math Captcha plugin.

REFERENCES

- [1] D. Stroud, "Social networking: An age-neutral commodity—Social networking becomes a mature web application", *Journal of Direct, Data and Digital Marketing Practice*, 9(3), 278-292, 2008.
- [2] L. Von Ahn, M. Blum, N. Hopper and A. Langford, "Captcha", [Online], Available: <http://www.captcha.net>
- [3] C. Demirel, D. Kılınc, "Captcha ile Güvenliğe Genel Bakış", *Türkiyede İnternet, inet-tr*, pp.1-4, 2016.
- [4] A.S El Ahmad, J. Yan, M. Tayara, "The robustness of Google CAPTCHA's", *Computing Science*, Newcastle University, 2011.
- [5] M. Imsamai, S. Phimoltares, "3D CAPTCHA: A next generation of the CAPTCHA", In 2010 International Conference on Information Science and Applications, pp.1-8, 2010.
- [6] M. Belk, C. Fidas, P. Germanakos, G. Samaras, "Do human cognitive differences in information processing affect preference and performance of CAPTCHA?", *International Journal of Human-Computer Studies*, 84, pp.1-18, 2010.
- [7] G. Garg, C. Pollett, "Neural network captcha crackers", In 2016 Future Technologies Conference (FTC), pp.853-861, 2016.
- [8] Wordpress, [Online], Available: <http://www.wordpress.org>
- [9] T. Tamang, P. Bhattachakosol, "Uncover impact factors of text-based CAPTCHA identification", In 2012 7th International Conference on Computing and Convergence Technology (ICCT), pp.556-560, 2012.
- [10] T. Ahmed, K.A. Tushar, S. I. Nova, M. M. Rahman, "Simple, Robust & User Friendly CAPTCHA 'InstaCap' for Web Security", *International Journal of Hybrid Information Technology*, 9(1), pp.163-182, 2016.
- [11] S. Sivakorn, J. Polakis, A.D. Keromytis, "I'm not a human: Breaking the Google reCAPTCHA", 2016.
- [12] C.J. Hernandez-Castro, A. Ribagorda, "Pitfalls in CAPTCHA design and implementation: The Math CAPTCHA, a case study", *computers & security*, 29(1), pp.141-157, 2010.
- [13] NLP Captcha, [Online], Available: <http://nlpcaptcha.in/>
- [14] S. Kulkarni, D.H. Fadewar, "CAPTCHA based web security: an overview", *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(11), pp.154-158, 2013.
- [15] V.P. Singh, P. Pal, "Survey of different types of CAPTCHA", *International Journal of Computer Science and Information Technologies*, 5(2), pp.2242-2245, 2014.
- [16] R. Lawson, "Web scraping with Python", Packt Publishing Ltd, 2015.

A Review on Nvidia GauGan

Y.SANTUR

Firat University, Elazig/Turkey, ysantur@firat.edu.tr

Abstract - In this study, the world's largest graphics card manufacturer Nvidia's deep learning assisted drawing software tool "GauGan" has been investigated. GauGan which is firstly announced Nvidia GPU Technology Conference (GTC 2019) is basically a drawing tool. The difference between the other drawing tools is that it can transform simple drawing lines into realistic nature pictures by using deep learning. It is improved on previous drawing software tool which is called "pix2pixHD". Nvidia has announced that it will use 1 million images on the Flickr platform as a set of data for the development of Deep Learning.

Keywords – Cuda, Deep Learning, GAN, GauGan, Nvidia

I. INTRODUCTION

There are many technological developments thanks to Deep Learning (DL) which is one of the sub-branches of Artificial Intelligence (AI). Today, many applications such as machine translation, object recognition, driver fatigue detection, autonomous robots, semantic text analysis, face recognition, picture subtitles can be easily performed by computers [1].

We basically owe these developments to two things. The first one is improved AI algorithms. But advanced AI algorithms needed large data sets and advanced hardware. Modern computers were not enough to be used in the training of these algorithms because of its complexity and size. The use of FPGA and / or Parallel programming for this purpose required additional workforce. Nowadays, the graphics processor units (GPU) have hundreds/thousands more cores compared to the central processor units (CPU) [2].

With the developing technology, the GPUs produced today are both very advanced and have reached a reasonable range in price. Nvidia GTX 20 series GPUs, the World's largest GPU manufacturer, have approximately 3000 cores (Intel I9 CPU has 9 cores) [3]. Nvidia has also developed Compute Unified Device Architecture (CUDA) architecture that will work on these GPUs [4]. In this way, parallel programming on GPU is easier in today's. The open source Tensorflow framework developed by Google is assisted for CUDA supported GPU [5]. Many machine learning algorithms, such as Convolutional Neural network (CNN) [6], in the Keras framework, which is using the Tensorflow as a backend, can be implemented in easily with much fewer lines program code [7]. Therefore, CUDA-supported Nvidia GPUs are very suitable for today's DL algorithms.

CPU and GPU difference is shown in Figure 1. According to the Nvidia, CPU is suitable for tasks to be processed sequentially, GPU is suitable for parallel processing tasks.

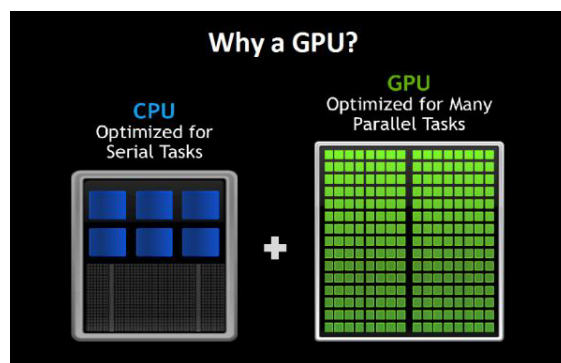


Figure 1: CPU and GPU difference [8]

One of the developments about deep learning is the drawing program "GauGan" which is developed by Nvidia takes its name from French painter Gauign and Generative Adversarial Networks (GAN) [9].

GANs are deep neural net architectures comprised of two nets, which is shown in Figure 2, pitting one against the other.

Generative: It is a neural network that produces synthetic/fake inputs similar to the inputs used in the machine learning model.

Adversarial: Outputs produced by Generative are compared with real inputs; if the probabilistic value obtained is below the threshold value then destroyed.

GANs were introduced in a paper by Ian Goodfellow and other researchers at the University of Montreal, including Yoshua Bengio, in 2014. Referring to GANs, Facebook's AI research director Yann LeCun called adversarial training "the most interesting idea in the last 10 years in ML" [9-13].

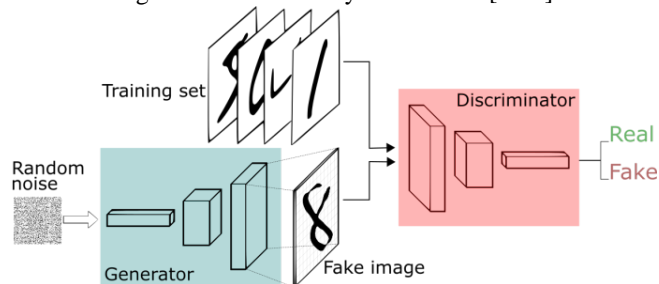


Figure 2: GAN generator Discriminator [11]

II. GAUGAN AND RELATED WORKS

GauGAN which is developed by Nvidia, is a DL and GAN based painting software that converts a few brush strokes into realistic nature landscapes such as selected mountains, rivers and seas.

A. Deep Learning

DL is part of a broader family of AI machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, reinforcement or unsupervised. Deep learning architectures such as deep neural networks, deep belief networks, Convolutional Neural network (CNN) and recurrent neural networks (RNN) was applied to many fields including computer vision, speech recognition, natural language processing, audio recognition, autonomous or humanoid robots, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, industrial diagnostics. Figure 1 shows a DL study performed by LeCun and others for image captioning [10]. In that study, CNN and RNN was used as a DL model. A CNN consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of convolutional layers, pooling layers, fully connected layers, normalization

layers and RELU layer for activation function. Figure 4 shows a CNN application [14, 15].

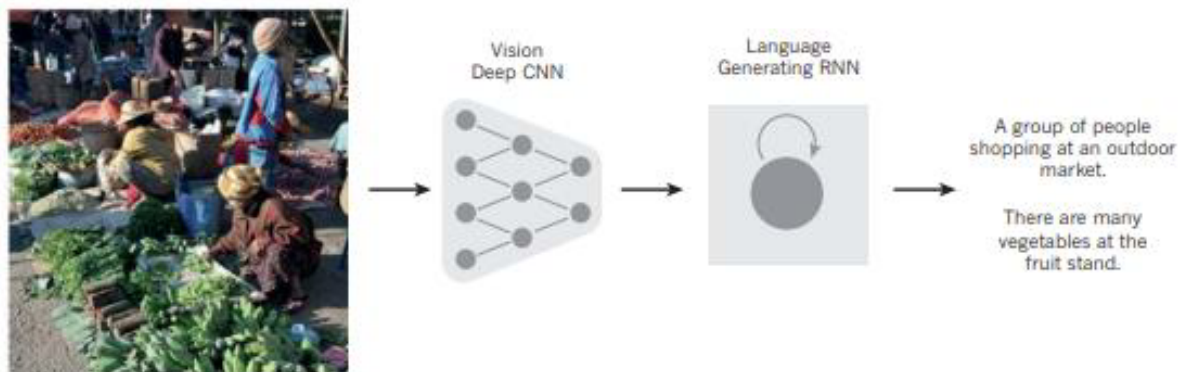
Convolutional: is the core building block of a CNN. It's parameters consist of a set of kernels that have a small receptive field but extend through the full depth of the inputs. During the forward pass, each filter is convolved across the width and height of the input size, computing between the entries of the filter and the input and producing a two-dimensional activation map of that filter.

Pooling: in this layer performs down sampling and feature selection. There are several methods for implementing pooling, among which max pooling is the most common.

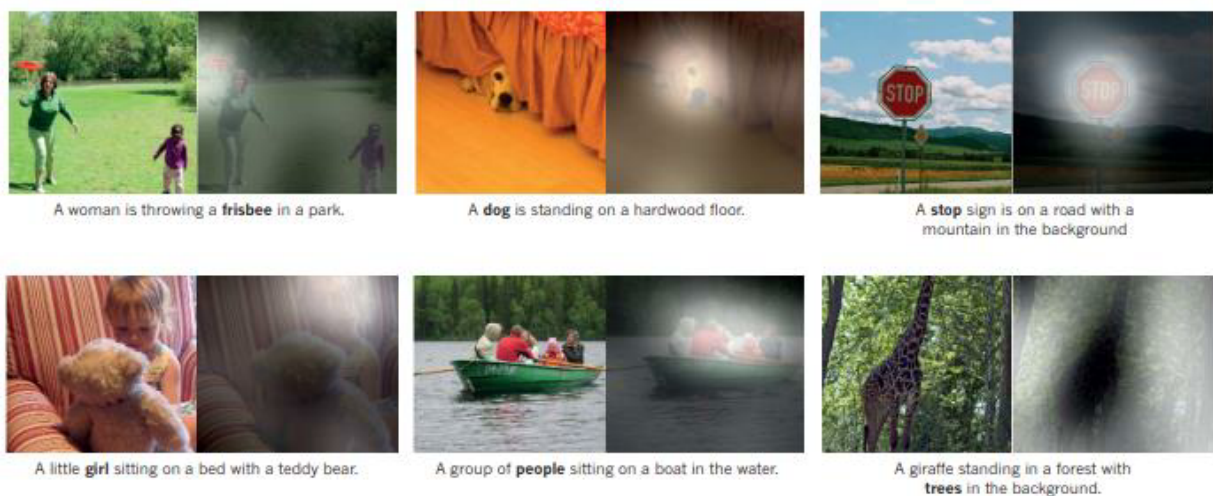
Fully Connected: with many convolutional and max pooling layers, the high-level reasoning in the neural network is performed via fully connected layers. Neurons in a fully connected layer are fully connected to all activations in the previous layer, as seen in neural networks.

Relu: is an activation functions that doesn't face gradient vanishing problem as with sigmoid and tanh function.

Normalization: to speed up training of CNN and reduce the sensitivity to network initialization, is use normalization layers between convolutional layers and output layers, such as ReLU layers.



(a)



(b)

Figure 3: DL for image recognition [10]

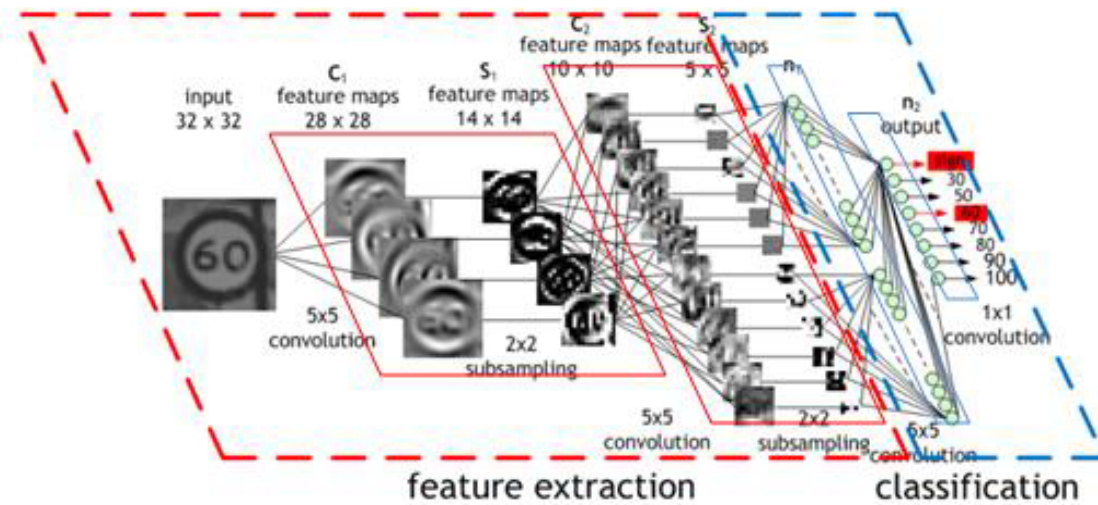


Figure 4: A CNN layers [14]

B. Generative Adversarial Networks (GANs)

This neural network type, which was invented by Google researcher Ian Goodfellow in 2014 [12], consists of 2 neural networks working opposite to each other. It was invented based on Nash equality in the game theory [13]. GANs typically based on unsupervised learning and learn itself how to generate similar given distribution of data. GANs have two neural networks as the generator and the discriminator. The generator is a type of CNN that will create new instances of an object. The discriminator is a type of deconvolutional neural network that will determine its authenticity whether or not it belongs in a dataset. GANs simply work in three steps:

- 1) In this step GAN is establishes to identify the desired end output and gather an initial training dataset based on those parameters. These data sets are then randomized and input into the generator until it acquires basic accuracy in producing outputs.
- 2) The generated images are fed into the discriminator along with actual data points from the original concept. The discriminator filters through the information and returns a probability between 0 (correlates as fake) and 1 (correlates as real) to represent each image's authenticity.
- 3) These values are then checked for success and repeated until the desired outcome is reached

The results obtained by GAN algorithm developed by Ian Goodfellow are given in Figure 5. In the figure, rightmost column shows the nearest training example of the neighboring. In the figure (a) MNIST, (b) TFD, (c) and (d) CIFAR-10 data set were used. GANs have been a source of inspiration for many different works since 2012. For example, Kaplan (2017) used the GANs to produce synthetic retinal images for medical image analysis [15].

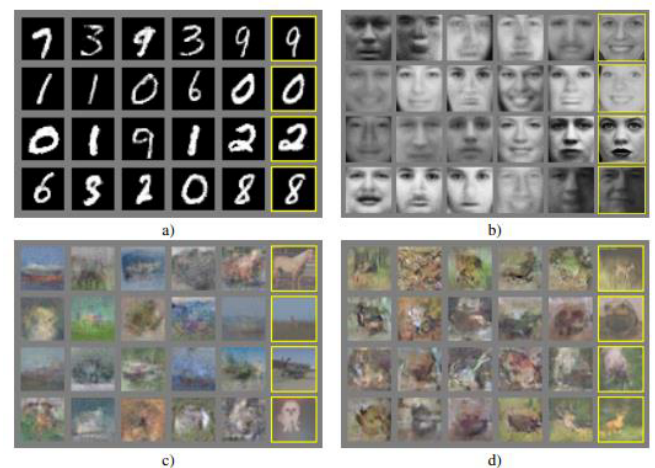


Figure 5: GANs [12]

C. GauGAN

GauGuin drawing software was also developed using GAN. GauGAN was trained with DL on a million images collected from Flickr [16]. According to Nvidia, "GauGAN allows users to draw their own segmentation maps and manipulate the scene, labeling each segment with labels like sand, sky, sea or snow (GauGAN totally have 180 models). The DL model then fills in the landscape with show-stopping results: Draw in a pond, and nearby elements like trees and rocks will appear as reflections in the water. Swap a segment label from "grass" to "snow" and the entire image changes to a winter scene, with a formerly leafy tree turning barren"[9]. GauGAN improves on pix2pixHD, which is developed in 2018. It is improved adversarial loss. The new model has fewer parameters and more generatable image options. Figure 6 is a screenshot of the GauGAN software.

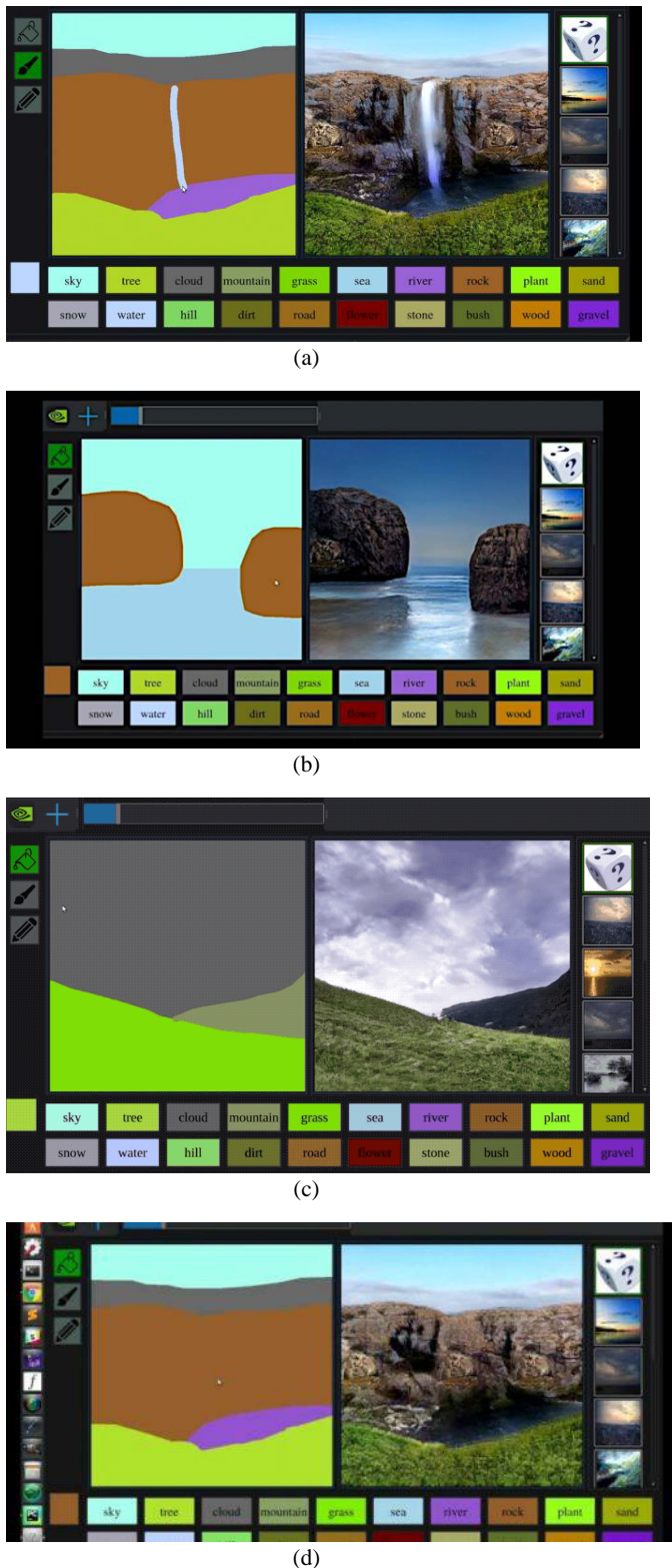


Figure 6: GauGAN [16]

The GauGAN software screen has a menu on the left, a canvas in the middle and models to be selected at the bottom. Menu includes paint bucket, a pencil and a brush. On the left side of the canvas, the user creates a synthetic

shape or line with any of these three tools, then then selects one of the models like rock, tree below. Finally, GauGAN creates realistic images by taking into account the user's synthetic input on the right side of the canvas.

III. CONCLUSION

In this study, AI and GAN based drawing software GauGAN developed by Nvidia were reviewed. It reveals the point where technology comes from. The GauGAN that was developed by Nvidia in 2019, research paper has been accepted by CVPR 2019 as an Oral Session. It is also improves on pix2pixHD, a NVIDIA research paper accepted to CVPR 2018. Although GauGAN is still a proof-of-concept and not commercial product, the method has potential applications in gaming, object detection, medical diagnostic, film-making, and many other visual presentation and editing areas. Moreover, DL and GAN, which is used GauGAN, can be used to produce new datasets in many areas such as medical diagnostics and industrial diagnostics. Because it is quite difficult to obtain real data sets in such areas.

REFERENCES

- [1] I. Goodfellow, Y. Bengio, A. Courville, "Deep learning", MIT press, 2016.
- [2] L. Yeager, J. Bernauer, A. Gray, M. Houston, "DIGITS: the deep learning GPU training system", In ICML 2015 AutoML Workshop, 2015.
- [3] Nvidia, [Online], Available: <http://www.nvidia.com>
- [4] D. Kirk, "NVIDIA CUDA software and GPU parallel computing architecture", In ISMM, (7), pp. 103-104, 2007.
- [5] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Kudlur, "Tensorflow: A system for large-scale machine learning", In 12th Symposium on Operating Systems Design and Implementation, (16), pp.265-283,2016.
- [6] A. Krizhevsky, I. Sutskever, G.E. Hinton, "Imagenet classification with deep convolutional neural networks", In Advances in neural information processing systems , pp.1097-1105, 2012.
- [7] F. Chollet, "Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek", MITP-Verlags GmbH & Co. KG, 2018.
- [8] "CPU vs GPS performance", [Online], Available: <https://blogs.nvidia.com/blog/2009/12/16/whats-the-difference-between-a-cpu-and-a-gpu/>
- [9] "Gaugan", [Online], Available: <https://blogs.nvidia.com/blog/2019/03/18/gaugan-photorealistic-landscapes-nvidia-research/>
- [10] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning", nature, 521(7553), 436, 2015.
- [11] "GAN", [Online], Available:<https://skymind.ai/wiki/generative-adversarial-network-gan>
- [12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, Y. Bengio, "Generative adversarial nets. In Advances in neural information processing systems", pp.2672-2680, 2014.
- [13] H.H. Chen, Y. Li, Y. Jiang, Y. Ma, B. Vucetic, "Distributed power splitting for SWIPT in relay interference channels using game theory", IEEE Transactions on Wireless Communications, 14(1), 410-420, 2015.
- [14] "Convolutional Neural Networks", [Online], Available: <https://developer.nvidia.com/discover/convolutional-neural-network>
- [15] S. Kaplan, "Deep generative models for synthetic retinal image generation", 2017.

A simulation based harmony search algorithm for part routing optimization problem

Ayşe Tuğba Dosdoğru¹, Aşlı Boru¹, and Mustafa Göçken¹

¹Adana Alparslan Türkeş Science and Technology University, Adana/Turkey, adosdogru@atu.edu.tr

¹Adana Alparslan Türkeş Science and Technology University, Adana/Turkey, aboru@atu.edu.tr

¹Adana Alparslan Türkeş Science and Technology University, Adana/Turkey, mgocken@atu.edu.tr

Abstract – In production systems, one of the main part routing problems is the determining the best process sequence and the best machine-operation pairing. In real world systems, flexibility is important to provide alternative part routings. The system is considered flexible when certain operation of a part can be processed on more than one machine. However, flexibility increases the complexity of system. This paper deals with assignment of operations to a machine and determining the best process plan for each part under flexible environment. We used a harmony search algorithm to cope with highly complex and nonlinear problem in which makespan is considered as the objective function and evaluated by using a simulation model.

Keywords - Harmony search, simulation, part routing, flexible job shop environment.

I. INTRODUCTION

THIS paper is concerned with two sub-problems: 1) assignment of operations to a machine, and 2) sequencing operations in the context of flexible job shop environment. Machine selection and operation sequence determination makes the problem much more complex. In this case, flexible job shop scheduling problem (FJSP) is considered as an NP-hard combinatorial optimization problem. Thus, using metaheuristics becomes inevitable to solve the FJSP.

There have been many studies which focus on solving FJSP by using metaheuristics. Rossi and Dini [1] presented an ant colony optimization based software system. Saidi-Mehrabadi and Fattahi [2] analyzed the FJSP using a tabu search algorithm. The results of the proposed method were compared with a classical branch and bound method. Liouane et al. [3] integrated the ant system optimization with local search methods. Pezzella et al. [4] employed a genetic algorithm for the FJSP with makespan criterion. A hybrid tabu search and particle swarm optimization was developed by Zhang et al. [5] to analyze the multi-objective FJSP. Wang et al. [6] developed an artificial bee colony algorithm for FJSP. Among these studies, a few study considered a harmony search (HS) algorithm for FJSPs. Yuan et al. [7] proposed the HS algorithm to investigate the FJSP. They combined HS algorithm and a local search procedure to enhance HS. An integrated search heuristic is proposed for large-scale FJSP by Yuan and Xu [8]. They integrated HS algorithm with a large

neighborhood search. Gao et al. [9] presented a discrete HS for FJSP.

From the literature, it is clearly seen that metaheuristic algorithms are successful at achieving efficient solutions. Besides, some studies used simulation as an evaluation function for the optimization problem as a set of values given to the decision variables [10]. Azadivar and Tompkins [11] integrated genetic algorithm and simulation to evaluate various configurations. Halim and Seck [12] proposed a simulation-based evolutionary optimization methodology for a multi-objective supply chain optimization problem. Kulkarni and Venkateswaran [13] integrated simulation with an optimization model to analyze the job shop scheduling problems. In this study, simulation ensures feasible solutions for each changing constraints in the optimization model.

This paper provides an approach to solve a part routing problem by minimizing makespan. Simulation based HS algorithm is used to solve part routing optimization. HS algorithm which is envisaged to be the optimization component of the methodology is integrated with the simulation model of the production system.

The problem description is given in section 2. In section 3, computational results are given. In this section, the algorithm is tested on several problem instances selected from literature. Finally, a conclusion is given in section 4.

II. PROBLEM DESCRIPTION

Both process plan and part routing optimization in FJSP are presented throughout this study to investigate the effectiveness of the methodology given here. Considered problem consists of two consecutive stages; (1) appropriate machine selection for each operation by using HS, and then (2) sequencing of each operation by using HS.

There are n parts $P = p_1, p_2, \dots, p_n$ indexed by i and each part has varying number of operations. There are m machines $M = \{M_1, M_2, \dots, M_m\}$ placed in the job shop environment. Each operation O_{ik} , i.e., the k th operation of part P_i , can be processed at any compatible machine among a subset M_{ik} . This phase is called as optimization phase. Second phase is the simulation phase where new sets of values for decision variables produced by the first phase (i.e., HS) are evaluated.

While the values of decision variables are generated through the first phase, the second phase evaluates these new values by using the simulation model of the system. In this respect, HS is developed to assign new values to selected decision variables (i.e., generating candidate solutions), and a simulation model is created to evaluate these new values of decision variables. This is an iterative process as given in Figure 1. The procedure continues until some termination criterion is met. The main goal is to determine the solution that optimizes the value of the objective function. In this paper, we determined the best operation sequence and machine-operation pairs while minimizing the total makespan.

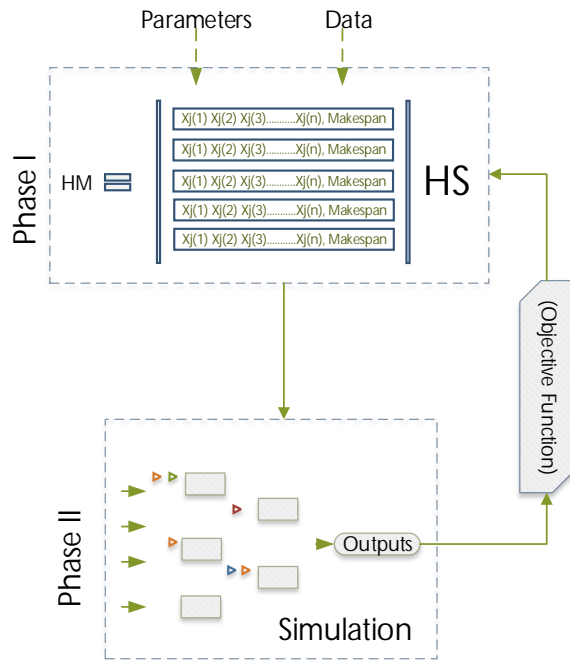


Figure 1: General framework of the proposed study.

A. Simulation Phase

In simulation, user can easily change the simulation model parameters. In addition, the system performance can be observed under different sets of parameters. At this point, it is natural to try to determine the parameter sets that optimize the performance of system. Therefore, we employed an Optimization via Simulation (OvS). It does not depend on strong problem structure and is tolerant of some sampling variability. Details of OvS can be found in [14].

In this paper, simulation model is employed to evaluate the performance of the part routings obtained by optimization module (HS). For this reason, a computer simulation model of flexible job shop environment is developed by using SIMIO 6.105.11267 Enterprise Edition.

Simulation model assumptions are determined as follows:

- All parts arrive to the shop at time 0.
- Initial sequence for each part's first operation is fed into simulation model.
- After completing their first operation, they move to the next machine queue. FIFO dispatching rule is used.
- At each arrival only one part arrives at the system (i.e., no batches considered).

- Preemption is not allowed.
- A machine cannot perform more than one operation at a time.
- Consecutive operations of parts can be processed on the same machine.
- The setup times and transportation times assumed to be zero and are excluded from the model.
- Machine breakdowns are not considered.
- Simulation ends with the completion of last operation, and end time corresponds to makespan.

B. Optimization Phase

The HS algorithm developed by Geem et al. [15] is one of the latest metaheuristic which is inspired by music improvisation process implemented by musicians.

The setting of parameters and the stopping criterion is considered as first step of the HS algorithm. Then, the Harmony Vectors (HM) are initialized. After, simulation is employed to evaluate the harmony vectors in the HM for the FJSP. Afterwards, the best and worst harmony vectors are determined. The next step is improvising a new harmony memory from the HM through a memory consideration, a pitch adjustment, and a random selection. The best harmony vector is compared with new improved harmony vector and the HM is updated. If new harmony vector is better than the best harmony vector, the new harmony vector is saved as the new best harmony vector. The process is repeated until the stopping criterion is met.

In this study, HS algorithm is designed to integrate with model and to change the values of the decision variables of the model directly (i.e., by this way HS generates candidate solutions). The results of the simulation run is returned back to the HS as the most recent fitness functions to be evaluated, and HS once more tries to determine better decision variables to improve the performance of model.

III. COMPUTATIONAL RESULTS

The HS algorithm is coded using Microsoft Visual C# 2013, and flexible job shop environment is created using SIMIO 6.105.11267 Enterprise Edition. To test the performance of the proposed simulation based HS, four problem instances are used known as Fdata from [16].

Table 1: Problem parameters for selected benchmark problems.

Problem	n	m	LB
SFJS01	2	2	66
SFJS02	2	2	107
SFJS10	4	5	516
MFSJ01	5	6	468

Three problems are selected from small size FJSPs (i.e., SFJS01, SFJS02 and SFJS10), also a problem is selected from medium and large size FJSPs (i.e., MFSJ01). Problem parameters are given in Table 1. Problems are run on an Intel Core i7 2.20 GHz CPU with 8 GB RAM computer.

In Table 1, second column (i.e., n) gives the number of parts, third column (i.e., m) gives the number of machines, and

the last column demonstrates the lower bound for each problem. HS algorithm parameters are given in Table 2. Integrated HS-Simulation is terminated when the algorithm finds lower bound of the problem or when it reaches the maximum number of iterations (number of improvisations).

Table 2: HS parameters used in the study.

Parameter Name	Value
Harmony Memory Size	20
Pitch adjusting rate	0.3
Harmony memory consideration rate	0.95
Bound factor	1
Number of improvisations	1000

Algorithm is run for five times for each problem independently. Best result (the minimum makespan obtained through five runs) is selected, and given as BC_{max} in Table 3.

Table 3: Computational results for each problem instance.

Problem	BC_{max}	Number of Improvisations
SFJS01	66	37
SFJS02	107	2
SFJS10	516	588
MFSJ01	498	1000

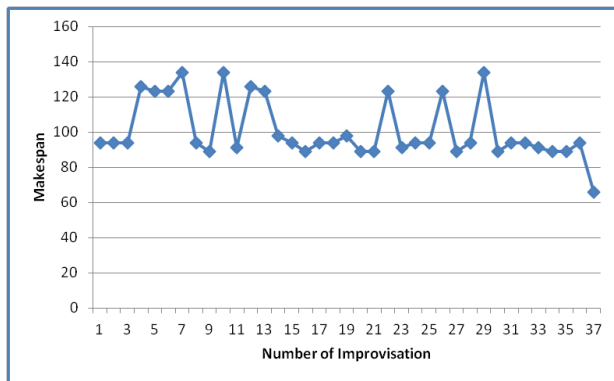


Figure 2: Makespan values obtained from simulation through iterations for problems SFJS01.

As can be seen from the Table 3, HS algorithm determines the lower bounds for the first two instances within less than 40 improvisations. Figure 2 shows the makespan values through iterations for problem SFJS01. For problem instance SFJS10, algorithm reaches the lower bound after 588 improvisations. HS algorithm finds the makespan 498 which one is quite close to the known best result (i.e., 468) for problem MFSJ01 after 1000 iterations. HS algorithm uses fewer parameters which makes it more convenient for practical use [7]. Nevertheless, when the problem size gets larger, the algorithm reaches better results with larger number of improvisations.

IV. CONCLUSION

This study proposed a simulation-based HS algorithm to solve part routing problems. Studies of Yuan, Xu, and Yang

[7], and Yuan and Xu [8] are the pioneering studies which solve flexible job shop scheduling problems by using HS. However, these studies are not considered the problem in the context of integration with simulation. Part routing problem consists of two sub-problems in our study. In this respect, two vector code is used in order to represent the problem. Objective function is determined to be makespan which is obtained from the simulation model. The HS and simulation model proceed successively. The proposed algorithm terminates when a pre-determined value (i.e., lower bound in this study) is obtained or a maximum number of improvisation is reached. Simulation is used to ensure feasible schedules for generated values by HS. Note that, the simulation model is built as a deterministic model which converts all solutions into a feasible schedule. The proposed algorithm is tested on selected problem instances from the literature. The computational results demonstrated that the algorithm is a promising candidate to solve production scheduling problems. In future studies, this proposed algorithm will be tested on problems with larger size, and detailed analysis will be performed.

To the best of our knowledge the solution to the considered test problems has not yet been considered via OvS approach. It should be noted that in spite of being solved via OvS approach all of the problem inputs considered to be deterministic. However, considered benchmark problems should be converted to its stochastic counterpart easily by defining some parameters, i.e., processing time, being stochastic. Such a conversion which enables one to analyze any type of system much more realistically is planned to be a future work.

REFERENCES

- [1] A. Rossi, and G. Dini, "Flexible job shop scheduling with routing flexibility and separable setup times using ant colony optimisation method," *Robotics and Computer Integrated Manufacturing*, vol. 23, pp. 503–516, 2007.
- [2] M. Saidi-Mehrabad, and P. Fattahi, "Flexible job shop scheduling with tabu search algorithms," *International Journal of Advance Manufacturing Technology*, vol. 32, pp. 563–570, 2007.
- [3] N. Liouane, I. Saad, S. Hammadi, and P. Borne, "Ant systems and local search optimization for flexible job shop scheduling production," *International Journal of Computers, Communications and Control*, vol. 2, pp.174–184, 2007.
- [4] F. Pezzella, G. Morganti, and G. Ciaschetti, "A genetic algorithm for the flexible job-shop scheduling problem," *Computers & Operations Research*, vol. 35, pp. 3202–3212, 2008.
- [5] G. Zhang, X. Shao, P. Li, and L. Gao, "An effective hybrid particle swarm optimization algorithm for multi-objective flexible job shop scheduling problem," *Computers & Industrial Engineering*, vol. 56, pp. 1309–1318, 2009.
- [6] L. Wang, G. Zhou, Y. Xu, S. Wang, and M. Liu, "An effective artificial bee colony algorithm for the flexible job shop scheduling problem," *The International Journal of Advanced Manufacturing Technology*, vol. 60, pp. 303–315, 2012.
- [7] Y. Yuan, H. Xu, and J. Yang, "A hybrid harmony search algorithm for the flexible job shop scheduling problem," *Applied Soft Computing*, vol. 13, pp. 3259–3272, 2013.
- [8] Y. Yuan, and H. Xu, "An integrated search heuristic for large-scale flexible job shop scheduling problems," *Computers & Operations Research*, vol. 40, pp. 2864–2877, 2013.
- [9] K. Z. Gao, P. N. Suganthan, Q. K., Pan, T. J. Chua, T. X. Cai, and C.S. Chong, "Discrete harmony search algorithm for flexible job shop scheduling problem with multiple objectives," *Journal of Intelligent Manufacturing*, pp.1–12, 2014.

- [10] J. April, F. Glover, J. P. Kelly, and M. Laguna, "Practical introduction to simulation optimization," in *Proceedings of the 2003 Winter Simulation Conference*, pp. 71–78, 2003.
- [11] F. Azadivar, and G. Tompkins, "Simulation optimization with qualitative variables and structural model changes: A genetic algorithm approach," *European Journal of Operational Research*, vol. 113, pp.169–182, 1999.
- [12] R. A. Halim, and M. D. Seck, "The simulation-based multi-objective evolutionary optimization (Simeon) framework," in *Proceedings of the 2011 Winter Simulation Conference*, pp. 2839–2851, 2011.
- [13] K. Kulkarni, and J. Venkateswaran, "Iterative simulation and optimization approach for job shop scheduling," in *Proceedings of the 2014 Winter Simulation Conference*, pp.1620–1631, 2014.
- [14] L. J. Hong, and B. L. Nelson, "A brief introduction to optimization via simulation," in *Proceedings of the 2009 Winter Simulation Conference*, pp. 75–85, 2009.
- [15] Z. Geem, J. H. Kim, and G.V. Loganathan, "A new heuristic optimization algorithm: harmony search," *Simulation*, vol. 76, pp. 60–68, 2001.
- [16] P. Fattahi, M. S. Mehrabad, and F. Jolai, "Mathematical modeling and heuristic approaches to flexible job shop scheduling problems," *Journal of Intelligent Manufacturing*, vol. 18, pp. 331–342, 2007.

Multi Class Tag Prediction on Stack Overflow Dataset

Y.SANTUR

Firat University, Elazig/Turkey, ysantur@firat.edu.tr

Abstract – In this study is focused on the approaches the multi class text document classification problem on Stack Overflow questions on published on the website Kaggle. The dataset contains a one hundred thousand questions and question's keywords on the Stack Overflow technology site. Main target of this study, using data from the Stack Overflow website, the problem is to predict the tags assigned to questions. For this purpose, a supervised learning has been performed by applying Natural Language Processing algorithms on the labeled dataset. In this proposed approach, 0.68 F1 score and 0.36 overall accuracy were obtained.

Keywords – Data Science, Document classification, Natural Language Processing, Text Processing

I. INTRODUCTION

NATURAL Language Processing (NLP) is a subfield of Artificial Intelligence (AI) that is focused on enabling computers to understand and process human languages, to get computers closer to a human-level understanding of language [1].

In NLP applications, the data source can be text and / or a conversation, but mostly it deals with documents that text based that called Corpus. NLP is considered a difficult problem in AI. It's the nature of the human language that makes NLP difficult. NLP can organize and structure knowledge to perform tasks such as summarization, author recognition, dialogue, information extraction, question-answering, semantic, part-of-speech tagging, document classification, machine translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, keyword/tag prediction and topic segmentation [2]. Therefore, NLP is one of the most important technologies of the information age in todays. Because there has been a striking growth in text data such as web pages, e-commerce, e-health, news, articles, messages, social media data, and scientific papers in the recent years [3]. NLP uses supervised, unsupervised, and reinforcement learning as a machine learning algorithm to overcome these problems.

That's why traditional machine learning algorithms could be insufficient. Therefore, to process, training and learn data on Corpus, which is consist of a series of words/sentences, there is often a need for DL algorithms, recurrent neural networks (RNNs) and convolutional neural networks (CNNs) [4, 5].

II. RELATED CONCEPTS AND TECHNOLOGIES

NLP, a sub-branch of artificial intelligence, contains many different fields of study. Document classification, machine translate, sentimental analysis, chatbots, question-answer systems are examples of this field [6].

Machine translation: is one of the most important application of NLP. Performing effective translation is very difficult task even for humans, requiring expert proficiency and prowess in areas such as morphology, syntax, and semantics [7]. Google translate, which can translate from many languages, is an example of this field. It can be used in very different area and web application using web service support [8].

Information extraction: is the process of using algorithms to extract patterns such as named entity recognition, event recognition and relationships extraction from text [9].

Summarization: is the task getting of the most important information from documents. In abstractive summarization is use DL approaches such as recurrent encoder decoder architectures [10]. There are two primary types of summarization techniques: extractive and abstractive.

Text classification: is assign predefined categories to free-text documents. In this field, n-gram, bag of words, word2vector algorithms are used [11].

Chatbots and question-answer systems: are a conversational agent that interacts with users in a certain domain or on topic, such as e-commerce, technology sites, as a e- assistant [12].

Some major areas of the above examined NLP have more subfields. One of them is the keyword prediction from the text. Automatic classification of these texts is an important application for tasks such as keyword extraction from text [13]. This problem is actually a supervised learning process that contains the keywords of a text and text. During the training, the input data of the model is text, and the outputs are keywords. In this study, question's text used as a dataset which is published on Stack Overflow [14]. The open source data set obtained from the Kaggle has 100000 questions and their keywords [15]. However, the proposed method is applicable to many areas of similar purpose. For the verification of the method, 70000 of the data set were reserved for training and 30000 were used for testing. TF-IDF, N-Gram, and Bag-of-Words algorithms were used for the vectorization of words in the training phase and .36 overall accuracy and .68 F1 value were obtained in the test phase [16].

III. PROPOSED METHOD

In this study, multi-class keyword estimation was performed on the stack overflow data set which is published on Kaggle. Dataset contains 100.000 question about computer science, such as databases, programming language, algorithms, bugs, errors and software development [15]. Each question includes the question text and the tags of the question. It is shown in Figure 1.

1	How to draw a stacked dotplot in R?	['r']
2	mysql select all records where a datetime field is less than a specified value	['php', 'mysql']
3	How to terminate windows phone 8.1 app	['c#']
4	get current time in a specific country via jquery	['javascript', 'jquery']

Figure 1: Stack Overflow dataset [15]

In this study 70 percent of this data set is reserved for training, 30 percent of it is reserved for test. The proposed method for this study is shown in Figure 2.

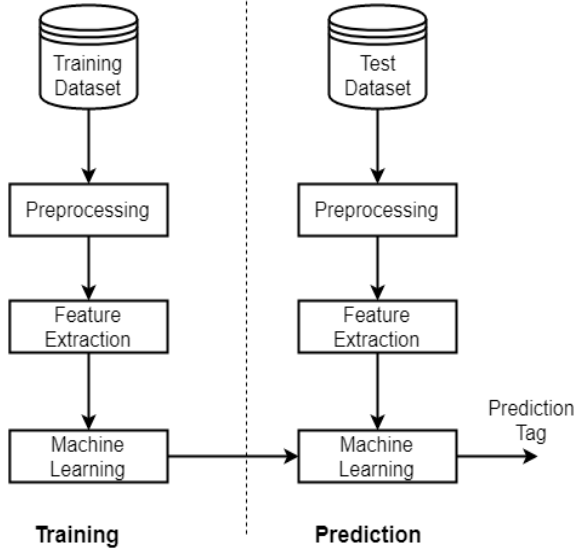


Figure 2: The proposed method.

The proposed method has a three main stages as a preprocessing, feature extraction and learning.

A. Preprocessing

Preprocessing has four processing in itself that is used in this study. The first step in the preprocessing is to break the question text and tag list apart into separate words. This process is also called tokenization. Tokenization is easy to do in English. It just split apart words if there's a space between them. The second is

called Lemmatization or Stemming. When working with text, it is helpful to know the base form of each word so that is known that both sentences are talking about the same concept. In NLP, it is called finding this process lemmatization or stemming figuring out the most basic form or lemma/stem of each word in the sentence. The third is stop words. Stop words are words which are filtered out before processing of natural language data. But there is no single universal list of stop words used by all NLP tools. After the first three stages, are deleted punctuation marks and stop words in the word lists, are all converted into capital letters.

B. Feature Extraction (Vectorization)

Machine learning cannot directly process text data such as words, sentences. In NLP applications, it is have to map those words to vectors that consist numbers. That is also called vectorization. In this study, more than one algorithms for the vectorization process were investigated.

TF-IDF: is used in a number of NLP techniques such as text mining, search queries and summarization. It uses all the tokens in the Corpus in calculation that shown in (1). Frequency of occurrence of a token from vocabulary in each document consists of the term frequency and number of documents in which token occurs determines the Inverse document frequency. What this ensures is that if a token occurs frequently in a document that token will have high TF but if that token occurs frequently in majority of documents then it reduces the IDF ,so stop words like and the in which occur frequently are penalized and important words which contain the essence of document get a boost. Both these TF and IDF matrices for a particular document are multiplied and normalized to form TF-IDF of a document.

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \quad (1)$$

Bag-of-Words: is a simplifying representation used in NLP. In this model, a text is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity. According to (2), when vectoring is done, if the word is existing in the corpus, it is assigned as a "1" and if it is not, it is assigned as a "0".

$$bow = \begin{pmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,n} \end{pmatrix} \quad (2)$$

$$x_{i,j} = \begin{cases} 1, & \text{if } x_{i,j} \text{ in } C \\ 0, & \text{Otherwise} \end{cases}$$

N-Gram: is a contiguous sequence of n items from a given sample of text. The n-grams typically are collected from corpus. The basis of the N-gram model is the chain rule of probability. In the study, n grams that given in (3) that is used find relationships between two questions or tags.

$$P(aUb) = P(a|b)P(b) \quad (3)$$

C. Machine Learning

In this study, a logistic regression is used which uses multi-class classifier as machine learning method. It consists in fitting one classifier per class that given in (4).

$$f(x) = \arg \max f_i(x) \quad (4)$$

D. Evaluation

Accuracy and F1 values were measured by creating a confusion matrix on the test data for evaluation that shown in (5). N is the total number of tests, TP is predicted True and the actual output was also True, TN is predicted False and the actual output was also False, FP is predicted True but the actual output was False, FN is predicted False and the actual output was also False. Acc is the ratio of number of correct predictions to the total number of test samples. $Precision$ is the ratio of correctly TP observations to the total predicted positive observations. $Recall$ is the ratio of correctly predicted positive observations to the all observations in actual class. $F1$ Score is the weighted average of $Precision$ and $Recall$.

$$Acc = \frac{TP + FN}{N}$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2x(Precision * Recall)}{Precision + Recall}$$

IV. EXPERIMENTAL RESULTS

In the first phase of the experimental studies, dataset was examined using python scientific data science tools. In figure 3, a word clouds that belongs to dataset is shown. Due to the size of the canvas, the word cloud was drawn according to first 1000 tags and number of usage, not for all corpus.

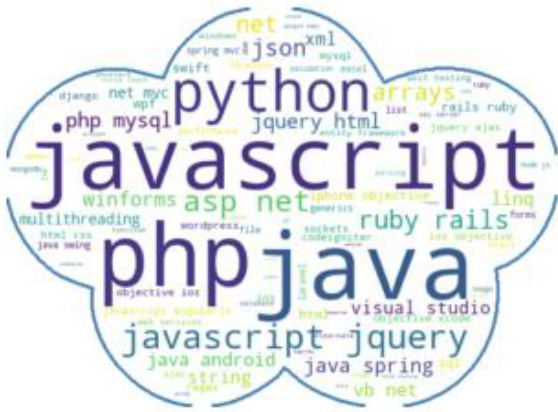


Figure 3: Word clouds for dataset

Dataset contains 100 unique tags. The least used tag in the data set is "mongodb" that repeated 350 times. The most used tag in the dataset is "javascript" that repeated 1978 times. According to Kaggle the most used tag is "java" [15]. It is believed that this is due to the fact that the word "java" has been passed twice in the "java" and "javascript" tags. However, in this study, two words are considered as different unique tokens. Figure 4 shows the 4 most commonly used labels and 4 least commonly used labels, and the number of repetitions of these labels.

Key	Type	Size	Value
mongodb	int	1	350
pointers	int	1	350
sorting	int	1	375
python-3.x	int	1	379
php	int	1	13907
java	int	1	18661
c#	int	1	19077
javascript	int	1	19078

Figure 4: Usage count of tags in dataset

Python scientific environment is used for experimental studies [18]. Metric values obtained using TF-IDF and BOW algorithms are shown in Table 1.

Table 1: Experimental results

Performance	Feature Extraction Method	
	TF-IDF	BOW
Acc	.34	.36
F1	.61	.65

Similar studies have been conducted with larger sized datasets and better or same results have been obtained in the literature [19-23]. The main objective in this study is to verify the method to be used. For example, in the Kaggle environment, there is also a Stack Overflow data set of 182 GB. However, a deep learning platform is needed to work with this data set. It is predicted that better results will be obtained with larger data sets in a deep learning platform.

Another experimental study is to find the words associated with each other. This process is important in data science for categorization of content, advertisement in commercial web sites and suggestions. Per question in the dataset, up to 9 keywords, at least 1 keyword and 2.8 (It was accepted as 3) words on average were used. Therefore, according to the n-gram algorithm n value 2 was taken. The results obtained are subjectively interpretable, so a confusion matrix is not created to obtain metric values such as accuracy. Some results are given in Table 2.

Table 2: Positive/Negative words

Input	Top positive words	Top negative words
Java	Spring, Android, Hibernate, Javafx	php, python, c#, javascript, c++
Php	php, laravel, codeigniter, yii, magento	java, c#, python, rails, aspnet
C#	c#, wpf, linq, aspnet, net	java, php, python, javascript, vbnet
Ruby	rails, nokogiri, sinatra, gem	rails, php, java, python, c#

In the first line in the Table 2, the predicted positive words for the "Java" input are "java" related technologies. Negative words, "java" related technology, framework or concept is not. Similar results are shown in the first three lines. All positive and negative words have been accurately estimated. But in the example given on line 4 in the Table 2, "rails" for the "ruby" entry was in both the positive and negative groups. In fact, "rails" is a framework used for web development with ruby programming language and should only be included in the positive group not negative group.

V. CONCLUSION

In this study, an experimental study was carried out to prediction the tag and to find the related tags on the "Stack Overflow" dataset which is published on Kaggle environment. Dataset contains one hundred thousand questions and labels of its. A supervised machine learning algorithm that uses a multi-class binary classifier on the data set has been performed. After the preprocessing on question text was vectorized using TF-IDF and BOW algorithms. The n-gram algorithm was used to find the words associated with each other. The study will be repeated using deeper learning algorithms on larger datasets in the future.

REFERENCES

- [1] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, P. Kuksa, "Natural language processing (almost) from scratch", *Journal of machine learning research*, 1, pp.2493-2537, 2011.
- [2] S. Sun, C. Luo, J. Chen, "A review of natural language processing techniques for opinion mining systems", *Information Fusion*, 36, pp.10-25, 2017.
- [3] J. Hirschberg, C. D. Manning, "Advances in natural language processing", *Science*, 349(6245), pp.261-266, 2015.
- [4] A. Khosravi, R. N. N. Koury, L. Machado, J. J. G. Pabon, "Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system", *Sustainable Energy Technologies and Assessments*, 25, pp.146-160, 2018.
- [5] W. Yin, K. Kann, M. Yu, H. Schütze, "Comparative study of cnn and rnn for natural language processing", *arXiv preprint arXiv:1702.01923*, 2017.
- [6] D. W. Otter, J. R. Medina, J. K. Kalita, "A survey of the usages of deep learning in natural language processing", *arXiv preprint arXiv:180.10854*, pp.1-35, 2018.
- [7] Z. Yang, W. Chen, F. Wang, B. Xu, "Generative adversarial training for neural machine translation", *Neurocomputing*, 321, pp.146-155, 2018.
- [8] "Google Translate", [Online], Available: <https://translate.google.com>.
- [9] S. Dhuria, H. Taneja, K. Taneja, "NLP and ontology based clustering An integrated approach for optimal information extraction from social web", *3rd International Conference on Computing for Sustainable Global Development*, pp.1765-1770, 2016.
- [10] M. T. Nguyen, D. V. Tran, C. X. Tran, M. L. Nguyen, "Learning to summarize web documents using social information", *28th International*

- Conference on Tools with Artificial Intelligence (ICTAI)*", pp.619-626, 2016.
- [11] X. Zhang, J. Zhao, Y. LeCun, "Character-level convolutional networks for text classification", *In Advances in neural information processing system*, pp.649-657, 2015.
- [12] J. Huang, M. Zhou, D. Yang, "Extracting Chatbot Knowledge from Online Discussion Forums", *In IJCAI*, 7, pp.423-428, 2007.
- [13] W. Maalej, Z. Kurtanović, H. Nabil, C. Stanik, "On the automatic classification of app reviews. Requirements Engineering", 21(3), pp.311-331, 2016.
- [14] "Stack Overflow", [Online], Available: <https://www.stackoverflow.com>
- [15] "Kaggle Stack Overflow Data Set", [Online], Available: <https://www.kaggle.com/badalgupta/stack-overflow-tag-prediction>
- [16] X. Xia, D. Lo, X. Wang, X. B. Zhou, "Tag recommendation in software information sites", *10th Working Conference on Mining Software Repositories*, pp. 287-296, 2013.
- [17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, J. Vanderplas, "Scikit-learn: Machine learning in Python. *Journal of machine learning research*", 12, pp.2825-2830, 2011.
- [18] K.J. Millman, M. Aivazis, "Python for scientists and engineers", *Computing in Science & Engineering*, 13(2), pp.9-12, 2011.
- [19] C. Stanley, M.D. Byrne, "Predicting tags for stackoverflow posts", *In Proceedings of ICCM*, 2013).
- [20] V. S. Rekha, N. Divya, P. S. Bagavathi, "A hybrid auto-tagging system for stackoverflow forum questions", *In Proceedings of the 2014 International Conference on Interdisciplinary Advances in Applied Computing*, 2014.
- [21] L. Short, C. Wong, D. Zeng, "Tag recommendations in stackoverflow", *San Francisco: Stanford University*, 2014.
- [22] S. Schuster, W. Zhu, Y. Cheng, "Predicting tags for stackoverflow questions", *CS229 Projects, Stanford university*, 2013.
- [23] J. Hong, M. Fang, "Keyword extraction and semantic tag prediction", 2013.

A Mobile Indoor/Outdoor Augmented Reality Application for Architecture

M.OKUR, M. BULUT, Y. SANTUR

Firat University, Elazig/Turkey, okurokur2014@gmail.com, muhtedibulut@gmail.com,
ysantur@firat.edu.tr

Abstract –Since the beginning of the 21st century, there have been revolutionary developments in technology with Industry 4.0. Artificial Intelligence has become a part of daily life in many fields, complex software has become capable of working in real time on devices such as mobile phones. Such advances in the technology get together humanity to Augmented Reality. In this field, both developed software are used for many purposes such as game, tourism, travel, medicine, military, industry, entertainment and education. In this study, a mobile application based on Augmented Reality working in both indoor and outdoor environment is presented in the architectural field.

Keywords – Augmented Reality, Architectural classification, Mobile Programming.

I. INTRODUCTION

WHEN the literature studies are examined, the subject "reality" is examined under three main headings as "Augmented Reality (AR)", "Virtual Reality (VR)" and "Mixed Reality (MR)" [1].

Peripherals such as helmets, goggles are based on the principle of using a computer-generated virtual world produced by the computer, completely abstracted from the real environment [2]. In other words, VR is based on the idea of deceiving human senses for various purposes. The basis of this idea was first introduced by computer scientist Sterling (2009) in 1965 [3]. With the development of screen technologies, this idea has been practically implemented and has become available in daily life in today's. In VR applications, virtual glasses or similar wearable technology is used. The user can thus be isolated from the real world. Its main applications are education, simulator, rehabilitation, game and entertainment [4]. Figure 1 shows a VR application using wearable technology and VR glasses.



Figure 1: VR application [5].

The AR is based on the principle of combining information such as pictures, video, audio, text produced by the computer on the real world image obtained by the mobile device or computer camera in a single scene and giving the user a different experience. In other words, AR brings the real world and the virtual world together. Thus, the user can interact with the virtual world without breaking away from the real world [6].

Today, AR technology is the most widely used for warning display in cars. Information such as vehicle speed, gear, oil, gasoline status can be traced from a panel that is traditionally placed on the vehicle's steering wheel. When the driver wants to access information, he has to keep his eye on the road. Another method that can be used is to give these warnings in audible intervals, but this method may distract the driver. The AR based warning screen is a built-in and / or projected on the windscreen. In this way, the driver can access his information on the panel without taking his eye off the road. Nowadays, many vehicle manufacturers now use this type of warning panels to reflect the vehicle's windshield. This increases driving safety. Figure 2 shows an example of the use of AR in vehicles.



Figure 2: AR application [7].

The combination of AR and VR with physical reality is defined as interaction. Environments which can place virtual objects in the real world, interact with physical reality and are evaluated between virtual and augmented reality are known. Another concept that emerges with the MR is the term immersion. While the user is in a virtual environment, he also feels in the real world with realistic virtual technology. The green screen technology used in the film industry, which enables the creation of advanced animation techniques, is the most popular example for the MR field [8].

II. LITERATURE OVERVIEW

In this section, a literature review about the use of AR in various disciplines has been done. One of the most widely used areas of AR is education [9, 10]. Wu and others(2013) outline technological, pedagogical, learning issues related to the implementation of AR in education [11].

Pikeraski and Thomas (2002) developed a desktop application that named "ARQuake" [12]. According to them, it is one of the first systems that allows users to play AR games outdoors—allowing them to move in the physical world, and at the same time experience computer-generated graphical monsters and objects [12].

Höllerer and others (1999) described "Mobile Augmented Reality System" testbed that employs different user interfaces to allow outdoor and indoor users to access and manage information that is spatially registered with the real world [13].

Vlahakis and others (2002) they proposed an approach that uses outdoor tracking, mobile computing, 3D visualization and AR techniques to enhance information presentation, reconstruct ruined sites, and simulate ancient life [14].

Webster and others (1996) described an experimental AR system that shows the location of columns behind a finished wall, the location of re-bars inside one of the columns, and a structural analysis of the column [15].

Wojciechowski and others (2004) conducted an AR-based study for museum exhibitions [16].

Wojciechowski and others (2004) described AR assisted laparoscopic adrenalectomy that the superimposition of virtual-reality reconstructions onto a real patient's images, in real time [17].

Ong and Nee (2013) viewed VR and AR applications in manufacturing [18]. A similar study was carried out by Nee and others (2012) [19].

Liarokapis and others (2004) present an educational application that allows users to interact with 3D Web content (Web3D) using virtual and augmented reality (AR) [20].

Michalos and others (2016) presented an AR tool for supporting operators where humans and robots coexist in a shared industrial workplace that provides AR visualization of the assembly process, video and text based instructions and production status updates [21].

Chong and others (2009) discussed the use of an augmented reality (AR) environment for facilitating intuitive robot programming, and presented a novel methodology for planning collision-free paths for a degree-of-freedom manipulator in a 3D AR environment [22].

Pettersen and others (2003) presented a demonstrator of a standalone AR pilot system allowing an operator to program robot waypoints and process specific events related to paint applications [23].

Agarwal (2016) examined the concept of AR and its various applications in civil engineering [24].

In this study, a mobile application which consists of a virtual model of AR-based architectural drawings in both indoor and outdoor environment was realized.

III. RELATED CONCEPTS

In this study, a mobile application which consists of a virtual model of AR-based architectural drawings in both indoor and outdoor environment was realized. Software tools and approach used in the following section are detailed.

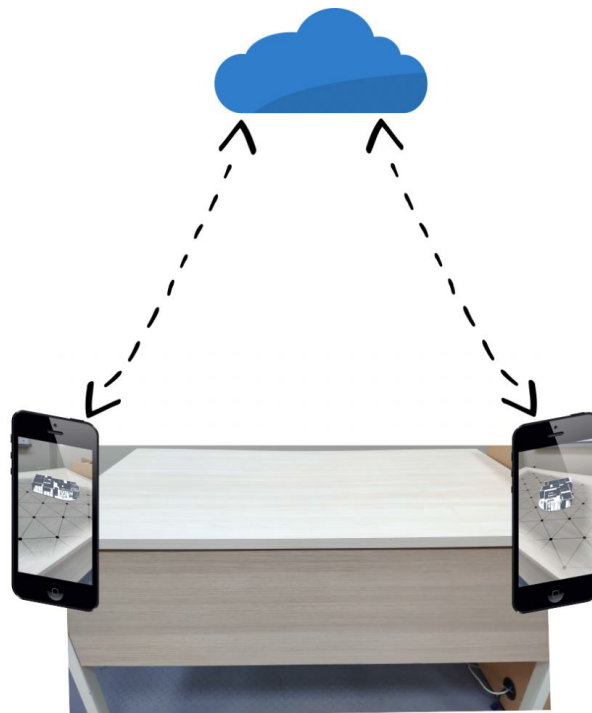


Figure 3: AR application for architecture

Figure 3 shows the main purposes of application, ARCore and ARKit is a technology that enables you to combine a virtual object extracted by Google into a real environment that used in the application. When using this combination, it uses three basic features [25, 26].

1.Motion Tracking: Identifies the location of the phone in the actual environment using operations called concurrent mapping. From the images it receives, it detects visually different features of the so-called feature points. It uses these points to calculate the change in position and to keep virtual objects constant at the specified point [26].

2.Environment Recognition: Provides usable information about flat surfaces such as floors, walls and tables, horizontally or vertically. It also determines the size of the surfaces and is made available for use in practice. The reason for using this feature is to enable virtual objects to be placed on flat surfaces [26].

3.Light Estimation: Collecting information about the lighting of the environment, it can provide the average intensity and color correction of the camera image. Thus, it provides a more realistic image by illuminating the virtual objects under the same conditions as the other objects in the environment [26].

Cloud Anchor, ARKit or any 3D virtual image created in ARCore allows multiple devices to use. This allows users to see and interact with 3D objects simultaneously.

With Cloud Anchor, users must connect to the server as in Figure 3 so that they can see virtual objects on their devices. Servers are used for "Host Anchor Sun and" resolve "Anchor Sun".

- **Host Anchor:** Provides visual map data obtained during the motion monitoring phase to be sent to the server from a user's environment. This data is processed on the server using a method called le Sparse Point Map Bu [27].

- **Resolve Anchor:** Virtual objects that the user has created in his or her environment are processed in Cloud. Other devices attempt to match this processed information to the ars Sparse Point Map physical method in their physical environment. Thus, in all devices connected to the Cloud, all the virtual objects created are placed according to their position [28].

In this study, the approach that used to implement ARCore technology is given in Figure 4. C# programming language and Visual Studio IDE was used and developed in the Unity game engine environment [29].

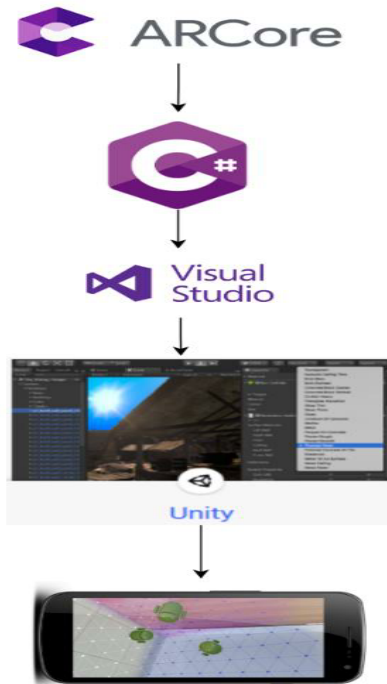


Figure 4: ARCore technology

As shown in Figure 5, the application will check if the ARCore SDK is installed when first opened. If installed, the user is asked to install the ARCore SDK. If the SDK is installed, the user will ask for permission to access the camera. After permission is granted, ARSession is created. The application detects the flat surfaces such as floor, wall and table from the images taken from the camera, horizontally or vertically, as in Figure 6.

In order to create virtual objects, the flat surfaces detected as shown in Figure 5 must be touched by the user. If the position the user touches is between the flat surfaces detected, it will create virtual objects.

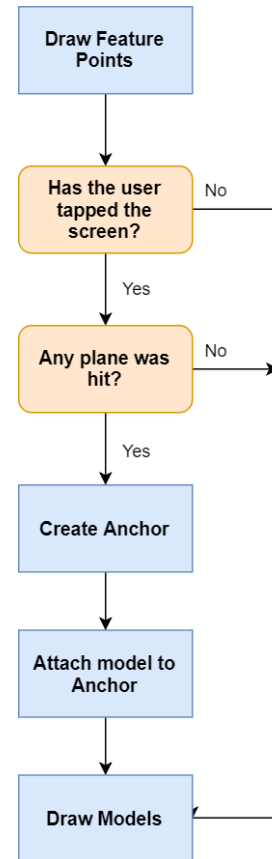


Figure 5: ARCore SDK

IV. EXPERIMENTAL RESULTS

In this section, indoor and outdoor examples of augmented reality based mobile application developed for architectural fields are given. Experimental results is shown in Figure 6. In the indoor environment, the floor is automatically detected and the table is created on its based augmented reality. Thus, the user can have an idea without the physical table about how the table will look in real environment.



Figure 6: An experimental result in indoor

A similar example is given in Figure 7. In this example, the application is tested in outdoor. In this scenario, a staircase and a single storey prefabricated house were formed on the parquet floor.



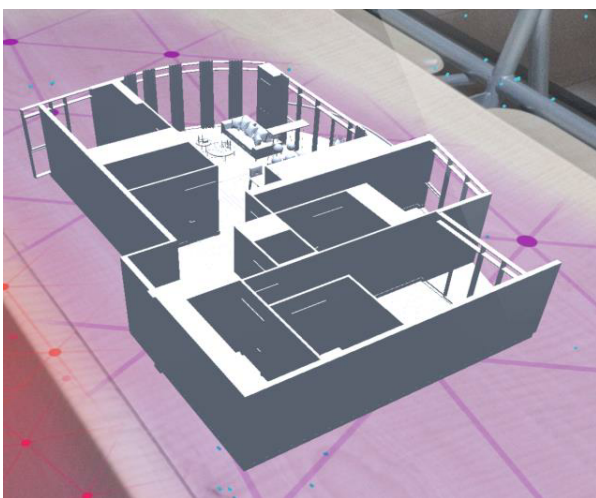
Figure 7: An experimental result in outdoor

The staircase is given closely in Figure 8. It is also shown with a human who touch it.

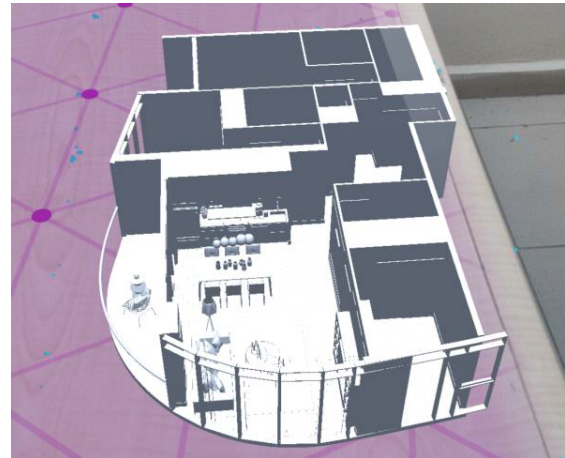


Figure 8: An experimental result in outdoor

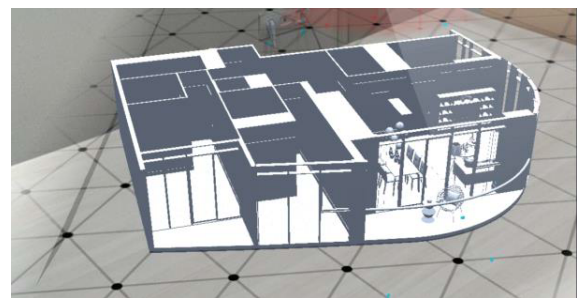
Some architectural models used in the application are given in Figure 9. These AR based models can be enlarged and scaled up to / down to according to the actual environment.



(a)



(b)



(c)

Figure 9: AR based architectural models

V. CONCLUSION

This application helps architects with AR technology to carry out project management processes easier and healthier. Before the realization of any architectural project, the project's 3D model is reflected in the real environment and provides the opportunity to see the problems that may occur before the project is implemented, thus reducing the project cost. On the other hand, it allows customers to see their own project in advance, allowing them to make the desired changes without realizing the project. This application will be re-develop through using cloud in the future. An API that for real-time mobile AR stream transmission between two users will be developed.

REFERENCES

- [1] S. K. Ong, A. Y. C Nee, "Virtual and augmented reality applications in manufacturing", Springer Science & Business Media, 2013.
- [2] W. R. Sherman, A. B. Craig, "Understanding virtual reality: Interface, application, and design", Morgan Kaufmann, 2018.
- [3] B. Sterling, "Augmented reality: 'The ultimate display' by Ivan Sutherland, 1965. Retrieved April, 25, 2017.
- [4] S. M. N. Glegg, D. E. Levac, "Barriers, facilitators and interventions to support virtual reality implementation in rehabilitation: a scoping review", *PM&R*, 10(11), pp.1237-1251, 2018.
- [5] [Online], Available: <http://www.teknolajilab.com>.
- [6] M. Billingham, A. Clark, G. Lee, "A survey of augmented reality", *Foundations and Trends® in Human-Computer Interaction*, 8(2-3), pp.73-272, 2015.
- [7] [Online], Available: <https://chipsnwafers.electronicsforu.com/2018/09/14/augmented-reality-head-up-display/>

- [8] L. Kobayashi, X.C. Zhang, S. A. Collins, N. Karim, D. L. Merck, "Exploratory application of augmented reality/mixed reality devices for acute care procedure training. Western Journal of Emergency Medicine", 19(1), 158, 2018.
- [9] M. Billinghurst, "Augmented reality in education", New horizons for learning, 12(5), 2002.
- [10] K. Lee, "Augmented reality in education and training", TechTrends, 56(2), pp.13-21, 2012.
- [11] H. K. Wu, S. W. Y. Lee, H. Y. Chang, J. C. Liang, "Current status, opportunities and challenges of augmented reality in education", Computers & education, 62, pp.41-49, 2013.
- [12] W. Piekarski, B. Thomas, "ARQuake: the outdoor augmented reality gaming system", Communications of the ACM, 45(1), pp.36-38, 2002.
- [13] T. Höllerer, S. Feiner, T. Terauchi, G. Rashid, D. Hallaway, "Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system", Computers & Graphics, 23(6), pp.779-785, 1999.
- [14] V. Vlahakis, M. Ioannidis, J. Karigiannis, M. Tsotros, M., Gounaris, D. Stricker, L. Almeida, "Archeoguide: an augmented reality guide for archaeological sites", IEEE Computer Graphics and Applications, 22(5), pp.52-60, 2002.
- [15] A. Webster, S. Feiner, B. MacIntyre, W. Massie, T. Krueger, "Augmented reality in architectural construction, inspection and renovation", In Proc. ASCE Third Congress on Computing in Civil Engineering, 1, p.1-7, 1996.
- [16] R. Wojciechowski, K. Walczak, M. White, W. Cellary, "Building virtual and augmented reality museum exhibitions", In Proceedings of the ninth international conference on 3D Web technology, pp.135-144, 2004.
- [17] J. Marescaux, F. Rubino, M. Arenas, D. Mutter, L. Soler, "Augmented-reality-assisted laparoscopic adrenalectomy", Jama, 292(18), pp.2211-2215, 2004.
- [18] S. K. Ong, A. Y. C Nee, "Virtual and augmented reality applications in manufacturing", Springer Science & Business Media, 2013.
- [19] A.Y. Nee, S. K. Ong, G. Chrysosolouris, D. Mourtzis, "Augmented reality applications in design and manufacturing", CIRP annals, 61(2), pp.657-679, 2012.
- [20] F. Liarokapis, N. Mourkousis, M. White, J. Darcy, M. Sifniotis, P. Petridis, P. F. Lister, "Web3D and augmented reality to support engineering education", World transactions on engineering and technology education, 3(1), pp.11-14, 2004.
- [21] G. Michalos, P. Karagiannis, S. Makris, Ö. Tokçalar, G. Chrysosolouris, "Augmented reality (AR) applications for supporting human-robot interactive cooperation", Procedia CIRP, 41, pp.370-375, 2016.
- [22] J. W. S. Chong, S. Ong, A. Y. Nee, K. B. Youcef-Youmi, "Robot programming using augmented reality: An interactive method for planning collision-free paths", Robotics and Computer-Integrated Manufacturing, 25(3), pp.689-701, 2009.
- [23] T. Pettersen, J. Pretlove, C. Skourup, T. Engedal, T. Lokstad, "Augmented reality for programming industrial robots", In The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, pp.319-320, 2003.
- [24] S. Agarwal, "Review on application of augmented reality in civil engineering", In International Conference on Inter Disciplinary Research in Engineering and Technology, 68, p.71, 2016.
- [25] J. Linowes, K. Babilinski, "Augmented Reality for Developers: Build Practical Augmented Reality Applications with Unity, ARCore, ARKit, and Vuforia. Packt Publishing Ltd, 2017.
- [26] [Online], Available:<https://developers.google.com/ar/discover>
- [27] [Online], Available:<https://developers.google.com/ar/develop/unreal/cloud-anchors/overview-unreal>
- [28] [Online], Available:<https://virocore.viromedia.com/docs/ar-cloud>
- [29] [Online], Available J. Liberty, "Programming C#: building. NET applications with C.", O'Reilly Media, Inc, 2015.

Reassembly of Synthetically Fractured Objects

Ö. ÇAKIR¹ and V. NABIYEV¹

¹ Karadeniz Technical University, Trabzon/Turkey, cakiro@ktu.edu.tr

¹ Karadeniz Technical University, Trabzon/Turkey, vasif@ktu.edu.tr

Abstract – In this paper we present a reassembly system for synthetically fractured 3D objects. Given as input synthetically fragmented parts of the tetrahedralized 3D vase model, for segmentation and matching stages we classify faces into two divisions whether they belong to the original model or fracture surface using indices of shared triangles as background knowledge. Then the vase model is reconstructed using an alignment algorithm based on unit sphere coordinates approach. The system is evaluated on several synthetically fracturing scenarios. The performance of the proposed system indicates that all reassembly computations are done in real-time.

Keywords – reassembly, fractured objects, 3d puzzle.

I. INTRODUCTION

THE reassembly of fractured 3D objects has gained an increasing importance in criminal policy, anthropology and especially in the field of archeology. It requires the solutions to two sub-problems, namely, matching and alignment to determine the shared faces between fragments and compute the inverse transformation matrices of the fragments prior to reassembly, respectively.

In this study a vase model is fractured synthetically. Triangles are used as the rendering primitive for fracture surfaces, so a volumetric triangulation method is needed for 3D points belongs to fracture surfaces. To this purpose, we use TETGEN [1] which generates not only tetrahedrons but also inner and outer triangles. As fracturing the vase model, the shared fracture surface of fragments is constructed using inner nearest triangles along the 3D fracturing path and duplicated for each fragment.

When fracturing is synthetical, the matching is simpler than the alignment because fracture surfaces are duplicated in each fragment. To classify a face as original versus fractured, we use the background knowledge of indices of shared triangles. For alignment, we use normal vectors of the fracture surfaces to compute the matrices of inverse rotation and translation. Each normal vector is matched with unit spherical (ϕ , θ) coordinates with respect to the reference vector. Details are discussed in section III.

In this paper, we present a reassembly system that automatically matches and aligns synthetically fractured 3D objects. We select a vase model and evaluate our reassembly system on several fracturing scenarios successfully. Performance results show that all reassembly computations are done in real-time.

II. RELATED WORK

The literature of the field of the reassembly of fractured 3D objects mostly focuses on the reconstruction of the broken archeological objects. To the best of our knowledge, the first work in reconstruction of the 3D objects is conducted by Papaioannou et al. [2] where fracture surfaces are assumed to be nearly planar and match each other one-to-one. The work in [2] use region-growing algorithm for surface segmentation. Then vertices of each face are projected in direction of the average face normal and the resulting depth maps are used for matching stage and extends matching method with boundary curve matching approach in [3].

Huang et al. [4] extract edges from point cloud data and faces are extracted from cycles of these edges. Original faces versus fracture faces segmentation are performed using graph cut algorithm runs on all faces of each fragment. For pairwise matching between two fractured surfaces they use patch-based surface features.

Winkelbach and Wahl [5] introduce a hierarchical clustering algorithm that decomposes each point set into a binary tree structure. The pairwise fragment matching is performed by descending the cluster trees simultaneously using depth-first search.

Altantsetseg et al. [6] proposed pairwise matching of 3D fragments using Fast Fourier Transform (FFT). They introduce two new descriptors: (a) the cluster of feature points extracted by using the curvature values of points and (b) curves along the principal directions of the cluster that are computed using Fourier series. They compare descriptor curves in terms of Fourier coefficients computed by FFT and total energies of curves.

Various other matching and alignment descriptors have been proposed such as curvatures [7], integral invariants [8] and RANSAC algorithms [9].

Recently, Son et al. [10] proposed reassembly of fractured objects using surface signature descriptor. The descriptor is based on whether a point on the fractured surface is convex or concave. The similarity between two fracture surfaces is calculated based on the spin images of feature curve points. For matching distance and normal deviation between two feature curves are used.

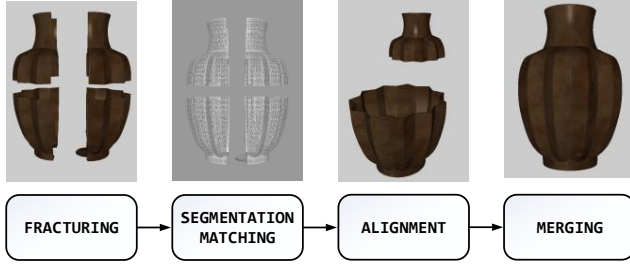


Figure 1: Overview of the reassembly system.

III. PROPOSED REASSEMBLY SYSTEM

Our reassembly system is composed of Fracturing, Segmentation, Matching, Alignment and Merging stages as shown in Figure 1.

A. System Inputs

Main input vase is a MAYA model that consists of 20494 vertices and 148692 triangles. Reassembly system requires various other inputs to be generated from this model:

- 1) *Shading Components*: Vase model is exported as *.OBJ file. Shading components like vertex, normal and texture coordinates are obtained from this *.OBJ file.
- 2) *Fragments*: Each fragment consists of original and fracture surfaces and each surface is defined as an array of indices of triangles.
- 3) *Inner Triangles*: Inner triangles are needed for the fracture surfaces. The model is exported as *.STL file and inner triangles are generated using TETGEN. Details are discussed in the following section.

B. Reassembly System

The fracturing is the first stage of the reassembly system where synthetical fracturing of the 3D vase model takes place. The vase is fractured using different planar fracturing scenarios including horizontal, vertical and both as shown in Figure 2. The system supports any planar fracturing. Vertices are split between fragments under regarding planar fracturing constraints.

We prefer triangles as a rendering primitive for fracture surfaces. To this purpose we use TETGEN which is a program to generate tetrahedral meshes. Besides tetrahedrons TETGEN also generates inner and outer triangles. We use inner triangles for fracture surfaces.

While fracturing, some vertices of the fracture surface are assigned to one fragment and some to other depending on regarding planar constraint. Please note that, inner triangles that have at least one vertex of two neighboring fragments are selected and duplicated in each fragment to generate fracture faces.



Figure 2: Various fracturing scenarios

In the stage of segmentation, every face is classified into two divisions whether it belongs to the original model or fracture surface using the background knowledge of indices of shared triangles because these indices are duplicated in fracturing stage. Background knowledge is also used for matching.

In the alignment stage, we use normal vectors of the fracture surfaces to compute the matrices of inverse rotation and translation. Each normal vector is matched with unit spherical (ϕ, θ) coordinates with respect to the *reference vector* $(1,0,0)$. To this purpose we generate a unit sphere that consists of 64440 points (ϕ varies in $[0,359]$, θ varies in $[1,179]$) as can be seen in Figure 3.

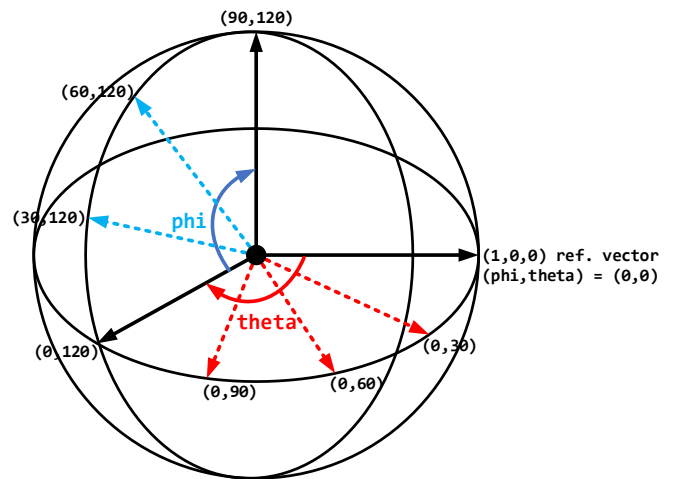


Figure 3: Unit sphere approach. Click sphere to watch animation.

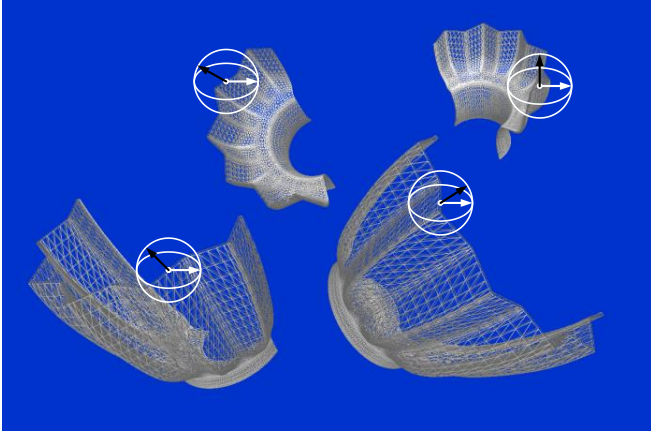


Figure 4: Spherical coordinates of the shared normal vectors.

Our reassembly system is implemented in DirectX 12. Interactive demo application runs on GeForce GTX 1050Ti GPU in real-time.

In demo application the vase model is firstly fragmented depending on regarding planar constraints. Then each fragment is rotated and moved to the arbitrary point in 3D space using keyboard. Using indices of shared triangles, the segmentation and matching tasks is performed. Then reassembly procedure executes and computes unit spherical coordinates of the shared normal vectors for each fragment as illustrated in Figure 4. Using spherical coordinates, the matrices of inverse rotation are calculated. Fragments are aligned using inverse rotation matrices and merged by translation.

Sample frames from the demo application are shown in Figure 5. In addition, a screen captured movie of the demo can be seen at: <http://ceng2.ktu.edu.tr/~cakir/icatces2019.html> The FPS values displayed in the movie, regarding the performance of the proposed system indicates that all computations are done in real-time.

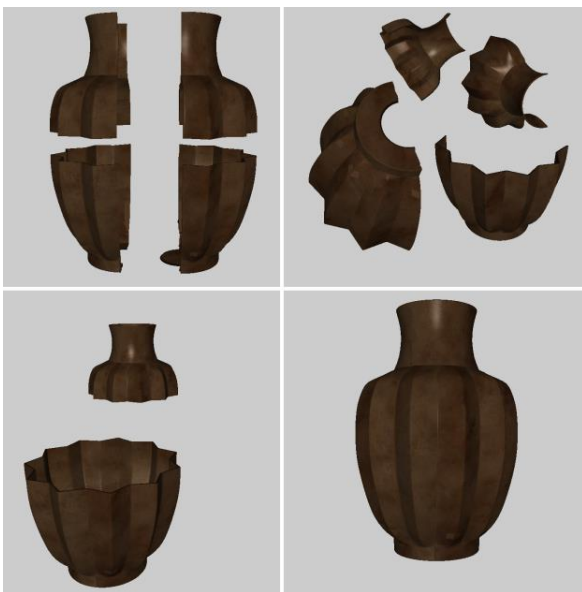


Figure 5: Sample frames from demo application.

IV. CONCLUSION

In this paper, a reassembly system that automatically matches and aligns synthetically fractured 3D objects is presented and its performance is evaluated on several fracturing scenarios successfully in real-time.

As a future work, a mechanism for reconstruction of real objects like 3D scanned archaeological remains will be developed. Since there are some erosions on fracture surfaces of archaeological remains, the reassembly processes including segmentation, matching and alignment becomes more difficult than presented system in this paper. We aim to propose a novel matching and alignment descriptor for lossy fragmented objects in the future.

REFERENCES

- [1] <http://wias-berlin.de/software/index.jsp?id=TetGen&lang=1>
- [2] G. Papaioannou, E. Karabassi, and T. Theoharis, "Virtual archaeologist: Assembling the past", IEEE Computer Graphics and Applications 21, 2, 53–59, 2001.
- [3] G. Papaioannou, E. Karabassi, "On the automatic assemblage of arbitrary broken solid artefacts", Image and Vision Computing 21, 401–412, 2003.
- [4] Q. X. Huang, S. Fly, N. Gelfand, M. Hofer, H. Pottmann, "Reassembling fractured objects by geometric matching". ACM Trans. Graph. (TOG) 25(3), 569–578, 2006.
- [5] S. Winkelbach, F. M. Wahl, "Pairwise matching of 3D fragments using cluster trees". Int. J. Comput. Vis. 78(1), 1–13, 2008.
- [6] E. Altantsetseg, K. Matsuyama, K. Konno, "Pairwise matching of 3D fragments using fast Fourier transform". Vis. Comput. 30(6), 929–938, 2014.
- [7] X Li, and I. Guskov, "Multiscale features for approximate alignment of point-based surfaces", In SGP'05, 217–226, 2005.
- [8] N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann, "Robust global registration", In SGP'05, 197–206, 2005.
- [9] M. A. Fischler, R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography". Commun. ACM 24(6), 381–395, 1981.
- [10] T. Son, J. Lee, J. Lim, K. Lee, "Reassembly of fractured objects using surface signature", Vis Comput. 34:1371–1381, 2018.

Classification of EEG patterns by using Katz Fractal Dimension

Ö. EMHAN¹, M. AKIN², M.S. ÖZERDEM³

¹ Dicle University, Diyarbakır/Türkiye, oemhan@dicle.edu.tr

² Dicle University, Diyarbakır/Türkiye, makin@dicle.edu.tr

³ Dicle University, Diyarbakır/Türkiye, sozerdem@dicle.edu.tr

Abstract – In this paper, EEG signals recorded from healthy individuals and epileptic patients were classified. In the literature, several complex domain transformations and statistical analyzes were used which increase the workload before feature extraction. We proposed a non-complex and low cost methods for automatic seizure detection. First, all EEG records were filtered with 0-40 Hz. Low pass filter (LPF). Then, features were extracted by Katz Fractal Dimension (KFD) method. Then, the extracted feature vectors were classified via the K-Nearest Neighborhood (KNN) and Support Vector Machine (SVM) algorithms. Finally, our proposed methods were compared among themselves and to other papers at the literature that have carried out on the same database.

Keywords – EEG, Epilepsy, Fractal Dimension, Support Vector Machine, K-Nearest Neighborhood

I. INTRODUCTION

EPILEPSY is one of the most common neurological disorder and is caused by abnormal activity of a group of neurons in the brain. More than 2% of the population worldwide has epilepsy at various levels. Approximately 2.4 million new cases are encountered each year [1, 2]. Epilepsy is the second most encountered neurological disorder after stroke. [3].

Epilepsy is usually diagnosed using electroencephalography (EEG) signals. Therefore, the extraction of effective features from EEG signals is a substantial process for correct classification. EEG is a method that measures brain's electrical activity. EEG signals are widely preferred for detection of neurological disorder [4] since it is low cost and painless method.

Trained professionals analyze the EEG records visually and try to detect seizures by means of this records [5]. Reviewing 24-h continuous EEG recording is a very costly method and prone to error. So automating the detection of epileptic seizures is quite important and accelerates and simplifies the process of treatment [6].

Therefore, in the literature, there are a lot of papers on automated EEG seizure detection. The latest directions in seizure detection area can be categorized several domains like *time domain*, *frequency domain*, *wavelet domain* and *empirical mode decomposition*. We presented 3 summaries of papers by this way and the last paper's summary contain fractal dimension feature.

Shanir and Kahn [7] proposed a method which uses two *time-domain* features called line length and energy. Linear discriminant analysis (LDA), Quadratic discriminant analysis (QDA) and (KNN) algorithms have been used and compared for classification. The best accuracy of 94.4% and 100% is achieved by the KNN classifier. Bhople et al. [8] proposed an epileptic seizure detection method by using fast Fourier transform (FFT) which is a *frequency domain* feature. They used FFT for feature extraction. Then features were fed to generalized feed-forward neural network (GFFNN) and multilayer perceptron (MLP) for classification. Bonn database was used and 100% accuracy achieved. Hasan et al. [9] proposed a seizure detection method by using *Wavelet* and *Hilbert transforms*. Mean, maximum, minimum, standard deviation and average power of absolute values of wavelet and Hilbert transform coefficients were calculated and were used as features. Each feature was fed to KNN classifier separately. The method was tested on publicly available Bonn dataset. Wavelet based features: 100% and 96% accuracy achieved on A-E and B-E comparisons respectively. Hilbert based features: 100% and 100% accuracy achieved on A-E and B-E comparisons respectively. Sharma et al. [10] proposed a novel method for detecting epileptic seizures using analytic time-frequency flexible wavelet transform (ATFFWT) and fractal dimension (FD). They have used ATFFWT to decompose EEG signals into the desired sub-bands. Then FDs of each sub-band were calculated and feature vectors were created. Finally, these feature vectors have been fed to the least-squares support vector machine (LS-SVM) classifier. The performance is tested on Bonn database and the accuracy of 100% normal-ictal, 98.67% non-ictal-ictal and 92.5% normal-non ictal achieved.

In this study, we proposed a non-complex and low cost automatic seizure detection method which make it attractive. Features were extracted with KFD method from raw EEG records. Extracted feature vectors were classified by KNN and SVM classifiers with 10-Fold cross validation. Fig. 1 shows flow chart of our proposed system.

The remainder of this paper organized as follows. Section II describes the dataset used in this paper. Feature extraction, classification algorithms and performance metrics will be explained in Section III. In section IV the results obtained with the proposed methodology will be presented. And paper will be concluded in section V.

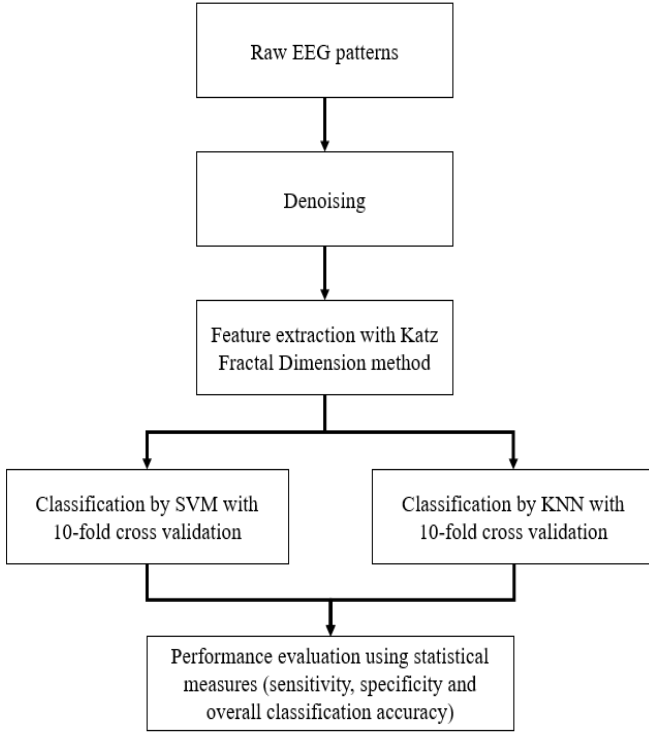


Figure 1: Flow chart of proposed system

II. DATASET

EEG data that used in this study has been acquired from the database of Epileptology Department of Bonn University [11]. There are five types of EEG signal (A, B, C, D, E)

Set A: five healthy subjects with eyes open,

Set B: five healthy subjects with eyes closed

Set C: seizure-free inter-ictal signals recorded from the hippocampal formation of the opposite hemisphere

Set D: seizure-free inter-ictal signals recorded from the epileptogenic zone

Set E: seizure signals

For each type of EEG signal, there are 100 single-channel scalp EEG segments. Each channel has 4096 samples. These EEG signals were sampled at 173.61 Hz for 23.6 s.

III. METHOD

The concept of ‘Fractal’ has been first mentioned by Mandelbrot to describe self-similarity of complex formations. The FD is used to measure dimensional complexity and self-similarity of non-stationary signals like EEG signals [12]. Since FD gives the measure of complexity of a signal, bio-signals like EEG, which are non-stationary and nonlinear can be analyzed using FD methods. FD can be calculated by several methods. In this study, KFD method were used for feature extraction.

A. Katz Fractal Dimension

Katz proposed a method for FD calculation: the ratio of the length of the curve L that calculated by Euclidean method, divided by the maximum Euclidean distance d that can be expressed as

$$FD = \frac{\log_{10}(L)}{\log_{10}(d)} \quad (1)$$

L is the total length of the curve or the sum of the Euclidean distances between successive points:

$$L = \sum_{i=1}^N \text{dist}(s_i, s_{i+1}), \quad i = 1, 2, \dots, N-1 \quad (2)$$

and d is the diameter of the curve, means the distance from the first point of the array to the point that is the farthest distance to the first point (Fig. 2). d can be expressed as

$$d = \max\{\text{distance}(s_1, s_i)\}, \quad i = 1, 2, \dots, N \quad (3)$$

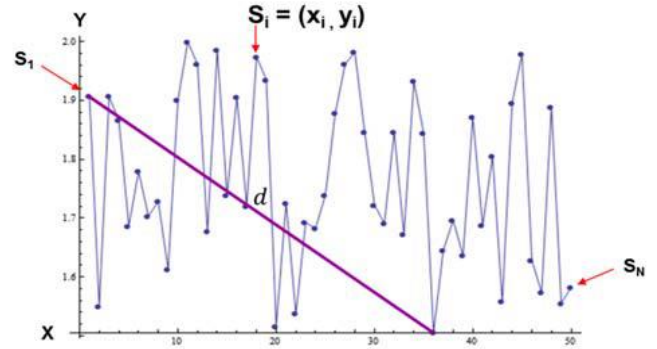


Figure 2: diameter of the curve d in the S array

Katz proposed to normalize d and L , the average distance between successive points, $a=L/N$, where N is the number of steps in the array. In this way finally KFD can be calculated as follows [13]

$$FD = \frac{\log_{10}(L/a)}{\log_{10}(d/a)} = \frac{\log_{10}(N)}{\log_{10}(N) + \log_{10}(d/L)} \quad (4)$$

B. Classification

Features extracted with KFD method from raw EEG dataset were classified by KNN and SVM algorithms with 10-fold cross validation in this study.

1) *K-Nearest Neighborhood (KNN)* is a non-parametric learning algorithm based on the principle of determining which class an unknown object (Fig. 3-a) belongs according to distance measurement (Fig. 3-b). The most common distance measurement methods are Euclidean, Mahalanobis, Minkowski, Manhattan etc. In the KNN algorithm k neighborhood value is chosen and then the unknown object is assigned based on majority rule to the lowest distance class (Fig. 3-c) [14].

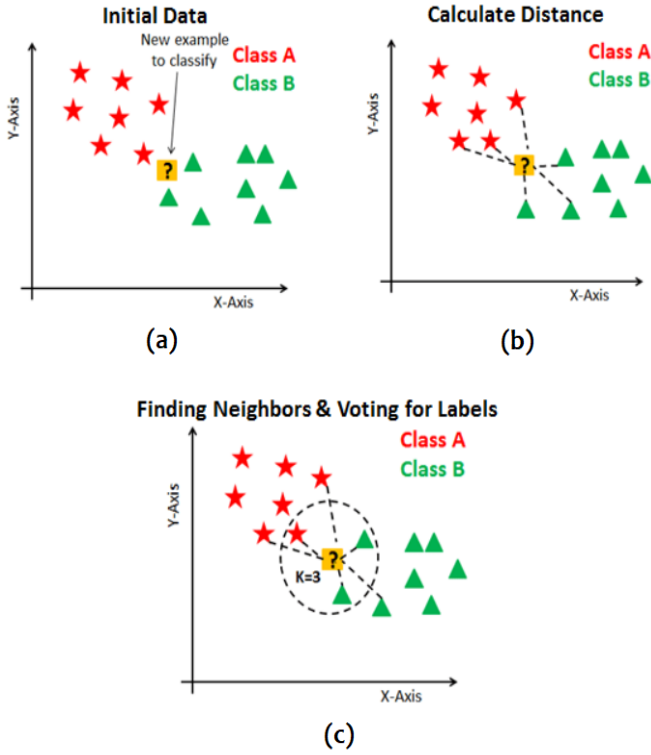


Figure 3: Explanation of KNN algorithm on X-Y axis [15]

2) *Support vector machine (SVM)* is a non-parametric supervised learning method, based on the structural risk minimization. SVM is particularly successful in binary classification problems. As can be seen in Fig. 4-a, several lines could be drawn in a way that is the farthest distance from both classes. In the case of linearly separated classes, ‘optimal’ hyperplane is defined as boundary which separates both training set and unknown samples. This boundary is the farthest distant hyperplane from both sets. The training vectors closest to the boundary are called support vectors (Fig. 4-b). The margin is the minimal distance from the separating hyperplane to the closest data points [14].

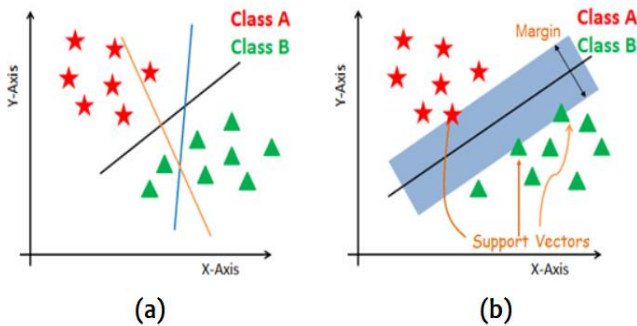


Figure 4: Explanation of SVM algorithm on X-Y axis [16]

K-fold cross validation

Cross-validation is a method that used to evaluate machine learning algorithms in case of limited data sample. In this study, a 10-fold cross validation was applied to test the robustness of the proposed model. The 100 EEG signals from each set were partitioned into ten disjoint part. 9 parts were used as training sets, while the remaining 1 part was used as testing set. The same process was repeated ten times part by part. So each of the 10 parts were used as the testing part as shown in Fig. 4. A total of ten trials were run and the average values for each performance metric is reported [17].

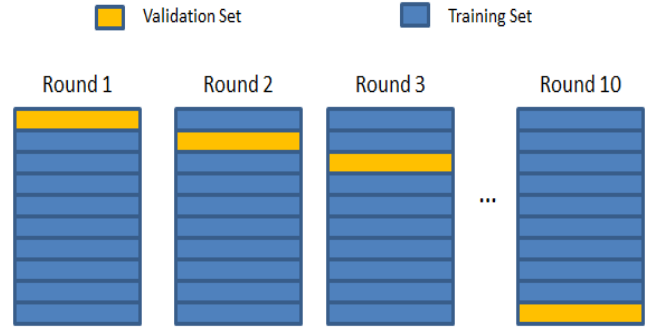


Figure 4: 10-fold cross validation [18]

C. Performance Measurement

In this paper we assessed the performance using criteria that are generally used in biomedical research such as *sensitivity* (proportion of the correctly classified normal class to the total number of labeled normal class), *specificity* (proportion of the correctly classified abnormal class to the total number of labeled abnormal class) and *accuracy* (proportion of the correctly classified EEGs out of the total number of EEGs). These performance measurements metrics could be formulized as follows

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \quad (5)$$

$$Specificity = \frac{TN}{TN + FP} \times 100 \quad (6)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (7)$$

where TP → true positive, is the number of epochs which are marked as seizure.

TN → true negative, is the number of epochs which are marked as non-seizure.

FN → false negative, the number of seizure epochs which are recognized as non-seizure but actually they are seizures.

FP → false positive, the number of non-seizures epochs which are recognized as seizure but actually they are non-seizures [19].

IV. RESULTS AND DISCUSSION

In this section, the results obtained with the proposed methodology are presented and analyzed.

The features extracted with KFD are presented at Fig. 6-9. It is clearly seams from these figures that the features extracted with KFD method are visually distinguishable.

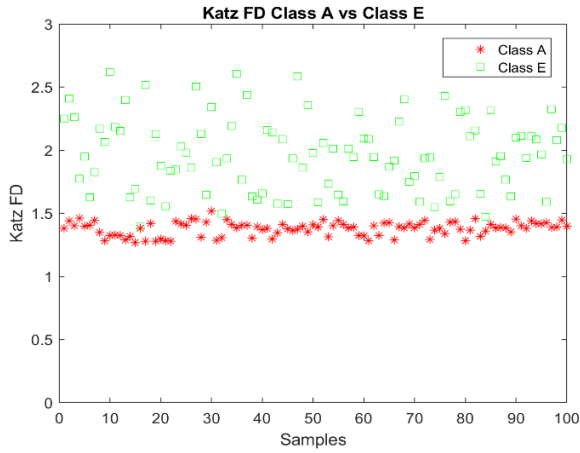


Figure 6: KFD value Class A vs Class E

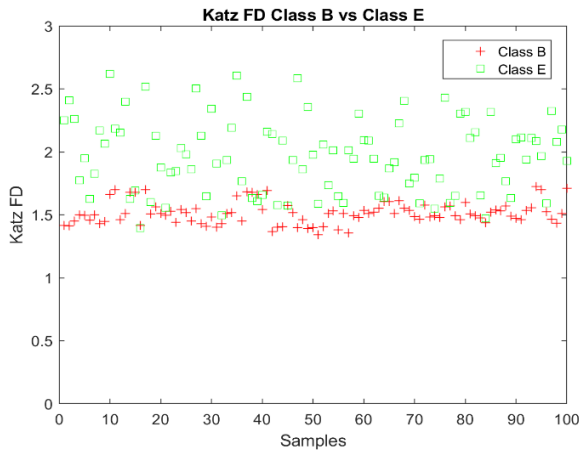


Figure 7: KFD value Class B vs Class E



Figure 8: KFD value Class C vs Class E



Figure 9: KFD value Class D vs Class E

In order to confirm the robustness of the classification performance measurement metrics, two different classification algorithms (KNN and SVM) were used with 10-fold cross validation. The KNN classification performance metrics were presented in Table 1 and SVM classification performance metrics were presented in Table 2 respectively.

The results were presented to represent class A and B healthy subjects, class C and D inter-ictal subjects and class E ictal subjects.

Case 1: healthy vs ictal (A vs E, B vs E, AB vs E)

Case 2: inter-ictal vs ictal (C vs E, D vs E, CD vs E)

Case 3: healthy & inter-ictal vs ictal (ABCD vs E)

Table 1: The performance metrics of KNN

Classes	Sensitivity %	Specificity %	Accuracy %
A - E	100	97	98,5
B - E	98	91	93
C - E	98	97	97,5
D - E	97	97	97
AB - E	99	91	95
CD - E	98,5	97	97,33
ABCD - E	98,75	92	96,8

Table 2: The performance metrics of SVM

Classes	Sensitivity %	Specificity %	Accuracy %
A - E	100	96	98
B - E	98	75	86
C - E	98	97	97,5
D - E	98	97	97
AB - E	99	75	91
CD - E	99	92	96,67
ABCD - E	99,75	74	94,6

The results presented in Table 1 and Table 2 clearly shows that KFD method is useful for classifying and diagnosing seizure and seizure free cases based on EEG. Highest accuracy of 98.50% is achieved with KNN classifier for case 1. Accuracy of 96.8% at KNN and 94,6% at SVM was achieved for case 3.

The overall accuracy and specificity metrics show that KNN classification algorithm is better than SVM classification algorithm. On the other hand, SVM is a little better than KNN in terms of sensitivity metric.

Especially, for case1 and case 3 KNN algorithm is more superior than SVM algorithm.

Finally, in Table 3 several automated EEG detection systems have been comparatively presented that have carried out on the same database [20]-[22]. According to this results it can be concluded that our proposed method has acceptable accuracy.

V. CONCLUSION

Automatic seizure detection is significant for the diagnosis of epilepsy and the reduction of massive workload for reviewing continuous EEG recordings. The main contribution of this paper is to provide a non-complex and effective automated feature extraction algorithm for epileptic seizure detection in EEG signals. We proposed KFD method for feature extraction from raw EEG records which is a time domain feature. According to the obtained results, it can be concluded that KFD features represent the characteristics of the EEG signal and also its ability to classify epileptic and normal segments with acceptable accuracy, less computational complexity, less time and low cost.

Table 3: Several studies employing with same dataset

Author	Methods and features	Cross Validation	Type of classification	Accuracy %
Gandhi et al.[20]	DWT and energy, std and entropy features using SVM and Probabilistic NN	Yes	ABCD-E	95,44
Nicolaou et al.[21]	Permutation entropy and SVM	No	A-E	93,55
			B-E	82,88
			C-E	88,00
			D-E	79,94
			ABCD-E	86,10
Swami et al.[22]	DTCWT and energy, std, Shannon entropy features using GRNN	Yes	A-E	100
			B-E	98,89
			C-E	98,72
			D-E	93,33
			ABCD-E	95,24
Our proposed method	KFD feature using KNN	Yes	A-E	98,5
			B-E	93
			C-E	97,5
			D-E	97
			ABCD-E	96,8

REFERENCES

- [1] National Institute of Neurological Disorders and Stroke. <https://www.ninds.nih.gov/Disorders/All-Disorders/Epilepsy-Information-Page> 2019 (online).
- [2] Alotaiby TN, Alshebeili SA, Abd El-Samie FE (2016) Channel selection and seizure detection using a statistical approach. *IEEE Explore*.
- [3] World Health Organization (WHO), Epilepsy fact sheet, 2019 (online). http://www.who.int/mental_health/neurology/epilepsy/en/
- [4] Zhou W, Liu Y, Yuan Q, Li X (2013) Epileptic seizure detection using lacunarity and Bayesian linear discriminant analysis in intracranial EEG. *IEEE Trans Biomed Eng* 60(12):3375–3381
- [5] H. Adeli, Z. Zhou, and N. Dadmehrc, “Analysis of EEG records in an epileptic patient using wavelet transform,” *J. Neurosci. Methods*, vol. 123, no. 1, pp. 69–87, 2003.
- [6] Y. Tang and D. M. Durand, “A tunable support vector machine assembly classifier for epileptic seizure detection,” *Expert Syst. Appl.*, vol. 39, pp. 3925–3938, 2012.
- [7] Shanir PP, Khan YU (2015) Time domain analysis of EEG for automatic seizure detection. *ETEEE*
- [8] Bhople AD (2012) Fast Fourier transform based classification of epileptic seizure using artificial neural network. *Int J Adv Res Comput Sci Softw Eng* 2(4). ISSN: 2277 128X
- [9] Polat H, Ozerdem MS (2016) Epileptic seizure detection from EEG signals by using wavelet and Hilbert transform. *MEMSTECH 2016, Polyana-Svalyava (Zakarpattya)*. 20–24 April 2016
- [10] Sharma M, Pachori R, Rajendra U (2017) A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension. *Pattern Recognition Letters* (94):172-179
- [11] Andrzejak, R. G., Lehnertz, K., Rieke, C., Mormann, F., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity. *Physical Review E*, 64, 061907.
- [12] Roberto A., Salazar R. Classification of EEG signals using fractal dimension features and artificial neural networks. *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*
- [13] S. M. Fernandez-Fraga, J. Rangel (2017). Comparison of Higuchi, Katz and Multiresolution Box-Counting Fractal Dimension Algorithms on EEG Waveforms Signals Based on Visual Evoked Potentials. *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 15, No. 7, July 2017
- [14] Luis A. Berrueta, Rosa M. Alonso-Salces, K'aroly H'eberger b.2007 Supervised pattern recognition in food analysis. *Journal of Chromatography A*, 1158 (2007) 196–214.
- [15] <https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn>
- [16] <https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python>
- [17] Zainuddin Z. On the use of wavelet neural networks in the task of epileptic seizure detection from electroencephalography signals. *Procedia Computer Science* 11 (2012) 149 – 159
- [18] <https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61be44b6>
- [19] Paul, Y.: Various epileptic seizure detection techniques using biomedical signals: a review. *Brain informatics* 5(2), 6 (2018)
- [20] T. Gandhi , B.K. Panigrahi , S. Anand , A comparative study of wavelet families for EEGsignal classification, *Neurocomputing* 74 (17) (2011) 3051–3057 .
- [21] N. Nicolaou , J. Georgiou , Detection of epileptic electroencephalogram based on permutation entropy and support vector machines, *Expert Syst. Appl.* 39 (1) (2012) 202–209 .
- [22] P. Swami , T.K. Gandhi , B.K. Panigrahi , M. Tripathi , S. Anand , A novel robust diagnostic model to detect seizures in electroencephalography, *Expert Syst. Appl.* 56 (2016) 116–130 .

Learning Semi-Supervised Nonlinear Embeddings for Domain-Adaptive Pattern Recognition

Elif Vural

Middle East Technical University, Ankara/Turkey, velif@metu.edu.tr

Abstract—We study the problem of learning nonlinear data embeddings in order to obtain representations for efficient and domain-invariant recognition of visual patterns. Given observations of a training set of patterns from different classes in two different domains, we propose a method to learn a nonlinear mapping of the data samples from different domains into a common domain. The nonlinear mapping is learnt such that the class means of different domains are mapped to nearby points in the common domain in order to appropriately align the two domains. Meanwhile, unlabeled samples are also used in the computation of the embedding via an objective term representing the preservation of the global geometry of the data. Along with the mapping of the training points, we also learn a linear classifier in the common domain, which allows an accurate estimation of the unknown class labels. We evaluate the performance of the proposed algorithm in domain-adaptive face and object recognition experiments. Experimental results show that the proposed method yields quite promising performance, outperforming baseline domain adaptation methods.

Keywords—Domain adaptation, pattern recognition, nonlinear dimensionality reduction, semi-supervised learning.

I. INTRODUCTION

Traditional machine learning methods are based on the assumption that the training and the test samples are sampled from the same distribution. Meanwhile, in many practical pattern recognition problems, the distribution of the training samples may differ from that of the test samples. For instance, in a face recognition problem, the labeled training images of the subjects may be captured from a certain camera angle, whereas the unlabeled test images might be captured from a different camera angle. Machine learning algorithms that aim to address classification problems when data is sampled from two or multiple different domains are called domain adaptation algorithms. Domain adaptation methods use the training data from a source domain where class labels are mostly known, in order to improve the performance of classification in a target domain where very few class labels are known.

The main approaches in domain adaptation are as follows. Sample reweighting algorithms are preferable when the deviation between the source and the target domains is relatively small [1], [2]. In cases where the deviation between the two domains is significant, some methods map the data into a common space of large dimension via feature augmentation [3], [4]. Meanwhile, one of the most common approaches is to map the source and the target data into a common domain with a projection in order to align them [5]. Such projections can be learnt in an unsupervised way [6], [7], while some

methods learn a classifier as well by making use of labeled data [8], [9].

Most of the domain adaptation methods aligning the data in a common domain learn linear mappings [8], [9]. Some methods employing nonlinear mappings exist; however, most of them are simply based on kernel extensions of linear methods [5], [10]. Restricting the learnt mapping to linear projections may result in poor classification performance especially in settings where the discrepancy between the source and the target distributions is large. Nonlinear kernel extensions usually perform better; however, the learnt representations are still restricted to a certain family of functions. In this work, we propose to learn nonlinear representations of data for domain adaptation in order to circumvent such shortcomings of the previous methods. The representations learnt with the proposed method are completely nonparametric. This feature allows the proposed algorithm to adapt well to the particular geometric structure of the data, and also successfully deal with the problem of large discrepancy between the source and the target distributions.

The proposed method is semi-supervised, thus, can make use of both labeled and unlabeled data from the source and the target domains. We embed the source and the target samples into a common domain via a nonlinear mapping by computing a new representation of each data sample in the common domain. When learning the new coordinates of data samples, the following are taken into account: First, the source and the target domains are aligned by minimizing the deviation between the means of data samples from the same class but different domains, where we propose to use a modified supervised version of the common MMD (maximum mean discrepancy) measure [5], [10], [11]. Unlabeled data samples are included in the learning objective with a term preserving the geometric structure of data samples from both domains. The parameters of a linear classifier are also learnt along with the coordinates of the data samples. We form a learning objective that jointly optimizes the embedding of data samples and the linear classifier parameters. The embedding and the classifier parameters are optimized with an alternating optimization scheme.

The proposed method is tested in domain-invariant pattern classification experiments in a face recognition and an object recognition problem. The experimental results show that the proposed semi-supervised learning framework outperforms baseline domain adaptation approaches in both data sets. Our results can be interpreted in the way that the nonlinear and

nonparametric representations learnt with the proposed method are quite flexible and give high classification accuracy despite the fact that the end-stage classification is done with a simple linear classifier.

II. PROPOSED METHOD

A. Problem Formulation

We first set the notation and then formulate our problem for learning domain-adaptive nonlinear mappings.

Let $\{x_i^s\}_{i=1}^{N_s} \subset \mathcal{X}^s$ be a set of source data samples in a source domain \mathcal{X}^s ; and $\{x_j^t\}_{j=1}^{N_t} \subset \mathcal{X}^t$ be a set of target data samples in a target domain \mathcal{X}^t . Without loss of generality, we assume that the first M_s samples $\{x_i^s\}_{i=1}^{M_s}$ in the source domain and the first M_t samples $\{x_j^t\}_{j=1}^{M_t}$ in the target domain are labeled, and the remaining data samples are unlabeled.

In our study, we consider a setting with a large discrepancy between the source and the target distributions. In order to make use of the information in both domains, we propose to learn a nonlinear mapping of the samples into a new common domain and learn a classifier in the new domain. Hence, we aim to map each source sample x_i^s and target sample x_j^t , respectively to the new coordinates $z_i^s \in \mathbb{R}^d$ and $z_j^t \in \mathbb{R}^d$ in a common domain \mathbb{R}^d . Our method then consists of learning the new coordinates $\{z_i^s\}_{i=1}^{N_s}$ and $\{z_j^t\}_{j=1}^{N_t}$ as well as the parameters W of a linear classifier in \mathbb{R}^d . The studied problem is illustrated in Figure 1.

Let Z^s and Z^t respectively denote the matrices consisting of the embeddings of the source and the target data samples as

$$Z^s = [z_1^s \dots z_{N_s}^s] \in \mathbb{R}^{d \times N_s}, \quad Z^t = [z_1^t \dots z_{N_t}^t] \in \mathbb{R}^{d \times N_t}.$$

Let us also denote as $Z = [Z^s \ Z^t]$ the matrix consisting of all coordinates to be learnt. We assume that the data samples in both domains belong to the classes $c = 1, \dots, K$.

When forming the optimization problem for learning the nonlinear embedding, we consider the following objectives:

Alignment of the source and the target domains: In the computation of the new data coordinates Z , we aim to align the source and the target domains so that a common classifier can be learnt. For this purpose, we include the following alignment objective in our objective function

$$f_{align}(Z) = \sum_{c=1}^K \left\| \frac{1}{M_s^c} \sum_{i:C(x_i^s)=c} z_i^s - \frac{1}{M_t^c} \sum_{j:C(x_j^t)=c} z_j^t \right\|^2 \quad (1)$$

where $C(\cdot)$ denotes the class label of a sample; and M_s^c and M_t^c denote the number of labeled samples from class c , respectively in the source and the target domains. The alignment objective $f_{align}(Z)$ is the total distance between the means of the source and the target samples from the same classes. Thus, it aims to align the two domains so that the class means coincide as much as possible in the common domain. Note that the proposed alignment objective bears some resemblance to the MMD (maximum mean discrepancy)

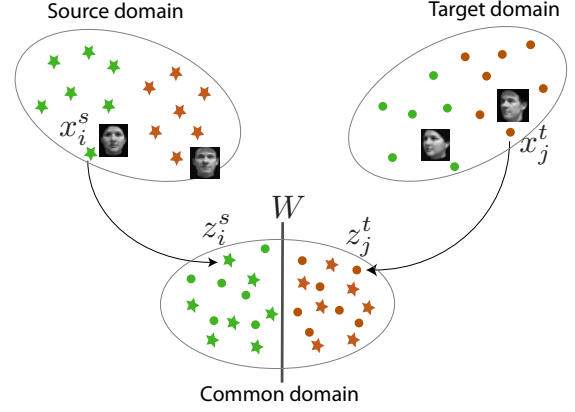


Figure 1: Illustration of the proposed method. The data samples in the source and the target domains have different distributions (e.g., in this illustration the source domain consists of front-view images of two different subjects, whereas the target domain consists of profile-view images of the subjects). In order to efficiently learn a common classifier, the samples in the two domains are aligned in a common domain via nonlinear mappings.

measure widely used in the KMM (kernel mean matching) methods in the literature. The commonly used MMD measure is an unsupervised metric that simply represents the distance between the means of the embeddings of the samples in the two domains. The proposed alignment objective differs from KMM methods in two aspects: First, it uses a supervised extension of MMD that aims to align not only the two domains but also the individual classes in the two domains. Next, the learnt embedding is nonparametric in the proposed method, whereas KMM methods use a parametric kernel function.

In order to write (1) in a more convenient form, we define a matrix $E \in \mathbb{R}^{(N_s+N_t) \times K}$ such that

$$E_{ic} = \begin{cases} 1/M_s^c, & \text{if } C(x_i^s) = c \\ -1/M_t^c, & \text{if } C(x_j^t) = c, \text{ for } i = j + N_s \\ 0, & \text{otherwise} \end{cases}$$

where $(\cdot)_{ij}$ denotes the (i, j) -th entry of a matrix. This definition allows us to write the subtraction in (1) in the form of a matrix product, hence resulting in the reformulation

$$f_{align}(Z) = \|ZE\|_F^2 \quad (2)$$

where $\|\cdot\|$ denotes the Frobenious norm of a matrix.

Linear separability of the classes: We would like to compute the embedding Z in such a way that a linear classifier can accurately estimate the class labels. Let Z_l^s and Z_l^t denote the submatrices of Z^s and Z^t consisting only of the labeled samples

$$Z_l^s = [z_1^s \dots z_{M_s}^s] \in \mathbb{R}^{d \times M_s}, \quad Z_l^t = [z_1^t \dots z_{M_t}^t] \in \mathbb{R}^{d \times M_t}.$$

Let us define the one-hot label matrices $Y^s \in \mathbb{R}^{K \times M_s}$ and $Y^t \in \mathbb{R}^{K \times M_t}$ containing the known class labels such that

$$Y_{ci}^s = \begin{cases} 1, & \text{if } C(x_i^s) = c \\ 0, & \text{otherwise} \end{cases}$$

and

$$Y_{cj}^t = \begin{cases} 1, & \text{if } C(x_j^t) = c \\ 0, & \text{otherwise} \end{cases}$$

Let $W \in \mathbb{R}^{d \times K}$ denote the matrix containing the parameters of the linear classifier to be learnt in the common domain \mathbb{R}^d . Then, in order to have a small classification error in the common domain, we include the following regularized total quadratic loss in the learning objective

$$f_{loss}(Z, W) = \|W^T Z_l^s - Y^s\|_F^2 + \alpha_1 \|W^T Z_l^t - Y^t\|_F^2 + \alpha_2 \|W\|_F^2. \quad (3)$$

The term $\|W\|_F^2$ is included in the objective function for the purpose of regularizing the classifier parameters.

Preservation of the geometric structure: Last, we discuss how to make use of the unlabeled data in the learnt representations. Previous studies have shown that the preservation of the geometric structure when learning low-dimensional data representations has large benefits in classification and regression problems [12], [13], [14]. This is typically achieved by forming a neighborhood graph with the data in the original domain, and computing the embedding such that sample pairs connected with strong edges in the original graph are mapped to nearby coordinates in the new domain.

In order to formulate this, let W^s denote a source weight matrix such that

$$(W^s)_{ij} = \exp(-\|x_i^s - x_j^s\|^2 / \sigma^2)$$

gives the weight of the edge between the sample pair (x_i^s, x_j^s) . The diagonal degree matrix D^s is then defined as [12]

$$(D^s)_{ii} = \sum_j (W^s)_{ij}.$$

Finally, the source graph Laplacian matrix is given by $L^s = D^s - W^s$. The weight matrix W^t and the graph Laplacian matrix L^t of the target domain are defined similarly. Graph Laplacian matrices have the following important property: For a graph function $z \in \mathbb{R}^N$ defined over an N -node graph with weight matrix W and graph Laplacian L , the expression

$$\sum_{i,j=1}^N (z_i - z_j)^2 W_{ij} = z^T L z$$

gives a measure of how fast the function z varies over the graph [12]. Motivated by this observation, we propose to include the following smoothness objective in our optimization problem

$$f_{smooth}(Z) = \text{tr}(Z^s L^s (Z^s)^T) + \beta \text{tr}(Z^t L^t (Z^t)^T) \quad (4)$$

where $\text{tr}(\cdot)$ denotes the trace of a matrix. The term in (4) enforces the coordinates of the embedding to vary slowly

on the data graph, which encourages the algorithm to map nearby data samples in each domain to nearby coordinates in the common domain.

Overall optimization problem: We now combine the objectives in (2)-(4), which results in the overall optimization problem

$$\min_{Z, W} \|ZE\|_F^2 + \mu_1 \|W^T Z_l^s - Y^s\|_F^2 + \mu_2 \|W^T Z_l^t - Y^t\|_F^2 + \mu_3 \|W\|_F^2 + \mu_4 \text{tr}(Z^s L^s (Z^s)^T) + \mu_5 \text{tr}(Z^t L^t (Z^t)^T). \quad (5)$$

Here $\mu_1, \mu_2, \mu_3, \mu_4$, and μ_5 are positive weight parameters that balance the contributions of different terms. The overall problem in (5) aims to compute the embedding coordinates Z of all data samples jointly with the parameters W of the linear classifier.

B. Solution of the Optimization Problem

The optimization problem in (5) is not jointly convex in the coordinates Z and the classifier parameters W ; therefore, it is hard to solve. Nevertheless, the objective function is quadratic and convex when regarded as a function of only Z or only W . In our method, we minimize the objective function with an alternating optimization approach. In each iteration of the method, we first fix Z and optimize W ; and then fix W and optimize Z . The iterative optimization procedure is explained below.

Initialization of the algorithm: In the beginning of the algorithm, neither the embedding Z nor the classifier parameters W are known. Hence, we propose to initialize the method by optimizing the terms that contain only Z in the objective function (5). We thus get the problem

$$\min_Z f_1(Z), \text{ subject to } ZZ^T = I_d \quad (6)$$

where $I_d \in \mathbb{R}^{d \times d}$ denotes the identity matrix and

$$f_1(Z) = \|ZE\|_F^2 + \mu_4 \text{tr}(Z^s L^s (Z^s)^T) + \mu_5 \text{tr}(Z^t L^t (Z^t)^T).$$

The optimization constraint in (6) has the purpose of normalizing the coordinates in order to avoid the trivial solution $Z = 0$. We can rewrite $f_1(Z)$ as

$$f_1(Z) = \text{tr}(ZAZ^T)$$

where $A = EE^T + L$ and

$$L = \begin{bmatrix} \mu_4 L^s & 0 \\ 0 & \mu_5 L^t \end{bmatrix}.$$

The solution of (6) is then given by the matrix Z whose k -th row consists of the eigenvector of A corresponding to its k -th smallest eigenvalue.

Optimization of W in the main loop: In each iteration of the algorithm, we first fix Z and optimize W . The problem in (5) then becomes

$$\min_W f_2(W) \quad (7)$$

where

$$f_2(W) = \mu_1 \|W^T Z_l^s - Y^s\|_F^2 + \mu_2 \|W^T Z_l^t - Y^t\|_F^2 + \mu_3 \|W\|_F^2.$$

We evaluate and set the gradient of $f_2(W)$ to 0 as

$$\frac{\partial f_2(W)}{\partial W} = 2BW - 2C = 0$$

where

$$B = \mu_1 Z_l^s (Z_l^s)^T + \mu_2 Z_l^t (Z_l^t)^T + \mu_3 I_d$$

$$C = \mu_1 Z_l^s (Y^s)^T + \mu_2 Z_l^t (Y^t)^T.$$

The solution of (7) is thus given by

$$W = B^{-1}C.$$

Optimization of Z in the main loop: In the next stage of an iteration, we fix W and optimize Z . The optimization problem (5) becomes in this case

$$\min_Z f_3(Z) \quad (8)$$

where

$$f_3(Z) = \|ZE\|_F^2 + \mu_1 \|W^T Z_l^s - Y^s\|_F^2 + \mu_2 \|W^T Z_l^t - Y^t\|_F^2$$

$$+ \mu_4 \text{tr}(Z^s L^s (Z^s)^T) + \mu_5 \text{tr}(Z^t L^t (Z^t)^T). \quad (9)$$

Computing the gradient of $f_3(Z)$ and equating to 0, we get

$$\frac{\partial f_3(Z)}{\partial Z} = 2ZA + WW^T ZS - C = 0 \quad (10)$$

where S is the block-diagonal matrix of the form

$$S = \begin{bmatrix} 2\mu_1 I_{M_s} & & & \\ & 0_{P_s \times P_s} & & \\ & & 2\mu_2 I_{M_t} & \\ & & & 0_{P_t \times P_t} \end{bmatrix}$$

with $0_{P \times Q} \in \mathbb{R}^{P \times Q}$ denoting a zero matrix, $P_s = N_s - M_s$, and $P_t = N_t - M_t$. The matrix C is given by

$$C = \begin{bmatrix} 2\mu_1 WY^s & 0_{d \times P_s} & 2\mu_2 WY^t & 0_{d \times P_t} \end{bmatrix}.$$

The equation (10) is linear in the entries of Z ; therefore, it can be rewritten as a linear equation system where the unknown is the vectorized form of Z . We compute the optimal solution of (8) by vectorizing Z , rearranging the matrices, and solving the resulting linear equation system.

Termination of the algorithm: In each iteration, the updates on both Z and W decrease the value of the objective function. As the objective function is decreasing and nonnegative, it is guaranteed to converge throughout the iterations. We terminate the algorithm when the decrease in the objective function between two consecutive iterations falls below a predetermined threshold.

Finally, the estimation $\hat{C}(x)$ of the class label of an unlabeled sample x (in either domain) is obtained with the linear classifier learnt in the algorithm. The estimated class label is

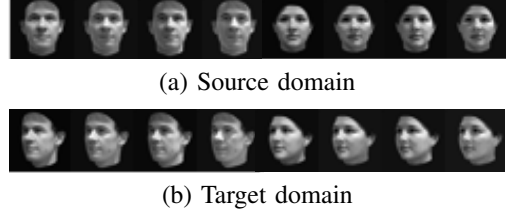


Figure 2: Example source and target images for two subjects from the MIT-CBCL face data set

simply given by the index of the largest entry of the classifier output

$$\hat{C}(x) = \arg \max_c y_c,$$

where $y = W^T z$ is the estimated label vector, y_c is the c -th entry of y , and z is the embedding of the data sample x .

III. EXPERIMENTAL RESULTS

The proposed method is tested in a face recognition and an object recognition application. We compare our method to the domain adaptation methods Geodesic Flow Kernel (GFK) [6], Subspace Alignment (SA) [7], and Domain Adaptation using Manifold Alignment (DAMA) [15]; as well as the basic Nearest Neighbor (NN) classification method.

The face recognition experiment is done on the MIT-CBCL data set [16], which consists of the images of 10 different subjects captured under 36 different illumination conditions and 9 pose angles. In the experiments, we consider each pose angle as a different domain, and test the algorithms by taking the images captured under the frontal view as the source domain samples, and the images under the profile view as the target domain samples. Example source and target images of two different subjects are given in Figure 2. All source images are labeled, and the ratio of the labeled target images is increased progressively in the experiment. The classification errors of the compared algorithms over the unlabeled target images is reported in percentage in Table I for different percentages of labeled target samples. The results show that the proposed method significantly outperforms the other algorithms in comparison, yielding a misclassification rate of 0% in all cases. The GFK and SA methods perform better than the other methods in comparison. This is due to the nature of the data set, consisting of the images of the same subjects captured from different angles, which can be conveniently aligned via linear projections as in these two methods. Meanwhile, the proposed nonlinear projection method outperforms these two linear methods, which suggests that the flexibility of nonlinear representations improves very much the classification performance.

Next, we compare the methods in an object recognition experiment done on the COIL-20 data set [17]. The data set consists of a total of 1440 images of 20 different objects captured from different camera angles. In order to increase the challenge of the problem, we consider a transfer learning

Table I: Misclassification rates (%) of the compared methods in the target domain for the MIT-CBCL face data set

Label ratio	2.78%	5.56%	8.33%	16.67%	25	%
Proposed	0	0	0	0	0	
GFK	9.69	5.71	2.85	0.47	0.19	
SA	12.2	7.88	3.79	1.17	0.07	
DAMA	78.69	73.21	73.00	54.97	44.00	
NN	15.43	8.15	3.36	0.73	0.04	



Figure 3: Example images from the COIL-20 data set. Each source object in the top row is matched to the target object right below it.

scenario in this experiment: The 20 different objects in the data set are divided into two groups such that the first and the second groups constitute the source and the target domains. The groups are formed such that each object in the source group is matched to the most similar object in the target group. The matching object pairs are then treated as if they have the same class label. Sample images from each object pair can be seen in Figure 3. The misclassification rates of the compared methods over the unlabeled target samples are presented in Table II. It is observed that the smallest misclassification rate is given by the proposed method. The source and the target samples from the same class are actually images of different objects in this experiment. This degrades the performance of the GFK and the SA methods, which are based on aligning the two domains with linear projections. On the other hand, the proposed approach of learning nonlinear projections seems to provide an effective solution in this challenging scenario.

IV. CONCLUSION

We have proposed a semi-supervised domain adaptation method that is based on learning nonlinear embeddings of the source and the target data into a common domain. The embedding is computed so that the same-class data samples from different domains are mapped to nearby coordinates, while the learnt coordinates respect the geometric structure of the data and also successfully allow the computation of a linear classifier. The proposed method is tested on face and object image data sets. Comparisons to reference domain adaptation methods show that the proposed nonlinear embedding approach is quite efficient for learning domain-adaptive representations for pattern recognition applications.

REFERENCES

- [1] J. Huang, A. J. Smola, A. Gretton, K. M. Borgwardt, and B. Schölkopf. Correcting sample selection bias by unlabeled data. In *Proc. Advances in Neural Information Processing Systems 19*, pages 601–608, 2006.
- [2] Q. Sun, R. Chattopadhyay, S. Panchanathan, and J. Ye. A two-stage weighting framework for multi-source domain adaptation. In *Proc. Advances Neural Inf. Proc. Sys.*, pages 505–513, 2011.

Table II: Misclassification rates (%) of the compared methods in the target domain for the COIL-20 object data set.

Label ratio	4.17%	6.94%	9.72%	13.89%	18.06%
Proposed	4.01	3.99	3.32	2.84	2.41
GFK	28.25	18.54	15.95	11.71	8.86
SA	30.19	21.64	18.35	13.34	9.53
DAMA	40.45	25.87	19.57	15.82	12.83
NN	28.74	18.79	15.95	12.29	9.03

- [3] H. Daumé III, A. Kumar, and A. Saha. Co-regularization based semi-supervised domain adaptation. In *Proc. Advances in Neural Information Processing Systems 23*, pages 478–486, 2010.
- [4] L. Duan, D. Xu, and I. W. Tsang. Learning with augmented features for heterogeneous domain adaptation. In *Proc. 29th Int. Conf. Machine Learning*, 2012.
- [5] M. Gong, K. Zhang, T. Liu, D. Tao, C. Glymour, and B. Schölkopf. Domain adaptation with conditional transferable components. In *Proc. 33rd Int. Conf. Machine Learning*, pages 2839–2848, 2016.
- [6] B. Gong, Y. Shi, F. Sha, and K. Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *IEEE Conf. Comp. Vis. Pat. Rec.*, 2012.
- [7] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. In *Proc. IEEE Int. Conf. Comp. Vis.*, 2013.
- [8] L. Cheng and S. J. Pan. Semi-supervised domain adaptation on manifolds. *IEEE Trans. Neural Netw. Learning Syst.*, 25(12):2240–2249, 2014.
- [9] T. Yao, Y. Pan, C. Ngo, H. Li, and T. Mei. Semi-supervised domain adaptation with subspace learning for visual recognition. In *IEEE Conf. Comp. Vis. Pat. Rec.*, pages 2142–2150, 2015.
- [10] K. Muandet, D. Balduzzi, and B. Schölkopf. Domain generalization via invariant feature representation. In *Proc. 30th International Conference on Machine Learning*, pages 10–18, 2013.
- [11] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(9):1853–1865, 2017.
- [12] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.
- [13] M. Sugiyama. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of Machine Learning Research*, 8:1027–1061, 2007.
- [14] C. Ornek and E. Vural. Nonlinear supervised dimensionality reduction via smooth regular embeddings. *Pattern Recognition*, 87:55–66, 2019.
- [15] C. Wang and S. Mahadevan. Heterogeneous domain adaptation using manifold alignment. In *Proc. Int. Joint. Conf. Art. Intel.*, 2011.
- [16] MIT-CBCL face recognition database. Available: <http://cbcl.mit.edu/software-datasets/heisele/facerecognition-database.html>.
- [17] S. A. Nene, S. K. Nayar, and H. Murase. Columbia Object Image Library (COIL-20). Technical report, Feb 1996.

Tree-Seed Programming for Symbolic Regression

Parvana Yunusova¹ and Mustafa Servet Kiran¹

¹ Konya Technical University, Konya/Turkey, parvanayunusova@gmail.com

¹ Konya Technical University, Konya/Turkey, mskiran@ktun.edu.tr

Abstract - Tree-seed algorithm (TSA), proposed for solving continuous optimization problems by inspiring relationship between trees and their seeds, is a metaheuristic optimization algorithm. Being used the interaction between trees and seeds, real-valued optimization problems are solved. In recent research, some discrete versions of TSA have also been developed. In this study, a novel version of tree-seed algorithm for programming is introduced as a product of intersection with swarm intelligence and automatic programming. Automatic programming solves problems by automatically generated computer programs without a need of knowing a form of a solution in advance. One of the successful tools of automatic programming is genetic programming. To obtain the proposed approach, the solution generation mechanism of genetic programming is embedded into TSA. To evaluate the performance of the proposed method on symbolic regression, 10 benchmark problems which consist of polynomial, trigonometric, logarithmic, square-root and bivariate functions have been considered. Moreover, a parameter analysis for the proposed method has been conducted and the obtained results show that the proposed algorithm produces acceptable solutions.

Keywords - tree-seed algorithm, genetic programming, automatic programming, swarm intelligence

I. INTRODUCTION

There is a growing interest in the intersection between swarm intelligence (SI) and automatic programming in which automatic programming algorithms use SI metaheuristic algorithms as the search technique [1]. The Automatic programming solves problems by automatically generated computer programs without a need of knowing a form of a solution in advance. One of the successful tools of automatic programming is genetic programming (GP) proposed by Koza [2]. GP added as a new technique and extension of the genetic algorithm to evolutionary computing. GP was developed by changing the representation of the individuals in GA which was the main issue to solve more complex problems. Hence, they were changed from fixed-length binary strings to compositions of primitive functions and terminals with different structures such as trees, linear sequences of instructions, graphs etc. [3]. SI is an artificial intelligence technique that aims to design intelligent multi-agent systems based on collective behaviors of ants, bees, birds, fish etc. [4]. SI algorithms have been applied successfully in many areas such as optimization, data mining, clustering, pattern recognition, machine learning etc. [5].

However, there are a few research using SI algorithms in automatic programming, especially using GP with SI algorithms and testing on symbolic regression. The focus of this study is in this direction. Although GP is a powerful technique in automatic programming, it is known as computationally expensive based on the number of individuals in the population [6]. One of the approach to overcome this drawback is the intersection of SI algorithms with GP. In this paradigm, SI algorithms such as ant colony optimization (ACO) [5], artificial bee colony (ABC) [7], particle swarm optimization (PSO) [8] etc. are used as a search method. For example, [9] artificial bee colony programming based on ABC, [10] dynamic ant programming based on ACO, [11] hybrid genetic programming with particle swarm optimization is introduced for the symbolic regression problem, [12] gravitational search programming based on gravitational search algorithm has been employed to symbolic regression and feature construction.

One of the recently proposed swarm intelligence algorithms is tree-seed algorithm [13] (TSA) for short, developed for solving continuous optimization problems by inspiring relation between trees and their seeds. According to this structure, the tree produces seeds, and the best seed location is replaced with tree location. Later, TSA was introduced for solving constraint optimization [13, 14], binary optimization [15, 16] and other research areas resulting in significant works such as [17-20]. In this paper, a new version of TSA named tree-seed programming, TSp for short, is developed for automatic programming. The algorithmic structure of TSp comes from TSA and the solution generation mechanism comes from GP. The proposed method and its performance has been investigated on symbolic regression.

The rest of the paper is as follows: Section 2 shows the brief introduction of TSA, TSp is introduced in Section 3, experiments and results are presented in Section 4 and conclusion is given in Section 5.

II. TREE-SEED ALGORITHM

Tree-seed algorithm, which is a novel population-based heuristic search algorithm inspired by the relationship between trees and their seeds and firstly introduced in 2015 [21]. Population in TSA is called stand and at the beginning of the search, stand consists of tree locations obtained by using Eq.(1)

$$T_{i,j} = L_{j,\min} + r_{i,j}(H_{j,\max} - L_{j,\min}) \quad (1)$$

where, $T_{i,j}$ represents the j th dimension of the i th tree, $H_{j,\max}$ and $L_{j,\min}$ is higher and lower bound of search space, respectively. $r_{i,j}$ is a random number generated in the range of $[0,1]$. The algorithm tries to solve the problem by producing seeds for each tree and comparing tree with its best seeds. If the best seed is better than the current tree, the best tree is added to the stand and the current is removed from the stand. Here the locations of trees and seeds are considered a solution to the problem. At each iteration, the number of seeds generated from each tree can be different and it is determined by selecting randomly in the range of $[10\% \text{ of stand size}, 25\% \text{ of stand size}]$. There are two equations (Eq.2 and Eq.3) proposed for production of seed and selection of equation depends on a control parameter named search tendency (ST) in the range of $[0,1]$. If a random number produced from $[0,1]$ is smaller than ST , then Eq.2, otherwise Eq.3 is used.

$$S_{k,j} = T_{i,j} + \alpha_{i,j}(B_j - T_{r,j}) \quad (2)$$

$$S_{k,j} = T_{i,j} + \alpha_{i,j}(T_{i,j} - T_{r,j}) \quad (3)$$

Here, $S_{k,j}$ is the j th dimension of the i th tree's k th seed, B_j indicates the j th dimension of best tree location obtained so far, $T_{r,j}$ is the j th dimension of a randomly selected tree from the stand, $\alpha_{i,j}$ is a random number in the range of $[-1,1]$, which is a scaling factor. Until the termination condition that is the maximum number of function evaluations (Max_Fes) is met, the generation of seeds continues [21].

III. TREE-SEED PROGRAMMING

TSp as an extended version of TSA is an automatic programming approach, which it uses tree-based GP for solution representation. In tree-based GP, individuals in a population which is represented by trees are computer programs consisting of functions and terminals appropriate to the problem. Functions can be arithmetic, mathematical, boolean etc. functions and terminals are variables, constant values or 0-arity functions. For instance, Figure 1 shows the tree representation of $xy + \sin y$.

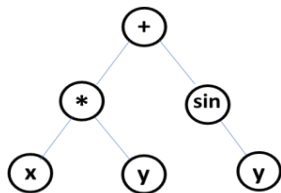


Figure 1: Tree representation of $xy + \sin y$.

GP changes population iteratively into other population using genetic operations such as crossover, mutation [22].

For developing TSp, the linear encoding and evolving rules are changed based on the tree encoding. Therefore initial stand in TSp is created by one of the methods using in GP (grow, full, ramped half-and-half) and because of tree encoding basic solution generation mechanism used for seed production in TSA cannot be directly applied for TSp. Therefore, seed production for each tree in Eq.2 can be thought as an interaction of each tree producing seed with an interaction of the best and randomly chosen tree, in Eq.3 as an interaction of each tree producing seed with an interaction of this tree and a random tree. This approach can be handled in tree encoding via the crossover operator. Using crossover operation (generates one offspring) in GP, Eq.2 and Eq.3 are considered as follows in TSp:

$$S_{k,j} = \text{crossover}(T_{i,j}, \text{crossover}(B_j, T_{r,j})) \quad (4)$$

$$S_{k,j} = \text{crossover}(T_{i,j}, \text{crossover}(T_{i,j}, T_{r,j})) \quad (5)$$

To measure the performance of each computer program (tree, seed location) the sum of absolute errors is selected.

$$SAE_i = \sum_{j=1}^N \|(g_j - t_j)\| \quad (6)$$

Where SAE_i is an objective function of TSp, N shows the number of samples, g_j and t_j is actual and target output of j th case, respectively.

IV. EXPERIMENTS AND RESULTS

The effectiveness and performance of the proposed method have been investigated on symbolic regression problems which are widely used for testing automatic programming approach. In this study, 10 benchmark functions given in Table 1 are used and they are taken from the works [23] [9, 12].

Table 1: Symbolic regression functions

Functions	Description
$f_1 = x^3 + x^2 + x$	20 random points $\subseteq [-1,1]$
$f_2 = x^4 + x^3 + x^2 + x$	20 random points $\subseteq [-1,1]$
$f_3 = x^5 + x^4 + x^3 + x^2 + x$	20 random points $\subseteq [-1,1]$
$f_4 = x^6 + x^5 + x^4 + x^3 + x^2 + x$	20 random points $\subseteq [-1,1]$
$f_5 = \sin(x^2) \cos(x) - 1$	20 random points $\subseteq [-1,1]$
$f_6 = \sin(x) + \sin(x + x^2)$	20 random points $\subseteq [-1,1]$
$f_7 = \log(x + 1) + \log(x^2 + 1)$	20 random points $\subseteq [0,2]$
$f_8 = \sqrt{x}$	20 random points $\subseteq [0,4]$
$f_9 = \sin(x) + \sin(y^2)$	100 random points $\subseteq [-1,1] \times [-1,1]$
$f_{10} = 2 \sin(x) \cos(y)$	100 random points $\subseteq [-1,1] \times [-1,1]$

Symbolic regression is a process of finding a function that best matches a given dataset. Therefore, the relationship between input and output is not presented on predefined function form as in linear or nonlinear regression methods [9, 24]. The location of trees and seeds are computer programs

consisting of function and terminal set. These sets chosen appropriately to the ten selected test problems are as follows: Function set={+, -, ×, ÷, *sin*, *cos*, *exp*, *rlog*}, for functions with one arity (f1, f2, f3, f4, f5, f6, f7, f8): Terminal set={ x }, for functions with two arities (f9, f10): Terminal set={ x, y}.

To evaluate the effect of stand size and Max_Fes on the TSp performance their different values were set according to each ST parameters (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9). Thus

each function ran 10 times for every ST and stand (50, 100, 250, 500) according to 3 different Max_Fes values (10000, 50000, 100000). Therefore, we have totally 108 different tests and each of them is run 10 times with random seed. In the experiments, while Max_Fes is 100000, the better values are obtained, and due to page limitation, the analysis of Max_Fes is not given in the present study. The rest of the experimental results are given in Table 2 – 4.

Table 2: The experimental results (ST=0.1, 0.2, 0.3, Stand Size= 50, 100, 250, 500).

ST=0.1 FES=100000					ST=0.2 FES=100000					ST=0.3 FES=100000				
pop	50	100	250	500	pop	50	100	250	500	pop	50	100	250	500
f1	7	10	10	10	f1	7	9	10	10	f1	4	7	10	10
f2	10	10	10	5	f2	10	10	9	5	f2	9	8	10	2
f3	10	10	10	5	f3	9	10	10	4	f3	9	10	8	4
f4	7	9	9	0	f4	7	6	6	0	f4	6	5	6	1
f5	0	2	0	0	f5	2	3	0	0	f5	1	1	0	0
f6	7	9	9	3	f6	8	9	10	4	f6	7	9	10	5
f7	0	1	1	0	f7	1	0	0	1	f7	0	0	0	0
f8	0	0	0	0	f8	0	0	0	0	f8	0	0	0	0
f9	8	9	10	3	f9	5	9	10	5	f9	5	8	7	3
f10	1	2	2	2	f10	0	1	4	0	f10	0	0	3	1
	50	62	61	28		49	57	59	29		41	48	54	26

Table 3: The experimental results (ST=0.4, 0.5, 0.6, Stand Size= 50, 100, 250, 500).

ST=0.4 FES=100000					ST=0.5 FES=100000					ST=0.6 FES=100000				
pop	50	100	250	500	pop	50	100	250	500	pop	50	100	250	500
f1	8	8	10	10	f1	7	7	9	10	f1	5	7	10	9
f2	10	8	10	7	f2	7	10	9	6	f2	8	5	10	3
f3	10	8	8	5	f3	9	8	9	4	f3	8	9	5	6
f4	8	5	5	0	f4	3	7	2	0	f4	1	2	2	0
f5	0	2	0	0	f5	0	1	0	0	f5	0	3	0	0
f6	7	7	9	8	f6	9	6	8	6	f6	7	5	8	7
f7	3	0	1	0	f7	0	1	0	0	f7	1	1	0	0
f8	0	0	0	0	f8	0	0	0	0	f8	0	0	0	0
f9	7	6	10	10	f9	4	10	8	6	f9	3	8	10	9
f10	0	2	2	2	f10	1	2	1	2	f10	0	1	4	1
	53	51	55	42		40	52	46	34		33	41	49	35

Table 4: The experimental results (ST=0.7, 0.8, 0.9, Stand Size= 50, 100, 250, 500)

ST=0.7 FES=100000					ST=0.8 FES=100000					ST=0.9 FES=100000				
pop	50	100	250	500	Pop	50	100	250	500	pop	50	100	250	500
f1	8	7	9	9	f1	5	6	10	9	f1	4	9	10	9
f2	7	8	7	4	f2	7	7	8	4	f2	2	3	7	3
f3	7	7	5	2	f3	5	5	6	4	f3	4	4	5	4
f4	2	2	2	0	f4	2	4	2	0	f4	5	2	1	2
f5	1	2	2	0	f5	2	1	0	0	f5	0	4	1	0
f6	6	6	9	10	f6	5	7	10	5	f6	3	3	8	10
f7	0	0	0	0	f7	0	2	0	0	f7	0	0	0	0
f8	0	0	0	0	f8	0	0	0	0	f8	0	0	0	0
f9	5	8	10	6	f9	2	7	10	7	f9	5	4	10	9
f10	0	2	2	2	f10	0	1	2	0	f10	0	0	2	4
	36	42	46	33		28	40	48	29		23	29	44	41

In Table 2-4, the number of hits is reported. As seen from these results tables, when we select the ST parameter as 0.1, the better results are achieved. In Table 2, it is expected that the higher population size is better than the others, unlike this

expectation, the highest stand size in TSp size shows the worst performance in Table 2. The appropriate value for the stand size is between 100 and 250 while when the stand size is 100, the best results in the experiments are obtained. For the functions,

the optimum results are obtained at each run for f1, f2, f3 and f9 functions in Table 2 while the same parameter set is used. Only, the proposed TSp cannot obtain the optimum result for f8 function under our experimental conditions. The other results can be seen in the tables.

V. CONCLUSION

In this paper, a new version of tree-seed algorithm named tree-seed programming is introduced. In the present approach, tree-seed algorithm is used for searching the solution space, and the solution generation mechanism of GP is integrated into TSA, because the solution update rule of basic TSA works on continuous solution space. To evaluate the performance of the proposed algorithm, a symbolic regression benchmark problem set are considered. The problems are numeric functions including polynomial, trigonometric, logarithmic, square-root and bivariate functions. In the experiments, tree-seed programming achieved good results for certain functions. However, the proposed approach can be improved in several ways such as using the different types of crossover, and more investigations can be done on the creation of tree-seed algorithm for automatic programming. Moreover, tree-seed programming use tree-based solution representation in this study, for future works, different solution representation approaches can be used for the proposed algorithm.

REFERENCES

- [1] J. L. Olmo, J. R. Romero, and S. Ventura, "Swarm-based metaheuristics in automatic programming: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Article vol. 4, no. 6, pp. 445-469, 2014.
- [2] R. Poli, W. B. Langdon, N. F. McPhee, and J. R. Koza, *A field guide to genetic programming*. Lulu. com, 2008.
- [3] J. R. Koza, "Genetic programming as a means for programming computers by natural selection," *Statistics and computing*, vol. 4, no. 2, pp. 87-112, 1994.
- [4] C. Blum and X. Li, "Swarm intelligence in optimization," in *Swarm Intelligence*: Springer, 2008, pp. 43-85.
- [5] M. Dorigo and G. Di Caro, "Ant colony optimization: a new meta-heuristic," in *Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406)*, 1999, vol. 2, pp. 1470-1477: IEEE.
- [6] W. B. Langdon and R. Poli, "Genetic programming bloat with dynamic fitness," in *Genetic Programming*: Springer, 1998, pp. 97-112.
- [7] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," *Journal of global optimization*, vol. 39, no. 3, pp. 459-471, 2007.
- [8] J. Kennedy and R. Eberhart, "Particle swarm optimization, proceedings of IEEE International Conference on neural networks (ICNN'95) in," ed, 1995.
- [9] D. Karaboga, C. Ozturk, N. Karaboga, and B. Gorkemli, "Artificial bee colony programming for symbolic regression," *Information Sciences*, vol. 209, pp. 1-15, 2012.
- [10] S. Shirakawa, S. Ogino, and T. Nagao, "Dynamic ant programming for automatic construction of programs," *IEEJ Transactions on Electrical and Electronic Engineering*, Article vol. 3, no. 5, pp. 540-548, 2008.
- [11] F. Qi, Y. Ma, X. Liu, and G. Ji, "A hybrid genetic programming with particle swarm optimization," in *International Conference in Swarm Intelligence*, 2013, pp. 11-18: Springer.
- [12] A. Mahanipour and H. Nezamabadi-pour, "GSP: an automatic programming technique with gravitational search algorithm," *Applied Intelligence*, pp. 1-15, 2018.
- [13] M. S. Kiran, "An Implementation of Tree-Seed Algorithm (TSA) for Constrained Optimization," in *Intelligent and Evolutionary Systems*: Springer, 2016, pp. 189-197.
- [14] A. Babalik, A. C. Cinar, and M. S. Kiran, "A modification of tree-seed algorithm using Deb's rules for constrained optimization," *Applied Soft Computing*, vol. 63, pp. 289-305, 2018.
- [15] A. C. Cinar, H. Iscan, and M. S. Kiran, "Tree-Seed algorithm for large-scale binary optimization," *KnE Social Sciences*, vol. 3, no. 1, pp. 48-64, 2018.
- [16] A. C. Cinar and M. S. Kiran, "Similarity and logic gate-based tree-seed algorithms for binary optimization," *Computers & Industrial Engineering*, vol. 115, pp. 631-646, 2018.
- [17] M. S. Kiran, "Withering process for tree-seed algorithm," *Procedia computer science*, vol. 111, pp. 46-51, 2017.
- [18] A. Cinar and M. Kiran, "A Parallel Version of Tree-Seed Algorithm (TSA) within CUDA Platform," in *Selçuk International Scientific Conference On Applied Sciences*, 2016.
- [19] V. Muneeswaran and M. P. Rajasekaran, "Performance evaluation of radial basis function networks based on tree seed algorithm," in *Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on*, 2016, pp. 1-4: IEEE.
- [20] V. Muneeswaran and M. P. Rajasekaran, "Beltrami-regularized denoising filter based on tree seed optimization algorithm: an ultrasound image application," in *International Conference on Information and Communication Technology for Intelligent Systems*, 2017, pp. 449-457: Springer.
- [21] M. S. Kiran, "TSA: Tree-seed algorithm for continuous optimization," *Expert Systems with Applications*, vol. 42, no. 19, pp. 6686-6698, 2015.
- [22] W. B. Langdon, R. Poli, N. F. McPhee, and J. R. Koza, "Genetic programming: An introduction and tutorial, with a survey of techniques and applications," in *Computational intelligence: A compendium*: Springer, 2008, pp. 927-1028.
- [23] N. Q. Uy, N. X. Hoai, M. O'Neill, R. I. McKay, and E. Galván-López, "Semantically-based crossover in genetic programming: application to real-valued symbolic regression," *Genetic Programming and Evolvable Machines*, vol. 12, no. 2, pp. 91-119, 2011.
- [24] C. Zhang, "Genetic programming for symbolic regression," *University of Tennessee, Knoxville, TN*, vol. 37996.

An Automated Deep Learning Approach for Bacterial Image Classification

M.TALO¹

¹ Munzur University, Tunceli/Turkey, muhammedtalo@munzur.edu.tr

Abstract – Automated recognition and classification of bacteria species from microscopic images have significant importance in clinical microbiology. Bacteria classification is usually carried out manually by biologists using different shapes and morphologic characteristics of bacteria species. The manual taxonomy of bacteria types from microscopy images is time-consuming and a challenging task for even experienced biologists. In this study, an automated deep learning based classification approach has been proposed to classify bacterial images into different categories. The ResNet-50 pre-trained CNN architecture has been used to classify digital bacteria images into 33 categories. The transfer learning technique was employed to accelerate the training process of the network and improve the classification performance of the network. The proposed method achieved an average classification accuracy of 99.2%. The experimental results demonstrate that the proposed technique surpasses state-of-the-art methods in the literature and can be used for any type of bacteria classification tasks.

Keywords – Bacteria classification, convolutional neural networks, deep learning, transfer learning, CNN.

I. INTRODUCTION

There are numerous bacteria species in our body and around us. Some bacteria species are beneficial for human life while others are harmful. Beneficial bacteria species help digestion of foods, fermentation of dairy products, drug production, etc. The harmful species of bacteria are the main cause of various diseases. Therefore, the classification of bacterial species is very important especially for human life.

Biologists try to identify and classify various bacteria types, which have different biochemistry and shapes. They use different attributes of bacteria for classification. For instance, the shape of the bacterial cells (spiral, cylindrical and spherical), the size and structure of the colonies formed by the bacteria are examined to differentiate bacteria species [1]. The cells of some bacteria types have different size and structure depending on environmental conditions. Some species of bacteria are very similar in shape. Although each bacteria species has its own characteristics, the biochemical reactions that bacteria carry out and the metabolic activities they perform together help to classify species [2]. However, the classification of bacteria species is not an easy task even by experienced specialists.

The classification of bacterial species with the help of computer-aided systems would provide great convenience for biologists. Accurate and rapid classification of bacterial species

is of great importance, particularly early detection and treatment of diseases caused by bacteria. Different image processing approaches, such as traditional machine learning methods and modern deep learning techniques, have been used in various studies for the recognition and classification of bacterial types. For the automated detection of bacteria types, Ahmed et al. [3], employed the Fisher discriminant analysis method to extract features and a Support Vector Machine (SVM) was used for classification. Bruyne et al. [4] utilized Random Forest and SVM machine learning methods to identify three different types of bacteria, *Fructobacillus*, *Leuconostoc*, and *Lactococcus*. The authors reported different recognition accuracies, which vary between 94% and 98%. Goodacre et al. [5], used different spectrometry and artificial neural networks (ANN) approaches to identify bacteria types. They have reached the highest identification accuracy of 80%. Nie et al. [6] employed convolutional deep belief networks (CDBN) to learn high-level representations from digital bacterial images. Then they trained a convolutional neural network (CNN) to classify and segment bacterial images. Fiannaca et al. [7], proposed a classification approach based on k-mer representation and used deep learning technique to classify bacteria sequences.

The classical machine learning techniques are mainly used for the purposed of feature extraction. Then, the learned representations are passed through a classifier, such as an SVM or random forest, for classification tasks. However, in deep learning, the representations that obtained from raw images and the classification process take place on a single structure.

In recent years, the deep learning algorithms, especially convolutional neural networks are commonly used for various image processing and computer vision tasks such as identification, classification, and segmentation [8]. CNN architectures have many successful applications in the fields of health, safety, language translation, pathology, microbiology, etc. They are usually trained on a large amount of labeled data (supervised learning) using parallel computation methods with the help of GPUs.

In this study, deep learning based pre-trained ResNet-50 architecture was used to classify bacterial species from digital images. The transfer learning technique was used for the accurate and robust training process. Transfer learning method enabled to train the model with fewer data samples. The proposed method has an end-to-end structure and classifies bacterial images without requiring any pre-processing step.

II. MATERIAL AND METHODS

A. Bacteria Species Dataset

In this study, DIBaS bacteria species dataset [9], was used to classify digital bacteria images. The dataset is publicly available and composed of 33 bacterial species. Each bacteria type has about 20 images. All the bacteria images have the resolution of 2048×1532 pixels. Several images from DIBaS dataset is shown in Figure 1. The different bacteria species from the dataset and the number of them are given in Table 1.

The samples from the dataset were stained with Gramm method. All bacteria images were acquired from the Olympus CX31 microscope.

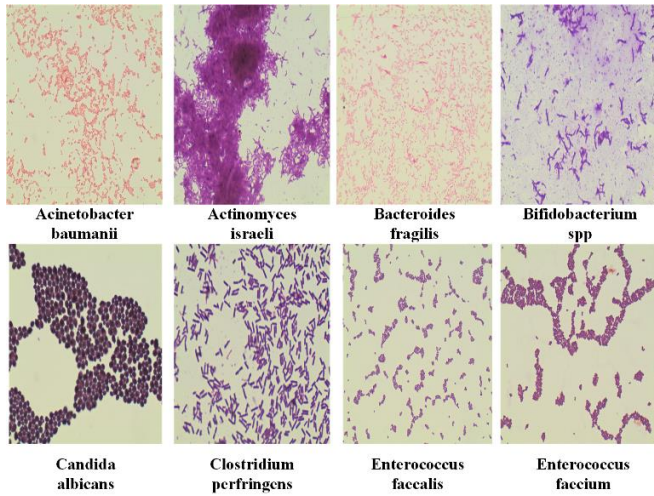


Figure 1: Sample bacteria images from DIBaS dataset

Table 1: The name of bacteria species and the number of samples from DIBaS dataset.

Species	Number
Lactobacillus johnsonii	20
Listeria monocytogenes	22
Propionibacterium acnes	23
Veionella	22
Staphylococcus aureus	20
Enterococcus faecium	20
Lactobacillus gasseri	20
Streptococcus agalactiae	20
Actinomyces Israeli	23
Fusobacterium	23
Pseudomonas aeruginosa	20
Lactobacillus plantarum	20
Lactobacillus reuteri	20
Clostridium perfringens	23
Neisseria gonorrhoeae	23
Proteus	20
Acinetobacter baumannii	20
Lactobacillus casei	20
Bacteroides fragilis	23
Porfyromonas gingivalis	23
Escherichia coli	20
Lactobacillus crispatus	20
Bifidobacterium spp	23
Staphylococcus epidermidis	20

Staphylococcus saprophiticus	20
Lactobacillus salivarius	20
Lactobacillus delbrueckii	20
Lactobacillus jehnsenii	20
Candida albicans	20
Lactobacillus rhamnosus	20
Micrococcus spp	21
Lactobacillus paracasei	20
Enterococcus faecalis	20

B. Deep Transfer Learning

Deep learning is a popular approach that commonly used in image processing tasks such as detection, classification, and segmentation. Deep learning algorithms, especially Convolutional Neural Networks (CNN), automatically learn high-level feature representations form the raw data without requiring any handmade feature extractions.

A CNN is constructed by stacking a block of convolutional, polling, and fully connected layers. There are also activation functions and some other layers such as normalization and dropout, which regularize the parameters of the network for efficient training which results in high performance. The convolutional layers learn representational information from the data by applying linear convolutional operations. The convolutional operations are performed by using different filters (kernels). A feature map is created by sliding the filter through the input data. The convolutional process is a mathematical linear operation obtained by an element-wise matrix multiplication between the input image and a filter. The sum of products is calculated and the results of convolutional operations are passed through the activation functions. An activation function enables a network to learn non-linear features. For example, the rectified linear unit (ReLU) activation function is commonly placed after the convolutional layer and describe as:

$$\text{ReLU}(x) = \max(0, x)$$

for a given input x . The Softmax activation function, which attached to the final layer of convolutional neural networks, outputs a probability distribution for each target class. A pooling layer reduces the size of the activation map spatially by extracting the important features from the image. The feature map size is reduced by pooling layer. Therefore, the pooling layer also decreases the computation cost of the network during training. For instance, the average pooling layer calculates the average of numbers in a feature map. The fully connected layers, which located at the end of the network, classify the images into various categories.

Training deep learning based CNN model from starch requires a large amount of labeled data and high computational power. The network weights (parameters) at the beginning of training are randomly initialized with some small values, which are usually between 0 and 1, as a starting point, then they are updated by the help of an optimization algorithm. This training process can take several weeks depending on the training data size and available computer hardware. The transfer learning technique [10] tackles these problems. Instead of training a model from scratch, a CNN model which was previously

trained for a different but related large-scale task can be used. It is not always possible to find numerous annotated data for training. The transfer learning technique enables to train networks which have a relatively small number of data.

In deep learning models, as the number of layers gets deeper, the accuracy of the model increases up to a point. Deep learning models learn the high-level features in deeper layers. However, an increase in the number of layers causes information loss after a certain point. This also complicates the training process of a network and the performance of the model may even degrade. ResNet architecture has overcome these problems. The ResNet architecture was released by He et al. [11], got the first place in ILSVRC and COCO challenges in 2015. ResNet introduced residual connections, also known as skip connection. The residual shortcuts add extra connections among residual blocks of a network for the information flow through the whole network. The skip connections were enabled to train deeper neural networks with even 1001-layers.

III. EXPERIMENTAL SETUP

In this study, ResNet-50 pre-trained CNN architecture is employed to classify bacteria species into 33 categories using the transfer learning technique. The convolutional and the pooling layers of ResNet-50 model are transferred to the new model. The fully connected layers of ResNet-50 is removed from the end of the network and replaced with a brand new fully connected layer that outputs a 33 unit tensors using the Softmax activation function to classify bacteria species. Additionally, two dropout layers have been added to the network to avoid overfitting during training. Once the overfitting occurs, the model does not generalize on unseen data, that is, it memorizes the training samples but fails on test data. The first and the second dropout rates are set to 25% and 50%, respectively.

The dataset consists of a total of 689 bacterial images. 80% of bacteria images (552) was allocated for the training set and the rest of the images, 20% (137) used for the validation set. The learning rate hyper-parameter is randomly set to $1e-3$. The learning rate hyper-parameter control the speed of the parameter update. If it is set to be too low, the model classification performance proceeds so slowly. On the contrary, if it is chosen to be too high, the model would have missed the local minimum and diverges. The RMSprop optimizer algorithm is used to update the parameters of the network during the training. The proposed ResNet-50 pre-trained model is trained and tested on Ubuntu 16.04 server using NVIDIA GeForce GTX 1080 TI graphic card. In the training and testing of the proposed ResNet-50 model, the Python programming language based PyTorch [12] libraries and Fastai [13] framework were used.

IV. RESULTS

5-fold cross-validation technique was used to evaluate the performance of proposed pre-trained ResNet-50 model. The experiments were repeated five times and the average of five trials on validation sets was given as a classification performance for the overall model. The ResNet-50 pre-trained model was trained for 50 epochs, that is, the model examined each training image 50 times. The 5-fold validation accuracy

and the training time for each fold are given in Table 2. A visualization of the validation accuracy graph for five folds and training and validation loss graph for Fold-5 are shown in Figure 2 and Figure 3, respectively.

Table 2: Training time and the classification accuracy for each fold.

Fold	Training Time (min : sec)	Validation accuracy (%)
Fold-1	31:55	99.27
Fold-2	32:17	99.27
Fold-3	31:30	98.54
Fold-4	31:42	99.27
Fold-5	31:35	99.27
Average	31:48	99.12

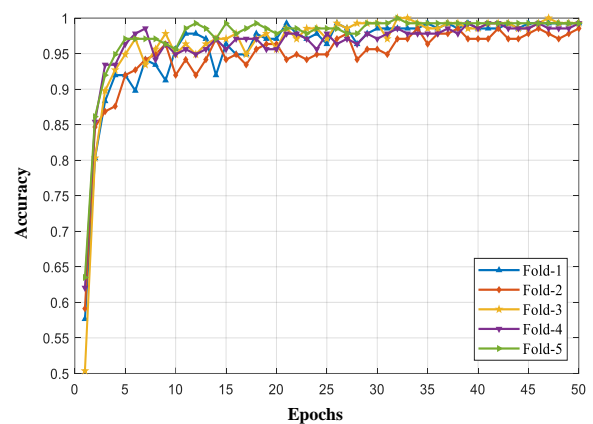


Figure 2: 5-Fold training accuracy of ResNet-50 model.

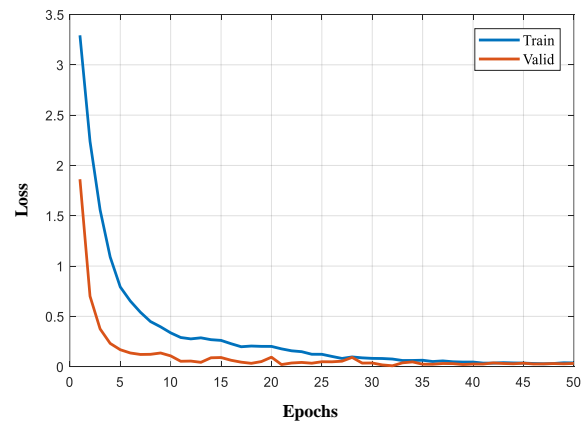


Figure 3: The training and validation loss graph for Fold-5.

As a result of all experiments, 99.12% average classification accuracy was obtained on the test sets. The average training time of ResNet-50 took about 31 minutes and 48 seconds. It can be seen from Figure-4 that, there is no overfitting during training, i.e., the training and validation losses were well-balanced and decreased during training. For the detailed performance analysis, the confusion matrix for Fold-5, which obtained using validation data, is given in Figure 4.

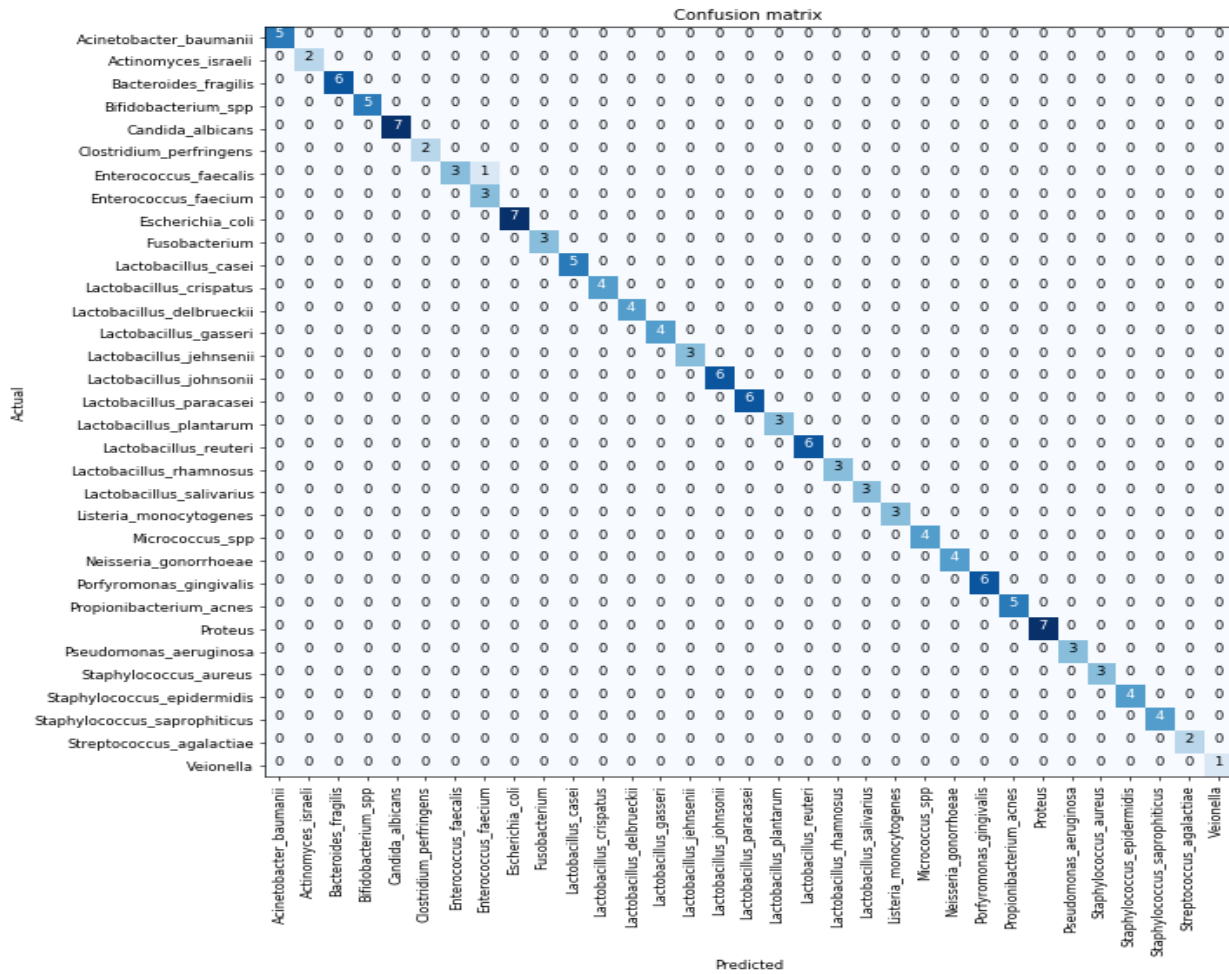


Figure 4: The obtained confusion matrix using the validation set for Fold-5.

The confusion matrix shows that, on the test set for Fold-5, all the bacteria images were classified correctly by the proposed ResNet-50 pre-trained model except one bacteria image, Enterococcus Faecalis. The model prediction for misclassified bacteria image was Enterococcus Faecium. Note that both bacteria types, Enterococcus Faecalis, and Enterococcus Faecium, comes from the enterococcus genre.

V.DISCUSSION

The classification performance of this study was compared with other state-of-the-art studies, which classified the bacterial images into different categories using DIBaS dataset. Table-3 shows the existing studies in the literature that conducted on the same bacterial image dataset.

In 2017, Zielinski et al. [9] applied convolutional neural networks to DIBaS bacterial image dataset. The authors applied fisher vector (FV), local image descriptors and the pooling encoder to acquire image descriptors. Then the Support Vector Machines (SVM) and Random Forest (RF) methods were used to classify species of bacteria into 33 classes. They have reported $97.24 \pm 1.07\%$ classification accuracy.

Mohamed et al. [14], proposed histogram equalization and Bag-of-words (BoW) methods to extract features from the DIBaS image dataset. They used Support Vector Machine to classify bacterial images into 10 categories. The authors provided an average classification accuracy of 97%.

In 2018, Nasip et al. [15], used deep learning based VggNet and AlexNet pre-trained CNN architectures to classify bacterial images into 33 categories. The authors obtained the highest classification accuracy of 98.25% via VggNet.

Table 3: The studies that are conducted on DIBaS dataset.

Study	Methods	Classes	Accuracy (%)
Zielinski et al. [9] (2017)	CNN, SVM, RF	33	97.24
Mohamed et al. [14] (2018)	BoW, SVM	10	97
Nasip et al. [15] (2018)	VggNet	33	98.25
The proposed (2019)	Resnet-50	33	99.12

Table 3 demonstrates that the proposed approach has achieved an average classification accuracy of 99.12%. The performance of the pre-trained ResNet-50 model also evaluated using precision, recall, and F1-score. The average precision, recall, and F1-score value for the validation sets was 99%. As a result, the proposed approach outperforms the existing studies in the literature for the same dataset.

VI. CONCLUSION

Bacteria are all over life. The beneficial species of bacteria have positive effects on human life. For example, they heal diseases and help digestion. However, harmful bacterial species affect human life negatively by causing diseases. The automated classification of the different type of bacterial species has great importance. In this study, ResNet-50 pre-trained CNN model is employed to classify bacteria species into 33 classes. This model has reached an average classification accuracy of 99.2%. The proposed study automatically classifies bacteria types without using any pre-processing technique. This model is ready to be examined in the field of clinical microbiology to automatically classify bacterial images.

REFERENCES

- [1] Nie, D., Shank, E.A. and Jovic, V., 2015, September. A deep framework for bacterial image segmentation and classification. In *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics* (pp. 306-314). ACM.
- [2] Hiremath, P.S. and Bannigidad, P., 2010, February. Automatic classification of bacterial cells in digital microscopic images. In *Second International Conference on Digital Image Processing* (Vol. 7546, p. 754613). International Society for Optics and Photonics.
- [3] W.M.Ahmed, B.Bayraktar, A.K.Bhunia, E.D.Hirleman, J.P.Robinson, and B.Rajwa. Classification of bacterial contamination using image processing and distributed computing. *IEEE Journal of biomedical and health informatics*. 17(1):232-239, 2013.
- [4] De Bruyne, K., Slabbinck, B., Waegeman, W., Vauterin, P., De Baets, B. and Vandamme, P., 2011. Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. *Systematic and applied microbiology*, 34(1), pp.20-29.
- [5] Goodacre, R., Burton, R., Kaderbhai, N., Woodward, A.M., Kell, D.B. and Rooney, P.J., 1998. Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. *Microbiology*, 144(5), pp.1157-1170.
- [6] Nie, D., Shank, E.A. and Jovic, V., 2015, September. A deep framework for bacterial image segmentation and classification. In *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics* (pp. 306-314). ACM.
- [7] Fiannaca, A., La Paglia, L., La Rosa, M., Renda, G., Rizzo, R., Gaglio, S. and Urso, A., 2018. Deep learning models for bacteria taxonomic classification of metagenomic data. *BMC bioinformatics*, 19(7), p.198.
- [8] LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, 521(7553), p.436.
- [9] Zieliński, B., Plichta, A., Misztal, K., Spurek, P., Brzychczy-Włoch, M. and Ochońska, D., 2017. Deep learning approach to bacterial colony classification. *PloS one*, 12(9), p.e 0184554
- [10] Pan, S.J. and Yang, Q., 2010. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), pp.1345-1359.
- [11] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.
- [12] Ketkar, Nikhil. "Introduction to pytorch." *Deep Learning with Python*. Apress, Berkeley, CA, 2017. 195-208.
- [13] Howard, J.a.o., fastai. *GitHub*, 2018
- [14] Mohamed, B.A. and Afify, H.M., 2018, December. Automated classification of Bacterial Images extracted from Digital Microscope via Bag of Words Model. In *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)* (pp. 86-89). IEEE.
- [15] Nasip, Ö.F. and Zengin, K., 2018, October. Deep Learning Based Bacteria Classification. In *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (pp. 1-5). IEEE.

Image Processing on Electrophoresis Image with Embedded System

K.ÖZKARA¹ and A.R. KAVSAOĞLU²

¹ Sinop University, Sinop/Turkey, kerimozkara@gmail.com

² Karabuk University, Karabuk/Turkey, kavsaoglu@karabuk.edu.tr

Abstract - The aim of this study is to analyze the agarose gel images in an embedded system with a program. Thanks to this program, the bands are found automatically and are made faster and smaller error rate than manual measurements and calculations. This program is designed using the python and openCV on Raspbian which is the Raspberry PI operating system.

Keywords – Electrophoresis Image Processing, Python, OpenCV, Raspbberri PI.

I. INTRODUCTION

Electrophoresis is based on the application of charged particles under the action of an electric field through the gel. These charged particles can be bacteria, viruses, protein molecules, nucleic acids or synthetic particles. The electrophoresis method is an electrokinetic method which is also preferred for reasons of effective separation, ease of application and cheapness. [1]

DNA gel electrophoresis is a molecular biology technique that separates different sizes of DNA particles. Between the Applications of DNA gel electrophoresis include DNA fingerprinting (genetic diagnosis), DNA separation for DNA size estimation. The correct interpretation of DNA band patterns on electrophoretic images can be error prone and inconvenient because there are many bands measured manually. Although many biological imaging techniques have been proposed, none of them can fully automate the spelling of the DNA, usually due to the complexity of the resulting migration patterns. [2]

II. ANALYSIS OF ELECTROPHORESIS IMAGES

The distance between heavy base pairs and light base pairs increases as a result of electrical pushing. The base pairs, which are heavy from the starting point, cannot travel too far, while the light base pairs will progress the further distances within the same period. A function is created between the distance travelled on the image and the number of base pairs on the ladder steps. With this function, the base pairs of the moving base pairs in other DNA samples are estimated by calculating the function of the pairs of base pairs. The accuracy of predictions is related to the closeness of R² value of function to 1. Therefore, the accuracy and sensitivity of measurements made before calculating the formula is very important. While the measurement with a ruler on an image can be in millimeter

sensitivity, measurements with a program can be reduced to micrometer and nanometer levels.

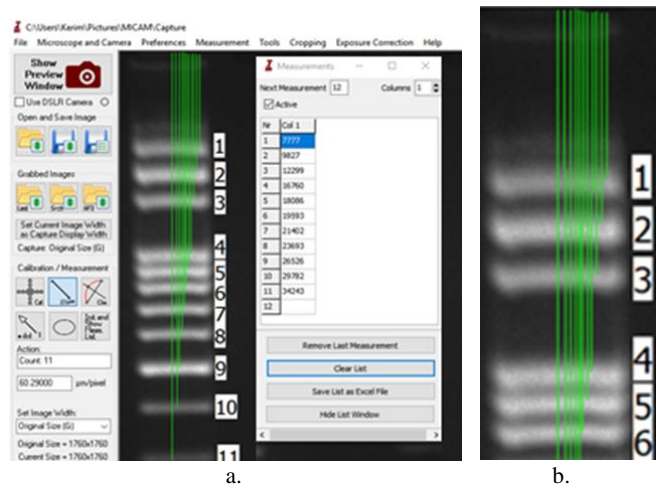


Figure 1:(a) Measure band distance with a measurement program.

(b) Starting points are not same level and finishing points are not at the fixed point of bands. Some measurements are not straight

This image in Figure 1 was captured by mini gel electrophoresis system based on embedded system [3]. The figure 1.a shows an agarose gel image measured as a micrometer by MiCam measurement program. As it is understood from the figure 1.b, it is observed that the reference points that are started to be measured cannot be taken at the same level, nor can it be observed that the equalization of the endpoints (start or midpoints of the bands) exactly for each band is not possible to measure. That is because of the type of mouse used and the pixel dimensions of the picture. Also, it is not easy to set the measuring lines straight. This situation affects actual measuring values. When all these errors are collected, the measurements can go away from their actual values.

The measurements in Figure 1 are shown in Table 1. According to Table 1, as a result of applied voltage to ladder DNA, it was measured that a band including 3000bp carried out 7777 μm , the band including 100bp carried out 48639 μm . An exponential function (1) which estimates the number of the bp depending on the way carried out was created. The R² value of this function is 0,968.

$$y = 1.10^9 \cdot x^{-1,702} \quad (1)$$

Table 1: Measured values and curve-fitting

	μm	Real Base
1	7777	3000
2	9827	2000
3	12299	1500
4	16760	1000
5	18086	900
6	19593	800
7	21402	700
8	23693	600
9	26526	500
10	29782	400
11	34243	300
12	40331	200
13	48639	100

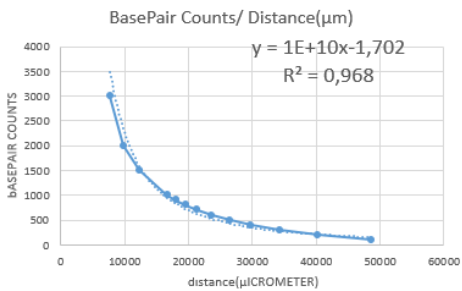


Table 2 shows the error percentages and the differences between the base pair numbers which was calculated by entering the distance travelled by base pairs in the function and the actual base pair numbers. Although the R^2 value is 0.968, the error values vary between 5%-40%. The average error value is around 31%.

Table 2: Calculated values from function and differences

	Um	Base Pair	Calculating	Difference	Error
1	7777	3000	2537	463	15,43
2	9827	2000	1603	397	19,85
3	12299	1500	1094	406	27,07
4	16760	1000	646	354	35,4
5	18086	900	568	332	36,89
6	19593	800	495	305	38,13
7	21402	700	426	274	39,14
8	23693	600	358	242	40,33
9	26526	500	296	204	40,8
10	29782	400	243	157	39,25
11	34243	300	191	109	36,33
12	40331	200	145	55	27,5
13	48639	100	105	-5	5
Average Error =					30,8554

III. ANALYSIS OF ELECTROPHORESIS IMAGE WITH ALGORITHM

The pixel values in the agarose gel image in grayscale are summed vertically and the edge points of the lanes are determined from the signal obtained. The pixel values in the lanes found are horizontally summed and the peaks of the bands are found from the signal obtained [4]. Row indices containing the peak points are defined as the distance in which base pairs are carried out. An exponential function(2) is created with the position pixels of the automatically detected bands.

$$y = 8.10^9. x^{-2,769} (2)$$

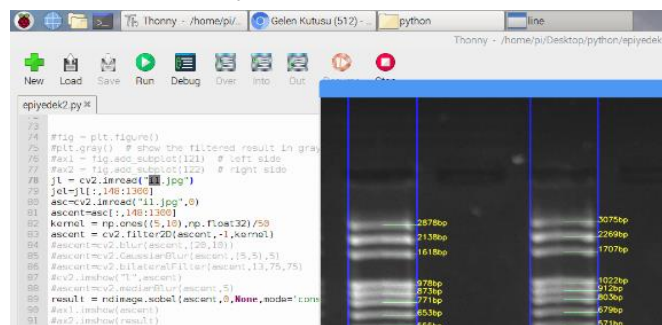
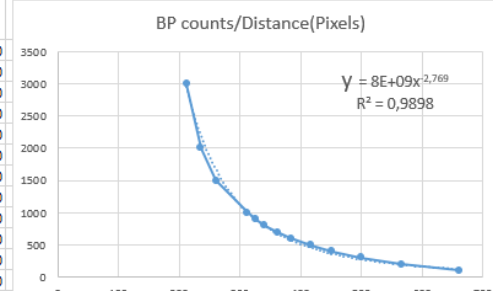


Figure 2: Auto detection location of band program in Raspbian

The R^2 value of the function was calculated as 0,9898 and the values in the table 3 were obtained.

Table 3: Detected pixel values and pixel function

	Pixel Position	Real Base
1	212	3000
2	236	2000
3	261	1500
4	312	1000
5	326	900
6	341	800
7	362	700
8	384	600
9	417	500
10	452	400
11	499	300
12	566	200
13	661	100



When these values are applied to the function, it is understood that the error values are between 2%-24% and the average error value is around 7,6% shown as the table 4.

Table 4: Calculated values from function and differences

	Pixel Position	Real Base	Calculating Base Pair	Difference	Error %
1	212	3000	2894	106	3,53
2	236	2000	2150	-150	7,5
3	261	1500	1627	-127	8,47
4	312	1000	993	7	0,7
5	326	900	879	21	2,33
6	341	800	776	24	3
7	362	700	658	42	6
8	384	600	559	41	6,83
9	417	500	445	55	11
10	452	400	356	44	11
11	499	300	270	30	10
12	566	200	191	9	4,5
13	661	100	124	-24	24
Average Error =					7,604615

IV. CONCLUSION

When R^2 values of both functions are compared, it is seen that R^2 value of the function calculated with the program is closer to 1. When the number of calculated base pairs is considered according to closeness to the actual values, it is seen that the values calculated with the program are closer to the actual values. According to these results, it is concluded that the interpretation of agarose gel images in an embedded system is better than manual imaging and measurements in terms of speed and accuracy. This system also consumes less power than a computer and SLR camera. The results are more quickly reached because they do not require manual labor.

ACKNOWLEDGEMENT

This study was supported by The Karabuk University Scientific Researches Projects (Project No: KBUBAP-17-YL-473) research grant. We are indebted for its financial support.

REFERENCES

- [1] L.Güneştutar, "New Approaches in Electrophoresis. Journal of Engineering and Natural Sciences(Sigma)". Pp.151-161, 2009.
- [2] A.Intarapanich, Automatic DNA Diagnosis for 1D Gel Electrophoresis Images using Bio-image processing Technique, Tokyo: BMC Genomics, 2015.
- [3] İ.Mersinkaya, A.R.Kavsaoglu, "Mini Gel Electrophoresis System Based on Embedded System", 26th Signal Processing and Communications Applications Conference (SIU), İzmir, pp. 1-4, IEEE, (2018)
- [4] M.K. Turan, "Identification of column edges of DNA fragment by using K-means clustering and mean algorithm on lane histograms of DNA agarose gel electrophoresis images", Proc. of SPIE Vol. 9631 963110-5

Feature Selection for Text Classification Based on Term Frequency and Inter-Class Standard Deviation

M.GÖÇENOĞLU¹ and İ.TÜRKER²

¹ Hakkari University, Hakkari/Turkey, mustafagocenoglu@hakkari.edu.tr

² Karabuk University, Karabuk/Turkey, iturker@karabuk.edu.tr

Abstract – Rapid evolution of Internet-based technologies induced outbreak of big data, which is handled with machine learning methods. High dimensionality of feature space is the most common problem for machine learning algorithms, especially in text classification studies. To overcome this problem, we proposed a new method based on maximizing the product of term frequency (TF) and inter-class standard deviation (SD) of term frequencies for each word, namely TF-SD. We performed text classification tasks for a subset of well-known 20Newsgroups dataset including 4 main classes. Applying various classification methods (J48, SMO, Naïve Bayes and Logistic Regression), we compared the success of this measure with term frequency-inverse document frequency (TF-IDF) based feature selection for various number of features. Results show that TF-SD based feature selection outperforms TF-IDF for all classification algorithms, exhibiting higher accuracy and also reaching a saturation of accuracy for very low term counts compared to TF-IDF.

Keywords – Text classification, feature reduction, term frequency, machine learning.

I. INTRODUCTION

RECENT advances in web technologies pioneered the availability of explosive amount of unstructured data which cannot be processed without machine learning algorithms. Many studies have been conducted to evaluate this “big data” with supervised or unsupervised learning algorithms. Well-known representors of text classification methods are k-nearest Neighbor (kNN) [1], Support Vector Machines (SVM) [2, 3], Naïve Bayes (NB) [4], Logistic Regression (LR) [5] and some decision tree based approaches like ID3, C4.5 etc [6, 7]. All these methods face a common challenge of large feature space, therefore feature reduction is the major problem involving text classification task [8].

Feature selection is generally defined as selecting representative features from the original feature space to reduce the dimensionality of feature space and improve the efficiency and precision of classifier. Feature selection algorithms are subdivided into many groups like filtering, wrapper, embedded and hybrid methods [8, 9]. Most common methods include term frequency (TF), document frequency (DF), information gain (IG), gain ratio, chi-square, relief, term strength etc.

DF thresholding is one of the most lightweight methods of

feature selection, defined as the number of documents in which a word is used. Terms having DF under a predetermined threshold is removed from the feature space, leading a simple technique available to scaling to a very large corpora, but not the best for aggressive reduction [10]. Another low-complexity method is TF based filtering, eliminating low frequency words from the feature space. These two methods are further combined as TF-IDF, a common approach used in information retrieval to specify importance of each term by the means of both appearance frequency in a specific document and a global frequency for that term implying how rare it occurs among documents [11, 12].

For more aggressive reductions, information gain is a more promising method, measuring the number of information bits for classifying a document subject to presence of a word in a document. Term strength is another method estimating term importance based on how common a term appears in closely-related documents [13]. These closely-related documents are determined after a training task of documents and assigning documents with cosine similarities above a threshold to the same group. Term strength is calculated as a conditional probability of a word to appear in a group of documents given that it appears in the other group [14].

Among the most common filtering methods, a statistical feature selection method, chi-square (χ^2) stands out with high success in reduction of feature space. It measures the dependence of a specific word with a document category, where χ^2 measure gives zero if the appearance of a word is independent from a category [15].

Our study focuses on a new feature selection method based on inter-class dependencies of terms, with a similar but lightweight way compared to chi-square. Our method incorporates global TF with inter-class standard deviation (SD) of term usage, promoting selection of terms with relatively high global usage together with high variations of usage among classes. The method is detailed in Section II, where results are presented in Section III and concluded in Section IV.

II. DATA AND METHOD

A. Dataset

For the text classification task, we selected a subset of well-known text corpus 20Newsgroups, including 4 main classes

listed as in Table 1. We pre-processed this raw dataset by removing terms including non-alphanumeric characters, html tags, rare words with global frequency less than 5, stop-words and short words having 1 or 2 letters. We later filtered out entries including less than 80 words. By the way we achieved a resulting dataset of 7209 entries from 4 main classes.

Table 1: Subset of 20Newsgroups dataset used in this study.

1	COMPUTERS comp.graphics, comp.os.ms-windows.misc, comp.sys.ibm.pc.hardware, comp.sys.mac.hardware,, comp.windows.x
2	POLITICS talk.politics.misc, talk.politics.mideast talk.politics.guns,
3	AUTO&SPORTS rec.autos, rec.motorcycles, rec.sport.baseball,, rec.sport.hockey
4	RELIGION talk.religion.misc, soc.religion.christian

B. Method for Attribute Selection

We propose a new method promoting two properties of terms: global frequency and inter-class standard deviation. The former one is known as TF, regarding the normalized count of usage of words in whole documents. The latter one is the inter-class standard deviation (SD), calculated for each word using the average TF for each class. We multiply these two parameters for each word to achieve the TF-SD measure as in Eq. 1.

$$TF_iSD_{ij} = \frac{f_{ij}}{\sum_{j=1}^m f_{ij}} \cdot \sqrt{\frac{\sum_{k=1}^c |TF_k - TF_{avg}|^2}{c-1}} \quad (1)$$

In Eq. 1, f_{ij} is the frequency of term t_j in document d_i and denominator is the total frequencies of all terms in document d_i . This left side corresponds to TF, while right side evaluates SD using TF_k for mean normalized frequency of the term for each class and TF_{avg} for the TF for that term averaged over c classes.

We sorted whole terms in descending order according to TF-SD to select the terms with highest score. To trace classification performance for various number of terms, we performed classifications in Weka environment [16] for 100 to 1000 terms with an increment of 100 terms at each run.

For comparative purposes, we also performed the similar procedure above for well-known TF-IDF based feature selection. TF-IDF score incorporates the term frequency with the importance of that term, as in Eq. 2.

$$TF_iIDF_{ij} = \frac{f_{ij}}{\sum_{j=1}^m f_{ij}} \log\left(\frac{|D|}{|d_i t_j \in d_i \in D|}\right) \quad (2)$$

Where $|D|$ is the total count of documents in the corpus and

the denominator term in the right side is the total count of documents containing term t_j . Details of TF-IDF based scoring for feature selection can be accessed via Ref. [17].

III. RESULTS

For each attribute set derived from TF-IDF and TF-SD methods, we performed classification tasks for 4 machine learning algorithms as: J48, SMO, Naïve Bayes and Logistic Regression. We applied cross-fold validation with 6 folds to test classification performances.

We present accuracy values for these 4 classification algorithms for various numbers of features in Table 2, also plotted in Fig.1.

Table 2: **Accuracy** results for 4 classification algorithms, each one executed with 2 feature selection methods and 10 feature sets. Maximal values for each feature set and feature selection method are given in bold fonts.

Num. Terms	J48		SMO		Naive Bayes		Logistic Reg.	
	TF-IDF	TF-SD	TF-IDF	TF-SD	TF-IDF	TF-SD	TF-IDF	TF-SD
100	41.99	77.49	44.32	81.32	42.74	79.49	45.51	81.78
200	51.40	81.28	55.46	86.79	48.90	84.79	56.50	86.30
300	56.95	81.50	61.87	87.63	57.38	85.59	62.08	86.16
400	60.21	81.46	66.96	87.50	60.61	85.64	66.74	85.22
500	62.36	81.78	69.93	87.94	63.01	86.50	68.72	85.32
600	64.16	81.99	71.72	88.04	65.51	86.44	70.83	81.68
700	64.81	81.99	73.90	87.68	67.28	86.57	71.76	77.01
800	66.63	82.20	75.72	88.22	68.52	86.68	72.59	75.20
900	67.52	82.11	76.40	87.6	70.01	86.73	72.54	73.46
1000	68.35	81.78	76.85	88.23	70.71	86.50	72.78	72.43

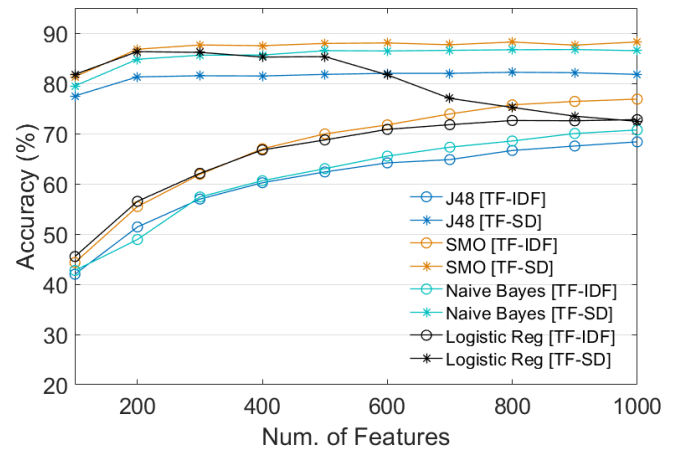


Figure 1: **Accuracy** plots for 4 classification algorithms, each one executed with 2 feature selection methods and 10 feature sets.

Accuracy results indicate that TF-SD based feature selection outperforms TF-IDF for all feature sets and classification algorithms. Beyond high accuracy, a noteworthy property of TF-SD is resulting the accuracy saturation for very low number

of features, generally for 200 or 300. Starting from these values, accuracy remains stable towards 1000 features. Accuracy for logistic regression decreases sensibly after 500 features. On the other hand, being far from reproducing the success of TF-SD, accuracy shows an increasing trend for TF-IDF based feature selection for all algorithms. The exception of this prospect is the right half of the logistic regression plot. To conclude, reaching about 80% accuracy for a 4-class dataset with 100 features seems satisfying, while ~200 features emerge as an optimal set for all algorithms if TF-SD method is employed.

Table 3: Weighted avg. **F1-Score** results for 4 classification algorithms, each one executed with 2 feature selection methods and 10 feature sets. Maximal values for each feature set and feature selection method are given in bold fonts.

Num. Terms	J48		SMO		Naive Bayes		Logistic Reg.	
	TF-IDF	TF-SD	TF-IDF	TF-SD	TF-IDF	TF-SD	TF-IDF	TF-SD
100	36,89	77,48	40,09	81,26	37,66	79,45	41,95	81,71
200	49,46	81,27	53,68	86,83	49,15	84,66	55,06	86,32
300	55,72	81,61	60,69	87,7	56,78	85,52	61,27	86,12
400	59,17	81,48	66,31	87,53	60,66	85,54	66,32	85,2
500	61,7	81,79	69,48	87,9	63,11	86,43	68,43	85,31
600	63,69	82,03	71,41	87,99	65,45	86,33	70,6	81,81
700	64,41	81,97	73,68	87,69	67,17	86,53	71,57	77,18
800	66,27	82,17	75,49	88,18	68,44	86,59	72,47	75,44
900	67,24	82,09	76,16	87,57	69,89	86,61	72,47	73,61
1000	68,13	81,77	76,66	88,18	70,69	86,39	72,8	72,46

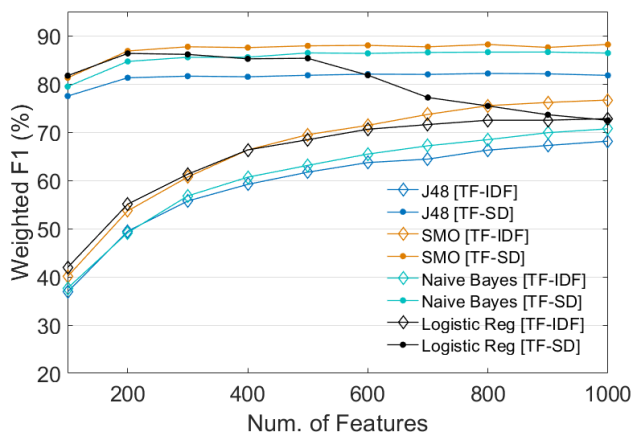


Figure 2: Weighted avg. **F1-Score** plots for 4 classification algorithms, each one executed with 2 feature selection methods and 10 feature sets.

We also present F_1 -score measures for the same experimental setup numerically in Table 3 and graphically in Fig. 2. This is the average weighted version of F_1 measure, which is the harmonic mean of precision and recall. Reader is referred to Ref. [18] for further details about this measure.

F_1 -score, as a better indicator of success in multi-class and unbalanced datasets [19], show very similar trends with

accuracy with slightly lower success levels. We achieve about 87-88% for SMO with TF-SD, for all feature sets larger than 100 words. Other algorithms similarly display stable results for these feature sets giving F_1 -scores slightly below SMO. The decreasing phase of logistic regression for higher feature sets is observed similarly, while a peak score for this algorithm is achieved for 200 features. To conclude, we can say that accuracy and micro- F_1 scores are very similar, consistent with statistical expectation.

IV. CONCLUSION

A new feature selection method based on statistical dependence of classes to terms together with term frequency is proposed. Results show that the method is very effective under low-dimension constraints. Classification success is notably high compared to TF-IDF based feature selection, also reaching a high saturation value for very low numbers of features like 100-200. This property enables usage of this method on aggressive feature reduction demands.

Comparisons between classification algorithms under same feature selection method condition shows similar results with recent studies, featuring high success for SMO and logistic regression, while other two algorithm has moderate outputs. Logistic regression algorithm, preserving increasing success trend for TF-IDF based feature selection for larger feature sets, emerges as inadaptate for large feature sets of TF-SD method.

A future work is planned by the authors to extend this comparison work to span more feature selection/reduction methods like chi-square, which is statistically similar with the approach proposed in this study.

ACKNOWLEDGEMENT

The dataset used in this study is a subset of 20Newsgroups dataset available at: <http://qwone.com/~jason/20Newsgroups/>

REFERENCES

- [1] M. R. Abbasifard, B. Ghahremani, and H. Naderi, "A survey on nearest neighbor search methods," *International Journal of Computer Applications*, vol. 95, 2014.
- [2] D. Meyer and F. T. Wien, "Support vector machines," *The Interface to libsvm in package e1071*, p. 28, 2015.
- [3] S. Suthaharan, "Support vector machine," in *Machine learning models and algorithms for big data classification*, ed: Springer, 2016, pp. 207-235.
- [4] B. Tang, S. Kay, and H. He, "Toward optimal feature selection in naive Bayes for text categorization," *IEEE transactions on knowledge and data engineering*, vol. 28, pp. 2508-2521, 2016.
- [5] F. E. Harrell, "Ordinal logistic regression," in *Regression modeling strategies*, ed: Springer, 2015, pp. 311-325.
- [6] S. Sathyadevan and R. R. Nair, "Comparative analysis of decision tree algorithms: ID3, C4. 5 and random forest," in *Computational Intelligence in Data Mining-Volume 1*, ed: Springer, 2015, pp. 549-562.
- [7] Y.-Y. Song and L. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai archives of psychiatry*, vol. 27, p. 130, 2015.
- [8] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2015, pp. 1200-1205.

- [9] N. Hoque, D. K. Bhattacharyya, and J. K. Kalita, "MIFS-ND: A mutual information-based feature selection method," *Expert Systems with Applications*, vol. 41, pp. 6371-6385, 2014.
- [10] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *Icml*, 1997, p. 35.
- [11] K.-C. Lin, K.-Y. Zhang, Y.-H. Huang, J. C. Hung, and N. Yen, "Feature selection based on an improved cat swarm optimization algorithm for big data classification," *The Journal of Supercomputing*, vol. 72, pp. 3210-3221, 2016.
- [12] M. Mohamad and A. Selamat, "An evaluation on the efficiency of hybrid feature selection in spam email classification," in *2015 International Conference on Computer, Communications, and Control Technology (I4CT)*, 2015, pp. 227-231.
- [13] R. B. Pereira, A. Plastino, B. Zadrozny, and L. H. Merschmann, "Categorizing feature selection methods for multi-label classification," *Artificial Intelligence Review*, vol. 49, pp. 57-78, 2018.
- [14] Y. Lin, J. Zhang, X. Wang, and A. Zhou, "An information theoretic approach to sentiment polarity classification," in *Proceedings of the 2nd joint WICOW/AIRWeb workshop on web quality*, 2012, pp. 35-40.
- [15] Y.-T. Chen and M. C. Chen, "Using chi-square statistics to measure similarities for text categorization," *Expert systems with applications*, vol. 38, pp. 3085-3090, 2011.
- [16] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, pp. 10-18, 2009.
- [17] L. H. Patil and M. Atique, "A novel approach for feature selection method TF-IDF in document clustering," in *2013 3rd IEEE International Advance Computing Conference (IACC)*, 2013, pp. 858-862.
- [18] V. Van Asch, "Macro-and micro-averaged evaluation measures," *Tech. Rep.*, 2013.
- [19] A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating probability with undersampling for unbalanced classification," in *2015 IEEE Symposium Series on Computational Intelligence*, 2015, pp. 159-166.

COMPARISON OF MACHINE LEARNING ALGORITHMS FOR FRANCHISE APPROVAL

W. Colley¹ and M. Göktürk²

¹Gebze Technical University, Gebze/Turkey, colley@gtu.edu.tr

²Gebze Technical University, Gebze/Turkey, gokturk@gtu.edu.tr

Abstract—Credit scoring is defined as a technique that enables banks or credit issuers in the granting of credit, these techniques decide who is to get credit or not. In the past this task used to be carried out by analysts but as the financial and credit granting institutions are increasing and the call for credit increases each passing day, a method or procedure to evaluate credit applicants becomes a challenging issue that needs to be addressed wisely and effectively. For this reason, credit scoring approaches will be used to rate branch office applicants for franchise operations, applicants that will or will not be able to successfully run a franchise as we believe franchise approval is very similar to personal credit scoring, hence six machine learning algorithms were selected for performance comparison since an atom of increase in accuracy is a significant accomplishment however, in these operations, normally fewer attributes are usually selected and used in classification from a higher number of attribute set, with the use Principal component analysis (PCA) we ended up with 31 features that contributed 95% of the variance and all six models attained accuracies far greater than the baseline accuracy of 75% with Support vector machine 98.80% and multilayer perceptron 98.74% as best respectively.

Keywords: franchise approval, machine learning, credit Scoring, PCA, grid search

1. Introduction

Credit scoring is defined as a technique that enables banks or credit issuers in the granting of credit to customers, these techniques decide who will get credit or not, the amount to get, and what useful strategies will help enhance profitability etc. [1] In this regard, this study will be using credit scoring approaches to rate branch office applicants for franchise operations. Applicants that will or will not be able to successfully run a franchise?

Franchising is based on a marketing concept that can be used by a company as a way for business expansion. When this strategy is implemented, a franchiser licenses its know-how, intellectual properties, use of business model, brand, and rights to sell its branded products and services to a franchisee. In return the franchisee pays certain fees and comply with certain obligations, typically set out in a franchise Agreement. For this reason, in our opinion, an

appropriate credit scoring technique is a crucial part towards a successful franchise approval and evaluation.

Abdou and Pointon [3] conducted an in-depth review of 214 articles and books concerned with applications of credit scoring in various areas of business. They found out that there is no technique considered overall best for credit scoring models across different datasets. Abdou and Pointon [3] in line with Hand and Henley [4] argue that performance of classification models is dependent on many factors ranging from available features in data set to the data structure or just the objective of classification etc.

On the other hand, Baesens et al [5] in an experiment with eight real life datasets embarked on studying the performance of several well known algorithms (e.g. logistic regression, discriminant analysis, k-nearest neighbor, neural networks and decision trees), together with, advanced kernel-based classification models such as support vector machines and least-squares support vector machines (LS-SVM) which lead to significant performance differences. It was found that both the LS-SVM and neural network models yielded very good performances, but also simple models such as logistic regression and linear discriminant analysis perform very well for credit scoring.

Several years later, Lessmann et al [6] set out to update Baesens et al. [5] and to explore the relative effectiveness of alternative classification models in retail credit scoring where they compared 41 models in terms of six performance measures across eight real-world credit scoring datasets. Their results suggest that several models predict credit risk significantly more accurately than Logistic Regression. Making mention of the good performance of heterogeneous ensembles models.

With time, more sophisticated, reliable and efficient models known as artificial intelligence, neural networks, ensembles, genetic algorithms among others continue to emerge, among these, the neural networks and ensembles have been used a lot and proven to be very promising Goonatillake and Treleaven, [7] Šušteršič et al [8] Fatemeh et al [9] Wang et al [10] Bensić et al [11] Harris [12] and Thomas et al [1]. Other artificial intelligence methods like support vector machines, decision trees were also adopted by many researchers in constructing credit scoring models.

Dealing with datasets of high dimensionality caused some authors/researchers like Šušteršič et al [8] Li et al [13] Fan et al [14] Fatemeh et al etc. [9] to consider the need for

feature selection/ extraction using Principal component analysis and Genetic algorithm to improve model accuracy as well as reduce the time needed to build classification models. few other researchers consider hyperparameter optimization as an important factor for performance improvement in some models like support vector machines and realized good results Xia et al[15] Šušteršič et al[8].Bischi et al[16] Wang and Ni [17].

All of the above are in relation to credit scoring in financial institutions like banks and credit bureaus but for the first time in this paper we will look in to franchise approval using credit scoring concepts. The main aim being to develop and compare performances of different machine learning models like neural networks, decision tree, random forest, logistic regression, naïve Bayes and support vector machines on a franchise approval dataset. but as we are dealing with large amount of data with 72 dimensions, the need for PCA dimensionality reduction is necessary in this study as well. Approaches taken towards this research is broken in to three main stages:

- 1) A brief description of the models used.
- 2) Steps towards Franchise Approval process(data pre-processing,PCA,hyperparameter turning)
- 3) Build the models and finally evaluate and compare their performances.

2. Machine Learning Techniques Used

2.1. Multilayer Perceptron (MLP)

A neural network (or an artificial neural network) [18] is inspired by the way the brain works in order to process information. It comprises of highly interconnected units called neurons working together to solve specific problems. Neural networks, like human beings, learn by examples. That is, they learn by experience, to make better and efficient decisions.[19] one of the famous types of neural networks are the multilayer perceptron (MLP) that consists of three layers: input layer, processor/hidden layer, and an output layer. where the three layers are interconnected in some way. The input layers represents the raw information fed into the model. Each hidden layer determined by the activities of the input units and the individual weights on their connections, The output layer otherwise called the layer that gives the result depends on the activity of the hidden layer and the weights between the hidden and output.[19]Neural networks are a widely used technique in that they are very efficient but they also have a disadvantage in their long model training phase.multilayer perceptron uses the back propagation approach that uses the gradient descent algorithm. the four main steps used by an MLP are:

- 1) get an input or inputs.
- 2) weight those inputs.
- 3) add the inputs
- 4) give an output.

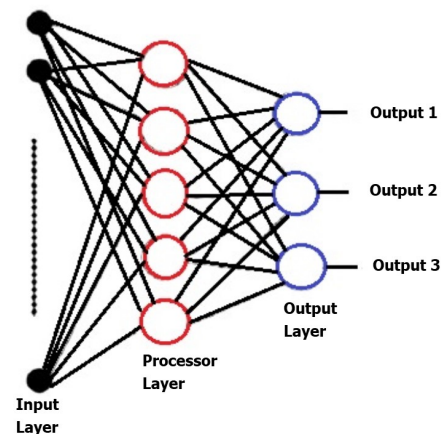


Figure 1. Three Layered Neural Net

2.2. Naive Bayes (NB)

A famous model based on the Bayesian theory in that they make independent assumptions that are indeed naive and yet provide interesting trade off on model complexity that sometimes turn out to be useful hence, the name naive. this assumption is called class conditional independence and assumes that every pair of features X_1 and X_2 are conditionally independent given the class C . Bayesian models are useful in that they are computationally efficient and easy to implement but also have a disadvantage with strong independence assumptions that reduce performance when dealing with highly correlated features.[20]

2.3. Logistic Regression (LR)

Logistic regression (LR) is a common machine learning technique very similar to linear regression except in how the line is fit to the data. A logistic regression model simply predicts if something is true or false by fitting in a logistic function, that means the curve tells the probability that a customer will be able to successfully run a franchise based on certain criteria or not.[20] for example if the probability of a customer is good is greater than 50% it classifies it good otherwise not. The major advantage of LR is that it has the ability to provide probabilities and classify new samples by not just using continuous measurements but also discrete ones.

2.4. Support Vector Machines (SVM)

Support Vector Machine (SVM) is a supervised Machine Learning method used for classification of both linear and nonlinear data. It can also be used for numeric predictions. SVM is known for transformation of data into a higher dimension so as to find a linear optimal separating hyperplane or decision boundary that separate classes. Such a hyperplane is the one that gives the widest separation between the

classes[21].Support vectors (dark blue ones and dark black zero in the diagram below) in SVM are so crucial that when found the rest of the points can be ignorable.figure 2 below gives a brief description.

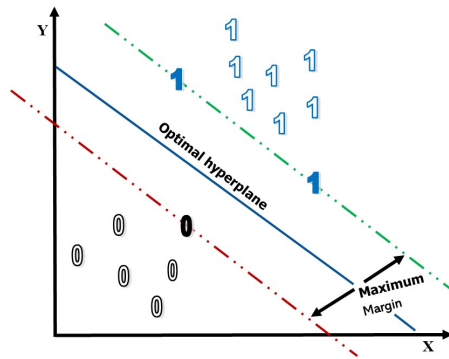


Figure 2. Support Vector Machine

2.5. Decision Tress (DT)

They are also one of the commonly used models for performing classification and regression tasks. Decision tree learning is a tree based model used to determine a course of action where each branch of the tree represents a decision and works for both numeric and categorical data.It is an enhanced version of C 4.5 which revolves on the Iterative Dichotomiser 3 (ID3) algorithm with some extra functionality, However, this technique can be time and space consuming. Initially, it builds a tree using the divide and conquer algorithm and then applies heuristic criteria. The rules according to which the tree is generated are precise and intuitive.The main advantage of the decision tree model is that they are simple to understand visualize and interpret and non linear parameters do not affect its performance. However, if not pruned adequately, trees can become complex and overfit the training data which is one disadvantage of decision trees. [22]

2.6. Random Forest Ensemble Learning (RF)

Ensemble as the name implies means grouping, so it is the grouping of multiple learning algorithms at the same time and then base the algorithm on a majority vote over the individual trees to obtain better accuracy. Random forests are an ensemble of decision tree where each tree is constructed by applying an algorithm A on the training set S and an additional random vector sampled independently and with the same distribution for all trees in the forest [23]. This model can be used for both classification and regression purposes. To understand Random Forest more intuitively it was necessary to understand decision trees that is why in this study random forest is used in comparison with decision tree model as its ensemble.

Figure 3 below explains simply how Random Forest algorithm works in the following steps:

- 1) Picks random K data points from the training data.
- 2) Builds a decision tree for these K data points.
- 3) Chooses the N tree subset from the trees and performs step 1 and step 2
- 4) Decides the category or result on the basis of the majority of votes.

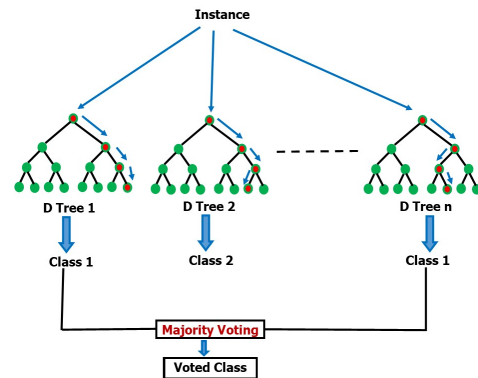


Figure 3. Random Forest Summarized

3. Steps Towards Franchise Approval Process

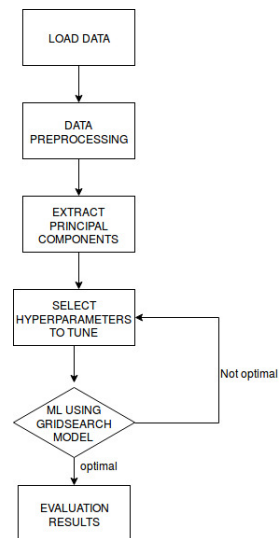


Figure 4. Experimental Procedure

four main steps where considered in carrying out this resaerch and they are briefly explained below

3.1. Data Description and Preprocessing

The dataset used for this study was collected from franchise application company to rate branch office applicants for franchise operations and consist of 20880 instances,71 features with a two-class target variable of positives(+) and negatives(-) where positives stand for people evaluated as

those who can successfully run a franchise constituting 15668 of the records and negatives for those that may not be able to successfully run a franchise with the remaining 5212 records. the dataset contains a good mix of numeric, categorical and string variable types. however, as real world datasets usually come in bulky, complicated and full of missing values that makes it a bit difficult to be directly fed in to a model, for this reason, the data had to go through a stage called the preprocessing stage where it was scanned for missing values and replaced with mean, outliers detected and handled, data transformation like changing target variables from +/- to 1/0, data encoding as required in machine learning, data scaling, feature selection which also helped in solving the data imbalance problem, hyper parameter tuning using grid search algorithm among others towards getting better classification results.

Name of Dataset	No. of Instances	No. of Features
Franchise Approval	20880	72

3.2. Principal Component Analysis (PCA)

A useful approach called Principal component analysis (PCA) according to Šušteršič et al[8] is a technique used mostly for reducing the dimensionality of a given dataset of correlated variables while maintaining as much of the feature's variability as possible. This efficient reduction of features is achieved by obtaining the so-called principal components (PCs). they further went on to say this is possible with a transformation of the features in to a new dimensional feature space. The transformation is done by rotation of the old coordinate system into the new one in such a way that most of the relevant information is collected around smaller dimensions (PCs) [8]. The first principal component been the most important and highly ranked, then the second been the second best, third and so on. In PCA high variance is very important and in this study 31 features out of the 72 that contributed 95% of the variance were selected.

3.3. Hyperparameter Turning

As the default parameter turnings were not quiet optimal, best hyperparameters for optimization were selected by using "grid search" algorithm. Grid-search is a way of selecting the best out of a family of models, parametrized by a grid of parameters. 5-folds cross validation was used accross all models except for naive Bayes to avoid over fitting. 5-folds was suitable for our case according to the size of our data.

4. Model Evaluation and Comparison

The principal objective in this study is to compare six prediction algorithms for a franchise approval and the algorithms used on the dataset are Naïve Bayes, Logistic Regression, MLP, SMO, Random forest and decision tree

and 5-fold cross validation was used.

Accuracy is most of the time, the first measure mostly used to evaluate and compare the performances of machine learning models. But given an unbalanced two-class data where most of the data falls in one class, a model can get the best accuracy by simply classifying the whole data as the class with most of the data. dummy classifier of sklearn is such an algorithm that only considers the class with the majority in classification and ignores the least, therefore it is used as our baseline. This however, shows that accuracy alone doesn't determine the actual capability of a model. Hence Precision, Recall and their combination F-measure are also used to make sure the models really performed well on the data.

Finally, all six models used in this research were build on python 3 on anaconda jupyter notebook and attained accuracies more than the baseline accuracy of 75% with all models having above 95% accuracy, precision, recall as well as f-measure, however, Support vector machine and multi-layer perceptron got the highest accuracy with 98.80% and 98.74% respectively.

Models	Accuracy	Precision	Recall	F-measure
NB	96.41%	96.00%	96.00%	96.00%
LR	98.16%	98.00%	98.00%	98.00%
SVM	98.80%	99.00%	99.00%	99.00%
DT	97.82%	98.00%	98.00%	98.00%
RF	97.39%	97.00%	97.00%	97.00%
MLP	98.74%	99.00%	99.00%	99.00%

Figure 5. model comparison

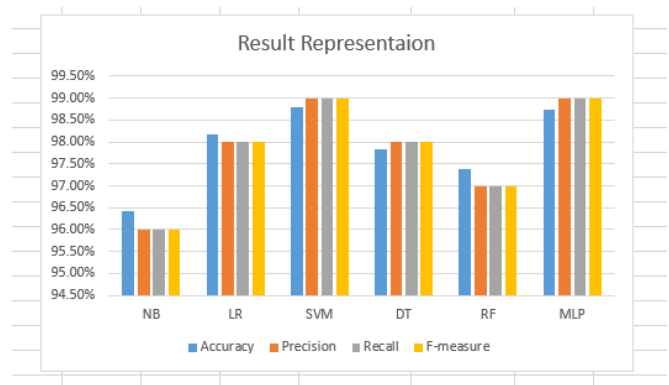


Figure 6. Graphical View of Results

5. Conclusion

In this paper, six different state-of-the-art machine learning models used mostly in credit scoring to enhance

faster, cheaper and easier decision making were compared in terms of performance to rate applicants for franchise operations and it was very effective. Moreover, PCA for dimensionality reduction along side hyper-parameter optimization are examined in the modeling procedure to evaluate the effectiveness and feasibility of the models on the franchise approval data set and results demonstrated that all six models performed far greater than the 75% baseline accuracy of dummy classifier with SVM and MLP been a little more than the rest with 98.80% and 98.74% respectively furthermore, precision recall and f-measure were also looked into to make sure the model really did perform on the dataset.

References

- [1] Thomas, L., Crook, J., Edelman, D. "Credit scoring and its applications (Vol.2.Siam)."2017
- [2] West, D. "Neural network credit scoring models. Computers and Operations Research"2000
- [3] Abdou, H. and J. Pointon. "Credit scoring, statistical technique and evaluation criteria: a review of the literature. Intelligent systems in accounting, finance and management 18(2-3):"2011
- [4] Hand, D. J., and Henley, W. E. "Statistical classification methods in consumer credit scoring: a review. Journal of the Royal Statistical Society: Series A (Statistics in Society), 160(3), 523-541."1997
- [5] Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., and Vanthienen, J. "Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the operational research society, 54(6), 627-635."2003
- [6] Lessmann, S., Baesens, B., Seow, H. V., and Thomas, L. C. "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), 124-136."2015
- [7] Goonatilake, S., and Treleaven, P. (Eds.). "Intelligent system for finance and business. Chichester: John Wiley and Sons."1995
- [8] Šušteršič, M., Mramor, D., Zupan, J. "Consumer credit scoring models with limited data. Expert Systems with Applications, 36(3), 4736-4744."2009
- [9] Koutanaei, F. N., Sajedi, H., and Khanbabaei, M. "A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. Journal of Retailing and Consumer Services, 27, 11-23."2015
- [10] Wang, G., Hao, J., Ma, J., and Jiang, H. "A comparative assessment of ensemble learning for credit scoring. Expert systems with applications, 38(1), 223-230."2011
- [11] Bensic, M., Sarlija, N., and Zekic-Susac, M. "Modelling small-business credit scoring by using logistic regression, neural networks and decision trees. Intelligent Systems in Accounting, Finance and Management: International Journal, 13(3), 133-150."2005
- [12] Harris, T. "Credit scoring using the clustered support vector machine. Expert Systems with Applications, 42(2), 741-750."2015
- [13] Li, W., and Liao, J. "An empirical study on credit scoring model for credit card by using data mining technology. In 2011 Seventh International "2011
- [14] Yan-qin, F., You-long, Y., and Yang-sen, Q. "Credit scoring model based on PCA and improved tree augmented Bayesian classification."2013
- [15] Xia, Y., Liu, C., Li, Y., and Liu, N. "A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. Expert Systems with Applications, 78, 225-241. "2017
- [16] Bischl, B., Kühn, T., and Szepannek, G. "On Class Imbalance Correction for Classification Algorithms in Credit Scoring. In Operations Research Proceedings 2014 (pp. 37-43). Springer, Cham. "2016
- [17] Wang, Y., and Ni, X. S. "A XGBoost risk model via feature selection and Bayesian hyper-parameter optimization. arXiv preprint arXiv:1901.08433. "2019
- [18] Haykin, S. "Neural networks: a comprehensive foundation (2nd ed.).New Jersey: Prentice Hall"1999
- [19] Tsai, C. F., Wu, J. W. "Using neural network ensembles for bankruptcy prediction and credit scoring. Expert systems with applications"2008
- [20] Yeh, I. C., Lien, C. H. "A The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Systems with Applications"2009
- [21] Han, Jiawei, Jian Pei, and Micheline Kamber. "Data mining: concepts and techniques. Elsevier"2011
- [22] Bunker, R. P., Zhang, W., Naeem, M. A. "Improving a Credit Scoring Model by Incorporating Bank Statement Derived Features. arXiv preprint arXiv:1611.00252."2016
- [23] Breiman, L. "Random forests. Machine learning, 45(1), 5-32."2001

Face Detection Using Forensic Software and Deep Learning Methods on Images and Video Files

M.KAYA¹, S.KARAKUŞ², F.ERTAM³

¹ Digital Forensic Engineering Dept. Technology Faculty, Firat University, Elazig, TURKEY, mkaya@firat.edu.tr

² Digital Forensic Engineering Dept. Technology Faculty, Firat University, Elazig, TURKEY, 161144101@firat.edu.tr

³ Digital Forensic Engineering Dept. Technology Faculty, Firat University, Elazig, TURKEY, fatih.ertam@firat.edu.tr

Abstract—Recently, face detection and classification have become a popular field of study thanks to the Convolutional Neural Networks (CNN), and datasets containing face images such as VGGFace. In this study, we aim to recognize faces within digital files obtained from the criminal objects using forensic software tools in judicial cases. To this end, we make this process autonomous by contributing to the examination of digital evidence via deep learning methods as detection and classification. For classification of the faces two different deep learning face classification methods presented in this paper, which are the cosine similarity and transfer learning method. For training of the CNN model presented in this paper, Resnet50 deep learning face recognition method has been used on the VGGFace2 dataset, and Vgg16 has been used on the VGGFace. The proposed model has obtained 100% success by classifying all of the 721 face images on face images larger than 40x40x3 size which can be detected by opencv. As a result of the work, we present multi classification as a deep learning method which can be used for autonomous face detection of persons on images and video files obtained from the digital evidence.

Keywords - Convolutional Neural Network, Digital Evidence, Face Detection, Image Classification

I. INTRODUCTION

Recent developments in the digital world have increased the prevalence of electronic devices, and there has been much progress in terms of processor power and storage. Electronic devices are defined as devices that can store data or connect data to remote servers, which contain voice, image, video and text-based information directly or indirectly on the subject of crime in the field of forensic information [1].

Electronic devices with data storage capabilities are very important in terms of forensic analysis as they carry a lot of information about the environment and users they use. The devices obtained from crime scene for investigation can contain tens of thousands of files, and these data are analyzed in the forensic information process and the criminal data or persons are identified and reported after being classified. Although this process consists of many sub-stages, it has been reported by Reith et al in 4 stages and is shown in Figure 1 [2].

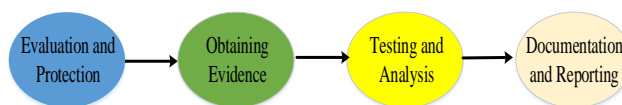


Fig. 1. Block Diagram of Forensic Science Investigation Process

In the examination of the tens of thousands of files during the forensic process, generally, evidence is obtained by indexing the metadata of files. In forensic information analysis, the information in the existing data files is obtained by indexing on the meta data. All of the forensic software used in the field of digital forensic does not have any content control on the data. Because content control cannot be done autonomously, it is done by experts using manual methods. Considering the devices contain thousands of images and for hours of video data, doing this examination by experts has a negative effect on both performance parameters such as time and accuracy. These images and video files must be extracted from the media on the electronic devices, and analyzed in the light of the information obtained. For example, on the image and video data, it is important to detect the human faces on the images and then to determine who the person is or who the person involved to this image.

Face detection and person recognition is an important issue in the field of computer science and computer vision. Although, it is not only to determine the content in the area of forensic information but also the person to whom the image belongs should be determined too. In the process, which requires such intensive work, things to do takes a lot of time even done by an expert. Therefore, there is a loss of attention and consequently the accuracy performance is low. It is often not possible to reach a conclusion.

In Turkey, the number of electronic devices confiscated for examination in 2016. According to information from the Police sources has become a huge work load increased by 30 times from a year prefix [3]. When all this information is considered together, while the content examination in the forensic computing process should be done by a computer-based decision support system, the detection and identification of the face from the images and video files that are the subject of this study will provide an important contribution to this process.

II. RELATED WORK

A. Computer Vision, Deep Learning and Face Recognition

The Computer vision is defined as the extraction and rendering of high-level images from digital images and videos [4]. The basis of the computer vision dates back to the early 1950s and has keeping up to evolve along with concepts such as artificial intelligence and neural networks [5]. F. Rosenblatt, who was also known as the inventor of artificial neural networks in 1957, established the Perceptron single-layer neural network machine by introducing the communication between the nerves in the brain. [6]. This network tried to perform the same process on computers by taking the samples that the neurons in the human brain had established with each other in the learning phase. In the 1970s, by developing the first commercial product of computerized vision, they developed the ORC (Optical Character Recognition) device so that the handwriting was recognized by the computer. In the 1980s, Hinton and friends presented the back propagation algorithm by performing a major breakthrough, then artificial neural networks were implemented in the field of computer vision [7]. In 1989, LeCun and his colleagues used the back propagation algorithm to recognize the handwriting obtained from the American Postal Service [8]. In 1988, Kirby and Sirovich began implementing linear algebra for facial recognition, and were able to extract a range of features from the faces they acquired [9]. In 1991, in an article named "Eigenfaces for recognition" published by Matthew Turk and Alex Pentland in the field of computer vision, enabling him to recognize the resulting features of face images by reducing the dimension in the vector space [10]. There are many algorithms used in face recognition in the field of computer vision. In 1998, Zhao et al. used the PCA (Principal Components Analysis) algorithm [11]. In 2000, Chen and his colleagues used the LDA (Linear Discriminant Analysis) algorithm [12]. In 2001 Viola and Jones used the edge extraction algorithm for face recognition [13], and in 2005 Zhang et al. used GABOR algorithm [14]. In 2006, Ahonen et al. Attempted to perform face recognition using the LBP (Local Binary Pattern) algorithm [15]. However, all algorithms were inadequate due to the effect of parameters such as exposure, light intensity and facial expressions. The field of computer vision has increased the interest in both face recognition and object recognition in the early 2000s. As a result of easy access to devices such as cameras and mobile phones, the amount of data started to increase and the face recognition studies gained speed after the article published in 2009 in the field of computer vision. Object detection and classification is seen as an important subject which has been studied since the first years when digital devices such as computers were first invented. In 2009, ImageNet data set was created by Jia Deng et al. via using more than 14 million pictures in 22 thousand categories with the contribution of more than 150 countries. It is aimed to extract objects from image data and be contributed to interpretation in this study. A competition called ILSVR (ImageNet Large Scale Visual Recognition) was organized so that people working in the field of computer vision could develop models in the area of 1000 class object detection[16]. In 2012, Krizhevsky and his friends created a model for object recognition by using convolution processes, and named AlexNet [17]. After this year, the computer vision was

introduced with CNN (Convolutional Neural Network), and now object recognition applications are developed with CNN algorithms. At the University of Oxford, 2.6 million face images were trained on the VGG (Visual Geometry Group) data set by using the CNN method, and a deep learning model with a class 2622 was created with a success rate of 98.95% [18]. It is called the VGGFace model. The group also expanded its data set in 2017 and developed a new model has 9131 class in 3.31 million face images [19]. When all information is evaluated together, it is seen that the classical algorithms are inadequate in face recognition, and the CNN algorithm used for object detection is successful in the classification of facial images.

III. PROPOSED SYSTEM

The face recognition model, which is formed by the deep learning methods from the image and video data obtained from the crime scene investigation in the field of forensic informatics, is added to the analysis and testing stage of the Forensic Evidence Examination Process. At this stage, person identification is done autonomously by making face classification. The block diagram of proposed model is shown in Figure 2.

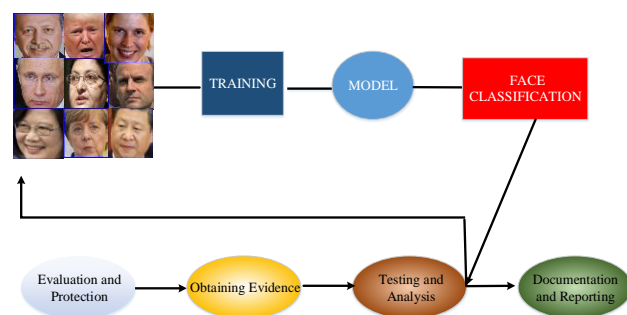


Fig. 2. Proposed System For Digital Forensic Process

This step, which was carried out by the expert investigators, has been made autonomous by the proposed system.

In this study, two different deep learning face classification methods are presented.

1. Transfer Learning Method
2. Cosine Similarity Method

For the transfer learning method on the VGGFace data set, the Resnet50 and Senet50 with the "notop" option (not including fully connection layer) ~ 98MB, while the Vgg16 model requires ~ 56MB storage. For the cosine similarity method, weight coefficient matrix of ~ 90,307MB is used to produce Resnet50 1x2048 dimension map on VGGFace2 data set. In the Vgg16 model, it uses a weight coefficient matrix of ~ 540,232MB in order to produce a 1x2622 dimension feature map. Therefore, Vgg16 model was used in Transfer Learning method and Resnet50 model was used in Cosine similarity method in order to optimize memory usage and speed parameters.

A. VGGFace Modelleri

On the VGGFace data set "Vgg16", "Resnet50" and "Senet50" deep learning models were trained. In the both different models, face images of 224x224x3 are given as input data; however, the

weight data of the model can be shaped according to the input data size we want. For each person, the VGGFace data set contains the URLs of the face coordinates and the images. The model was trained by the VGG with ~ 2.6 million face images of 2622 people.

The Vgg16 model are able to receive a minimum size of 48x48, and Resnet50 and Senet50 models data size of 197x197 for inputs. The Vgg16 model was trained with 2622 people on VGGFace's data set in 2015 [18]. The Vgg16 model uses a 38-layer softmax classifier with 2622 output. The diagram of the model is shown in Figure 3.

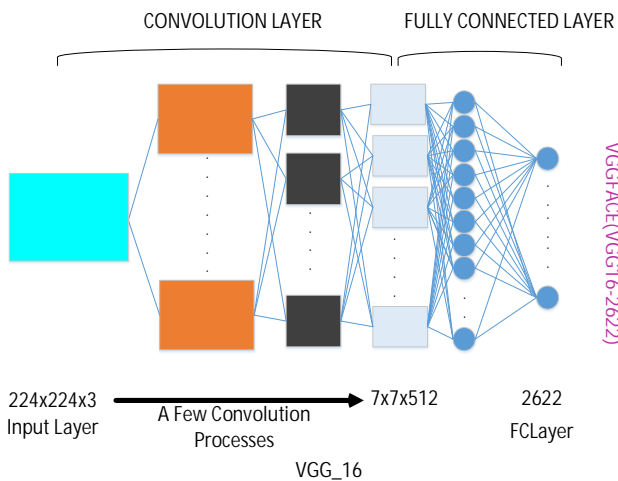


Fig. 3. The Model Structure of Vgg16 on VGGFace Dataset

Resnet50 and Senet50 models were trained on the VGGFace2 data set with 9131 people [19]. The VGGFace2 data set was published in 2017; Includes ~ 3.31 Million face image, containing 9131 people. Faces of different ages, races, genders and professions containing face images from 84-843 of each person were downloaded, and tagged via Google-Image-Search and then a VGGFace2 data set was created [19]. The model structure of Resnet50 on the VGGFace2 dataset is shown in Figure 4.

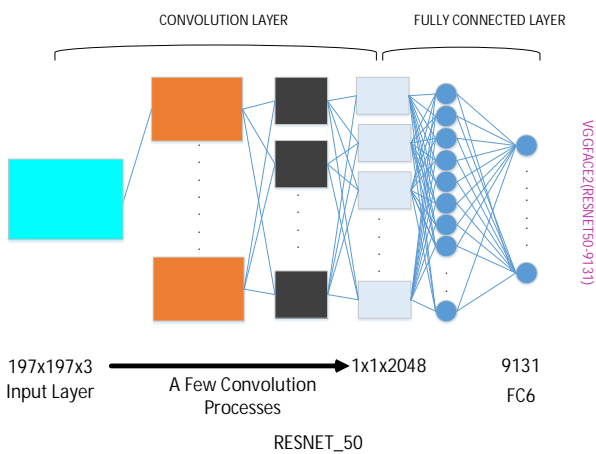


Fig. 4. Figure 4. The Model Structure of Resnet50 on VGGFace2 Dataset

Resnet50 consisted of 176 layers, and Vgg16, which consist of 38 layers, can successfully classify over one hundred images on which they are trained. Models on Imagenet represent the features of objects with 1000 different classes; however, VGGFace deep learning models uses facial data to extract the features. Therefore, Transfer-Learning and cosine similarity methods on our facial data set have been successfully applied.

B. The Transfer Learning Method

In this method; on the VGGFace data set, the Fully Connection Layer of the Vgg16 model was removed, and the classification layer with 10 outputs was added with the Transfer-Learning method, then the training was performed with this model.

Considering that millions of image data are trained for days on the GPUs that can perform parallel calculations, it is unnecessary and costly to collect millions of data and train for days to perform certain tasks. Therefore, since the data set that we will set up with 10 classes will consist of facial images, the use of the weights of the VGGFace model will enable to extract of facial features. The Transfer-Learning model is shown in Figure 5.

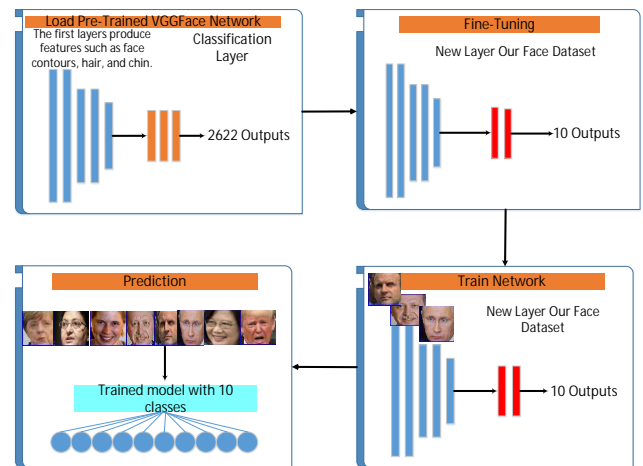


Fig. 5. The Model Structure of Transfer Learning on VGGFace

C. The Cosine Similarity Method

In this second method; on the VGGFace2 data set, the Fully Connection Layer of the RESNET50 model has been removed and a 1x2048 feature vector of each class is created for the model with 10 class. According to the feature vector of the new data, the cosine distances are estimated and their similarities are compared. Classes are determined by comparison results.

Cosine similarity is generally used in the field of computer science to determine the similarities of two documents in the vector space. In the field of text processing, it has been suggested that the document classification should be done with the cosine similarity in the studies conducted by Steinbach et al. [20]. In 2006, Mihalcea et al. demonstrated that the representation of documents in vector space can be classify and cluster by cosine similarity [21]. The mathematical expression of cosine similarity, which is still widely used in document classification in the field of deep learning, is shown by formula (1).

$$\cos(v1, v2) = \frac{v1 \cdot v2}{\|v1\| \|v2\|} \quad (1)$$

As can be seen from Figure 6, the ratio of the vector product of the V1 and V2 vectors to the scalar products will give the cosine of the angle between the two vectors.

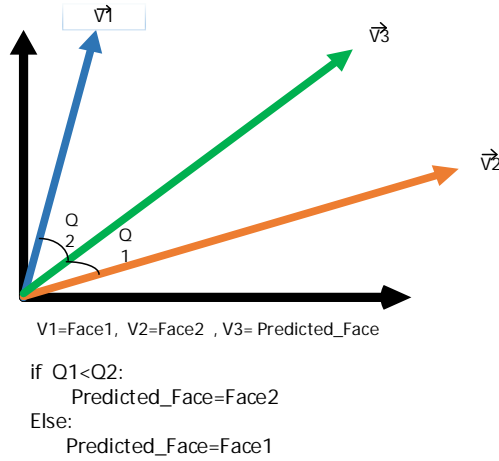


Fig. 6. Cosine Similarity in Vector Space

For vectors in the vector space of our model with 10 classes, the cosine value of the image vector is calculated for each vector to estimate the similarities. They are included in a class according the similarity values for facial classification. These operations and the structure of the method are shown in Figure 7.

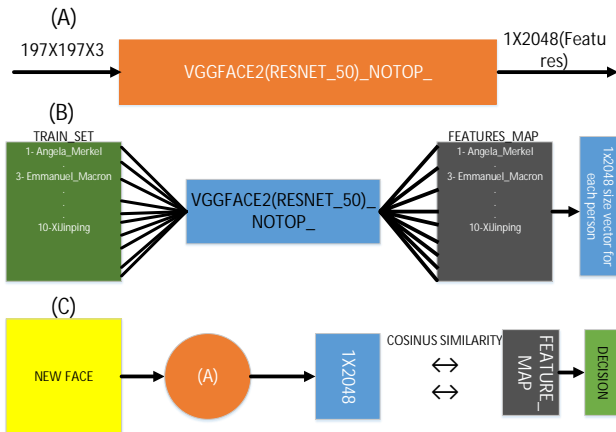


Fig. 7. Cosine Similarity Method Structure

D. Dataset

Dataset is very important in the application of deep learning methods. The difficulty of facial recognition appears to work successfully on images taken from the front face, not the different poses of the face. The reason for this is that VGGFace trained model is trained by using images trained with front face images, and updated the weights in the same way.

Dataset was created using the selenium library with python code via Google Search service as ten classes. The faces extracted for classes were searched with 10 names as search keywords from the member of the United Nations which was

randomly selected, and the first 100 image data were downloaded. Face detection process was applied on the images by using opencv library and the face images obtained were sorted and cleaned and ready for training. Sample images related to the training set in Figure 8, and the dataset properties are shown in Table 1.



Fig. 8. Dataset Image Samples

TABLE I. DATASET FEATURE

Features	Training_Set	Test_Set
Gender	6(M)-4(F)	6(M)-4(F)
Picture Per Class	~142	~71S

IV. IMPLEMENTATION

Deep learning model is realized as two different facial recognition application on VGGFace face images. In the first application, the fully connected layer was removed from the model by using the Transfer-Learning method, and a model with 10 outputs was subjected to training process.

In the second application, the Softmax classification layer in the last layer of the VGGFace model is removed and the feature map of each class sized as (1x2048) is extracted. Then the class is determined by calculating the cosine similarity of the predicted face image within the model that has ten classes.

The hardware and software used for applications are shown in Table II.

TABLE II. HARDWARE AND SOFTWARE

OS (Operating System)	Ubuntu 18.04.1 LTS
CPU (Central Processing Units)	Intel® Xeon(R) CPU E5-2640 v4 @ 2.40GHz × 20
Graphics	NV124
Operating System Type	64-bit
Programing Language	Python 3.7
RAM (Random Access Memory)	64GB

A. VGGFace Transfer-Learning

Training on the first application was realized at 686.32 second, and 100% accuracy and 0.0049 loss value were obtained. Our model has memorized all input data. Although there is an undesirable situation in deep learning studies, in the face recognition applications, the memorized model is thought to be useful to produce a better output unless there is a significant change in the face. The accuracy and loss values of the model are shown in Figure 9.

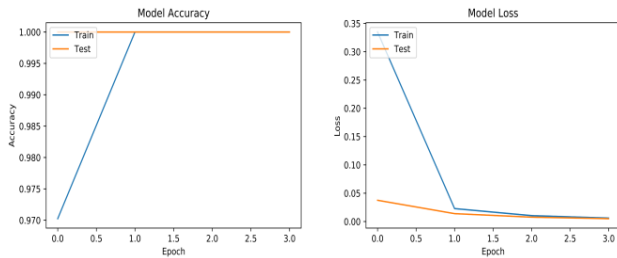


Fig. 9. Model Accuracy and Loss Values

The training of the model we have established completed successfully, and as a demo, the classification process has been successfully completed for real-time video images that are consist of 25 FPS (Frame Per Second). The classification process on the image data is shown in Figure 10.



Fig. 10. Prediction Result for Vgg16 Model

B. The Cosine Similarity of Image Vectors

For each class required to create the model, the train data set was in the first application was used. The vector average of the classes was calculated by estimating the vectors of all image data sized as (1x2048) for each class. After saving each vector information of ten classes to a file in json format, the test data used in the first application was realized on the data set. Based on the results obtained, it has 100% success by classifying all of the 721 face images on face images larger than 40x40x3 size which can be detected by opencv.

The created model file has been applied to 25 FPS video images and has achieved a successful result by classifying the images in each video frame correctly.

V. CONCLUSIONS

According to the results obtained in the applications on a single image data classification process takes time ~ 0.1 second. In terms of digital forensic investigation, on the forensic copy of a single image data, for the face detection and the classification on the image require approximately 2.5 second time.

When the existing models evaluated in terms of time parameters, a deep learning supported model that is 250 times faster is proposed in this study. As a result of the studies, we haven't realized exactly real-time face detection process on video images yet. There are partial delays. The reason of this situation can be shown that the model processed ~ 14 Million for each image data in the first application and ~ 138 Million parameters in the second application.

In future studies, it is aimed to increase the number of classes for the application and to create a real time face detection system.

(1) References

- [1] Casey, Eoghan. Digital evidence and computer crime: Forensic science, computers, and the internet. Academic press, 2011. page:7
- [2] REITH, Mark; CARR, Clint; GUNSCH, Gregg. An examination of digital forensic models. International Journal of Digital Evidence, 2002, 1.3: 1-12.
- [3] <http://www.haber7.com/yazarlar/dilek-gungor/2438824-feto-davalari-dijital-kabusa-donmesin> Son Erişim Tarihi:05.11.2018
- [4] Ballard, Dana H., and Christopher M. Brown. "Computer vision." Englewood Cliffs, NJ: Prentice-Hall. surfaces. Artificial Intelligence 17 (1982): 75-1.
- [5] ZHAO, Wenyi, et al. Face recognition: A literature survey. ACM computing surveys (CSUR), 2003, 35.4: 399-458.
- [6] Rosenblatt, Frank. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 1958, 65.6: 386.
- [7] Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors." nature 323.6088 (1986): 533.
- [8] LeCun, Yann, et al. Backpropagation applied to handwritten zip code recognition. Neural computation 1.4 (1989) 541-551.
- [9] Kirby, Michael, and Lawrence Sirovich. "Application of the Karhunen-Loeve procedure for the characterization of human faces." IEEE Transactions on Pattern analysis and Machine intelligence 12.1 (1990): 103-108.
- [10] Turk, Matthew, and Alex Pentland. "Eigenfaces for recognition." Journal of cognitive neuroscience 3.1 (1991): 71-86.
- [11] Zhao, Wenyi, et al. "Discriminant analysis of principal components for face recognition." Face Recognition. Springer, Berlin, Heidelberg, 1998. 73-85.
- [12] Chen, Li-Fen, et al. "A new LDA-based face recognition system which can solve the small sample size problem." Pattern recognition 33.10 (2000): 1713-1726.
- [13] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.
- [14] Zhang, Wenchao, et al. "Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition." Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. Vol. 1. IEEE, 2005.
- [15] Ahonen, Timo, Abdenour Hadid, and Matti Pietikainen. "Face description with local binary patterns: Application to face recognition." IEEE Transactions on Pattern Analysis & Machine Intelligence 12 (2006): 2037-2041.
- [16] DENG, Jia, et al. Imagenet: A large-scale hierarchical image database. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. Ieee, 2009. p. 248-255.
- [17] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.
- [18] Sun, Yi, et al. "Deepid3: Face recognition with very deep neural networks." arXiv preprint arXiv:1502.00873 (2015).
- [19] CAO, Qiong, et al. Vggface2: A dataset for recognising faces across pose and age. In: Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on. IEEE, 2018. p. 67-74.
- [20] Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. Introduction to Data Mining. Boston: Pearson Addison Wesley, 2006.
- [21] Mihalcea R., Courtney C., Carlo S., "Corpus-based and Knowledge-based Measures of Text Semantic Similarity", The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, July 16-20, 2006, Boston, Massachusetts, USA

Removing Redundant Primes in Multiple-Valued Logic Cover by Cube Algebra

I.SAVRAN

Karadeniz Technical University, Trabzon/Turkey, savran@ktu.edu.tr

Abstract - Multiple Valued Logic (MVL) is the generalized form of Boolean Logic. This generalization help us to generalize operations as well. Thus MVL can solve some boolean functions more efficiently. Here are some applications of MVL are optimization of finite state machines, testing and verification, design of PLA decoders. MVL also provides designing electronic circuits with more than two logic level.

MVL circuits have advantages over binary counterpart such as, it requires lower number of intercon-nections, and data density doubles in a memory cell when 4-valued logic is used. However, these potential advantages heavily depends on technology.

Formerly we shared primal results of our MVL-MIN application based on a heuristic algorithm which minimizes MVL functions. Even though MVL-MIN achieved 5x speedup over MVSIS, unfortunately, it may generate redundant prime covers. Here we propose a MVL method which guarantees a non-redundant prime cover when a prime cover is given for a MVL Function. The complexity of the new method is $|S_{pc}|^2$ where S_{pc} is the given prime cover set.

Keywords - Multiple-Valued Logic (MVL), Cube Algebra, Heuristic, MV Logic Optimization, prime cover, redundant prime.

I. INTRODUCTION

MULTIPLE-VALUED logic is the generalization of Boolean logic. In order to generalize Boolean logic, some properties must be redefined. MVL logic can encode multiple-output Boolean functions efficiently. MVL applications include optimization of finite state machines, testing and verification [1, 2].

Blif -MV is a format that is developed for encoding MVL functions and MV circuits [3]. Blif-MV has two main parts. One is header and the other contains one or more tables. In the header part, input and output variables are defined such that each variable may have its own range. A table contains a list of assignments which link the inputs to outputs.

A MVL function can be represented on an extended hypercube as well. Figure -1 depicts a 3-input 1-output MVL function.

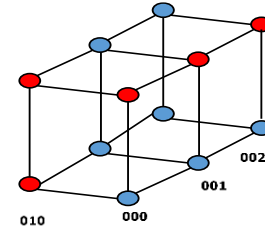


Figure 1: 3-Input 1-output MVL Function on Hypercube.

MVL-MIN reads the MVL functions from the blif-MV file and forms on-set and off-set as given below:

$$S_{on} = \{X_0^1 X_1^0 X_2^0, X_0^1 X_1^0 X_2^1, X_0^1 X_1^1 X_2^0, X_0^1 X_1^1 X_2^1, X_0^0 X_1^0 X_2^0, X_0^0 X_1^1 X_2^1\}$$

$$S_{off} = \{X_0^0 X_1^1 X_2^0, X_0^0 X_1^1 X_2^1, X_0^0 X_1^0 X_2^2, X_0^0 X_1^0 X_2^1, X_0^0 X_1^0 X_2^2, X_0^1 X_1^1 X_2^1\}$$

MVL-MIN tries to find out a list of primes that cover for a given on-set cube at a time [4]. Unfortunately MVL-MIN may generate redundant cover. A redundant cover is a cover that contains unnecessary primes in it.

II. METHOD

In order to compute non-redundant cover, we can use “Non-disjoint Sharp Operator” iteratively. The proposed algorithm (1) accepts a cover as its parameter. When we run the method, all primes are evaluated one by one. While any variable of a prime is removed, it is labeled with ‘s’. At the end of the operation when all the variables are labelled with ‘s’ that means the target prime is redundant. If any prime is redundant then it is removed from the prime cover. The result of the algorithm assures that we have a non-redundant prime cover.

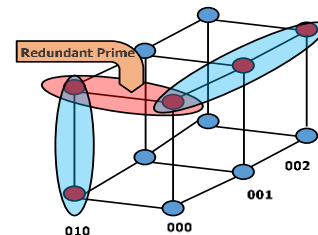


Figure 2: Redundant Cover of the given MVL Function

For example, let us remove redundant prime in the following prime cover set. Let $(S_{pc}=X_0^1 X_1^* X_2^0, X_0^* X_1^1 X_2^0, X_0^1 X_1^0 X_2^*)$ be the cover and we show how the method removes a redundant cube and keeps the essential ones.

```

    Input :  $S_{pc} = \{p_0, p_1, \dots, p_{m-1}\}$ 
    For  $i = 0 \dots |S_{pc}| - 1$  do
         $S_{pc} = S_{pc} / p_i$ 
        For  $j = 0 \dots |S_{pc}|$  do
             $p_i = p_i \# p_j$  //where  $p_j \in S_{pc}$ 
        Endfor
        If  $p_i \neq \emptyset$  then
             $S_{pc} = S_{pc} \cup p_i$ 
        Endfor
    
```

Algorithm 1: Remove redundant cubes from the Prime Cover Set.

Sample-1:

$X_0^1 X_1^* X_2^0 \# (X_0^* X_1^1 X_2^0 \# (X_0^1 X_1^0 X_2^*))$
 $X_0^s X_1^0 X_2^s \# (X_0^1 X_1^0 X_2^*)$
 $X_0^s X_1^s X_2^s$ Removed

Sampe-2:

$X_0^* X_1^1 X_2^0 \# (X_0^1 X_1^* X_2^0 \# (X_0^1 X_1^0 X_2^*))$
 $X_0^0 X_1^s X_2^s \# (X_0^1 X_1^0 X_2^*)$
 $X_0^0 X_1^s X_2^s$ Not Removed

The complexity of the proposed method is $|S_{pc}|^2$. We may improve the algorithm by considering the essential primes. Since essential primes are necessary to be in a cover, the complexity of the proposed algorithm decreases.

III. RESULTS

MVL-MIN is a heuristic MVL minimization algorithm. Since MVL-MIN may produce redundand primes that need to be removed. Therefore we presented a new method that produces a non-redundant cover for a given MVL function. It assures that the final prime cover is non-redundant. The complexity of the proposed method is as low as $|S_{pc}|^2$ where S_{pc} denotes a given redundant prime cover set.

REFERENCES

- [1] E. Dubrova, *Multiple-Valued Logic Synthesis and Optimization*. In: Hassoun S., Sasao T. (eds) Logic Synthesis and Verification. The Springer International Series in Engineering and Computer Science, vol 654. Springer, Boston, MA.
- [2] R. K. Brayton, M. Chiodo, R. Hojati, T. Kam, K. Kodandapani, R. P. Kurshan, S. Malik, A. L. Sangiovanni-Vincentelli, E. M. Sentovich, T. Shiple, K. J. Singh, and H.Y. Wang. *BLIF-MV: An Interchange Format for Design Verification and Synthesis*. Technical Report UCB/ERL M91/97, Electronics Research Lab, Univ. of California, Berkeley, CA 94720, Nov. 1991.
- [3] MVSIS (Version 3.0), University of Berkeley, <https://ptolemy.berkeley.edu/projects/embedded/mvsis>.
- [4] I. Savran, *MVL-MIN: A New Heuristic Minimization Tool for Multiple-Valued Logic Functions*. 18th IIE International Conference on Latest Trends in Eng. and Tech. (ICLTET-2018), Istanbul, Turkey, March 21-23 2018.

Detecting Point Mutations in Next-generation Sequencing Data

I.SAVRAN¹, M.E. ERCİN², F.C. EYÜPOĞLU³

¹ Karadeniz Technical University, Trabzon/Turkey, savran@ktu.edu.tr

² Karadeniz Technical University, Trabzon/Turkey, drmustafaemreercin@ktu.edu.tr

³ Karadeniz Technical University, Trabzon/Turkey, fcelep@ktu.edu.tr

Abstract - The improvement of next-generation sequencing (NGS) boosts medical and genomic research. NGS can now produce billions of short genome fragments. Thus, in a single run, a few terabytes of data is produced. A wide range of applications analysis processes this massive datasets. Moreover, the high-throughput feature of NGS decreases the sequencing cost. However, the NGS data is noisy due to faulty readings. There are typically 1-3% read errors. The design of big data algorithms is necessary for handling this non-clear huge amount of data.

A single nucleotide alteration in a sequence of genetic materials is called point mutation. There are mainly three point mutations (a) substitution (b) insertion (c) deletion. There are many genetic diseases such as Familial Mediterranean Fever (FMF) could be diagnosed earlier. FMF is an autosomal recessive disease caused by mutations in the FMF gene (*MEFV*) These mutations effects protein production. Once these mutations are determined, consequences are moderately predictable. However it is so difficult to extract point mutations in noisy NGS data. Since read errors are very similar to point mutations, we propose a scalable three-stage method to extract point mutations: (a) fix misreads in sequences, (b) categorize the sequences based on common shared substrings and finally (c) detect the point mutations. These procedures works based on pipeline. In order to save space and time, Bloom filter is determined as the main data structure. It is simple to implement and versatile to use in all the stage of the point mutation process.

Keywords - Next-generation Sequencing, Big Data, Familial Mediterranean Fever (FMF), *MEFV* gene, point mutation.

I. INTRODUCTION

MISREADS are mostly single base read errors during sequencing phase by NGS. Based on NGS technology, the types of read errors and error rates can be vary. Since obtained sequences are as short as 30 base-pairs (bp), the size of sequencing data of Human genome reaches up to billions of short sequences [1, 2].

II. BLOOM FILTER

The Bloom filter is a data structure which has probabilistic search feature. The base data structure of a Bloom filter is a *Bit Vector*. In order to add an element to a Bloom filter, the element is processed with a number of hash functions. Hash functions operate on input parameter and generate indices. Then the

indexed bits are set in the bit vector. The registering process continues until there is no token left.

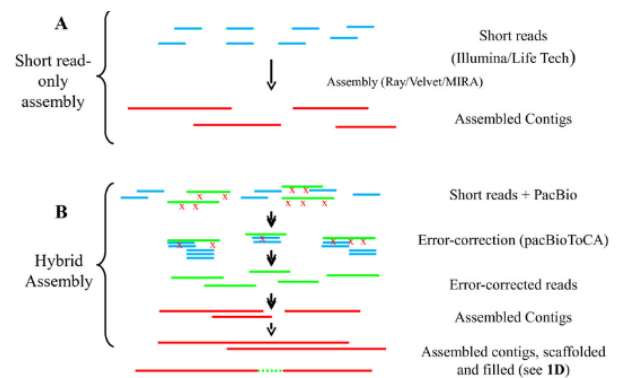


Figure 1: NGS Sequencing Technologies [2].

As an example, there is a 15-bit long bloom filter with two hash functions in Figure-2. In order to insert a “Hello” and “world” tokens into the filter, hash functions processes the first token and generate indices 14 and 2 respectively. Then the indexed bits are set. When the whole operation completed, we can use this bloom filter to check whether an unknown substring is encoded in it [3].

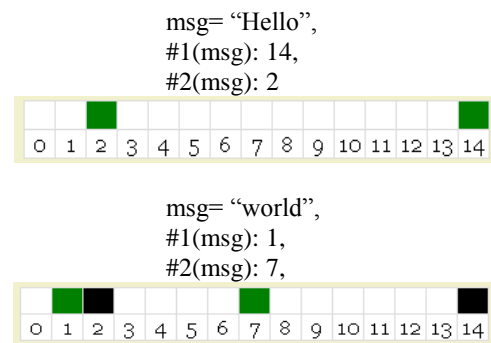


Figure 1: Bloom Filter encoding with hash function.

Designing a Bloom filter is dependent on some parameters. These parameters are the number of tokens (n), bloom filter size (m -bit) and the number of hash functions (k). The success of the bloom filter is dependent on these parameters. Thus correlation of these arguments specify the false positive rate (p).

III. *MEVF* GENE

Mediterranean fever gene (*MEFV*) is a human gene that provides instructions for making a protein called pyrin. Pyrin is produced in some white blood cells that play a role in inflammation and in fighting infection. The *MEFV* gene is located on chromosome 16p13.3 [4].

Until now, about 80 *MEFV* mutations are recorded. There are a few number of delete mutations on DNA from the *MEFV* gene, which can lead to an abnormally small, nonfunctional protein. Most *MEFV* gene mutations, change one of the amino acids used to make pyrin. The most common mutation replaces the amino acid methionine with the amino acid valine at protein position 694 (written as Met694Val or M694V). On the DNA, this mutation is encoded as "...A(M694V A>G)..." which indicates Adenin-Guanin alteration [5].

IV. METHOD

In order to speed up searching a specific pattern in big data, a practical data structure is necessary. Our method is based on encoding the known *MEFV* mutations on a bloom filter tape by a constant string of bases called "a window". The window size 'w' can be determined by experts. We need another auxiliary data structure to specify where the detected pattern is located on the gene.

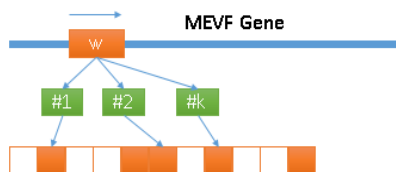


Figure 3: Encoding and detecting *MEVF* mutations. (a) Encoding *MEVF* mutations into Bloom Filter.

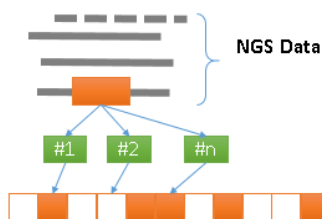


Figure 4: Searching whether NGS sequences has similar token.

Figure 3 explains the encoding the tokens of length w where each w contains a recorded mutation on *MEFV* gene. Encoding is an initial process and it is not a complex operation. The searching operation is the hard part because there are millions of NGS sequences to be evaluated (see Figure 4).

REFERENCES

- [1] B. Schmidt, A. Hildebrandt, *Next-generation sequencing: big data meets high performance computing* Drug Discovery Today, Vol 22, Issue Pages 712-717, 4, April 2017.
- [2] J. Powers, V. Weigman, J. Shu, M.J. Pufky, D. Cox, P. Hurban, *Efficient and accurate whole genome assembly and methylome profiling of E. coli.* BMC genomics. 14. 675. 10.1186/1471-2164-14-675.
- [3] Bloom Filters: <https://lilimlib.github.io/bloomfilter-tutorial/> (date: 21/02/2019).
- [4] *MEFV* gene, pyrin innate immunity regulator, <https://ghr.nlm.nih.gov/gene/MEFV> (date: 13/02/2019).
- [5] F. Salehzadeh and A. Fathi, *Patient with FMF and Triple MEFV Gene Mutations.* Medical archives (Sarajevo, Bosnia and Herzegovina), 69(4), 269-270. doi:10.5455/medarh.2015.69.269-270.

IP Packet Marking and Forwarding Based on Content Type with SDN

Mustafa Atalay¹ and Mehmet Demirci¹

¹Gazi University, Ankara/Turkey, atalaymst@gmail.com

¹Gazi University, Ankara/Turkey, mdemirci@gazi.edu.tr

Abstract – Software defined networking (SDN) enables various applications thanks to the flexibility and fine-grained control it provides. Efficient packet forwarding based on the type of content is an important capability for emerging technologies such as Internet of Things (IoT) and 5G, which benefit considerably from SDN. In such systems, numerous devices send many different types of data which may need to be processed on different specialized servers. Devices or hosts cannot be required to know about target servers for the data to be forwarded correctly. The purpose of this study is to facilitate inter-host network communication in such a setting and to direct different types of data to the relevant targets through the network using SDN. The proposed solution involves a packet marking methodology to indicate the type of content and utilizes these marks to forward packets to the appropriate hosts in the network using the Ryu controller. We evaluate our solution with both TCP and UDP on a network topology created on Mininet. We also considered ways of reducing the data transmission time.

Keywords – SDN, Ryu controller, IPv4 packet marking, IoT

I. INTRODUCTION

As the usage of mobile devices and Internet of Things (IoT) increases day by day, the amount of data collected by these devices skyrockets. The increase in the amount of data collected will also increase the workload of the hosts which will process the collected data. Considering this situation, the idea of processing different types of data on different hosts and customizing each host to process a data type has emerged. Thus, the hosts can be configured according to the characteristics of the data they will process. If sensors collecting different types of data are connected to a host, the host will have to send the data which is not the type of data it has to process to the corresponding hosts. This raises the problem of correct and efficient inter-host data transmission. In addition, the host does not necessarily have the target host information to send every type of data.

Software-defined networking (SDN) is a popular trend providing more flexible network management by separating the data layer and control layer. SDN can be leveraged for forwarding data based on various packet header fields to serve different purposes such as load balancing, filtering, cost minimization, security improvements etc. The goal of this study is to utilize SDN to facilitate inter-host network communication in an IoT setting and to direct different types of data to relevant

targets efficiently. We devise a packet marking methodology to label packets according to their content type and use these labels to forward packets to the required destination hosts with the Ryu SDN controller. We exhibit the feasibility and performance of our approach with simulations on Mininet [3].

The rest of this paper is organized as follows. Section II provides background information on the research topic, tools and techniques. Section III presents a review of related works about packet marking in SDN. Section IV describes the proposed packet forwarding approach and evaluation results. Finally, Section V summarizes and concludes the paper.

II. BACKGROUND, TOOLS AND TECHNIQUES

A. IoT

The Internet of Things (IoT) is a system of interconnected objects with data collection and/or processing capabilities. IoT devices need to have the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction [1]. Number of these devices are increasing day by day. Billions of IoT devices generate a massive amount of data, requiring efficient transmission and fast processing techniques [2]. Distributed data processing approach is at the heart of the proposed solutions for this problem. In this study, we designed a network infrastructure with SDN for data distribution.

B. SDN

SDN is an important paradigm in the current networking landscape and an enabler for the power of software in managing networks. SDN provides the ability to program a network from a central location via a controller software, separating network control tasks from data forwarding. It offers many advantages including easier management, faster response, more flexible forwarding actions, and reduced cost. Thus by using the facilities provided by SDN, we performed marked IP packet routing in the network with Ryu SDN controller.

C. Mininet

Mininet is an emulation environment for creating a realistic virtual network running real code on a single machine. It can create the desired network in seconds with a single command. Mininet is a commonly used tool for developing and experimenting with OpenFlow and SDN systems [3]. In this study, we created the entire network infrastructure on Mininet. This involves the creation of topology and host processes where IP packet creation, sending and receiving are performed.

D. Internet2 Network

Internet2 is a network and associated technical community founded in 1996 by several leading US universities [4]. It is utilized mainly as a testbed infrastructure for research and education. This network infrastructure has been used in many computer networking studies in the literature. We created our topology for this study based on the latest Internet2 network topology graph.

E. Netcat

Netcat is a computer networking utility for transmitting and receiving packets using TCP or UDP. It can be used directly as a command or driven by other programs and scripts. Additionally, it is useful for network debugging and investigation as it supports many types of connections with a number of built-in capabilities [5]. We used this tool to send packets from the hosts to the network and receive the incoming packets.

F. IPv4 packet structure

IP protocol is the main backbone protocol of the internet. On the other hand it is one of the main protocol inside the TCP/IP protocol. The IP protocol header is shown in Figure 1 [6].

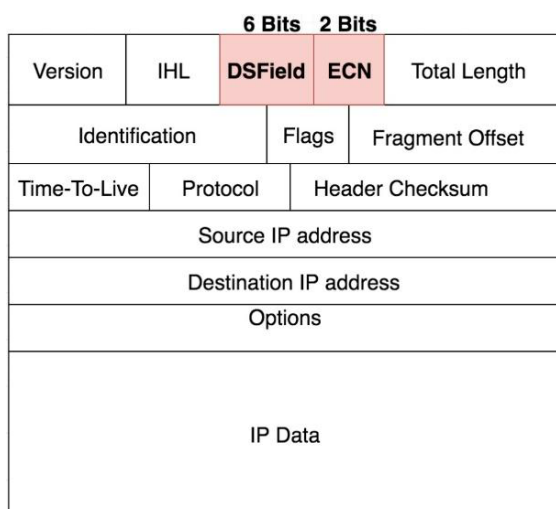


Figure 1: IP Packet structure [6]

The TOS (Type of Service) header field, which is 6-bit DSField and 2-bit ECN, consists of 8 bits. Although the TOS bits are used for QoS, these bit fields are usually ignored depending on the service provider. We used the 8-bit field for IP packet marking.

G. Spanning Tree

Spanning Tree Protocol (STP) is an important protocol in legacy layer 2 switches. By using STP, a legacy network can form a spanning tree topology without any loops, and the network could avoid broadcast flooding problem [7]. After network structure creation, we encountered broadcast flooding problem, because there are several closed loops in our topology. For this reason, we applied this protocol to our network infrastructure.

III. RELATED WORK

The process of routing packets to a different host by modifying the destination IP address with SDN controllers is generally used in the SDN load balancer operations. In load balancing, routing is performed according to the load status of the target servers. The studies reviewed in this section differ in terms of the criteria to be considered, the SDN controller used, the network topology examined, and the routing algorithm used in the network.

Prakash and Deepalakshmi [8] used Floodlight as the OpenFlow controller. Target servers are selected according to the load status of the target servers. They performed their work by connecting 2 servers and 4 clients to a network topology where 2 switches are connected linearly. Because the network topology consisted of 2 switches and no loops, they did not need a routing algorithm.

In another study, Hamed et al. [9] used Ryu as the OpenFlow controller. Target server selection was made based on bandwidth. The authors created their network topology consisting of 1 client and 2 servers. Since there is only one switch in the network, the routing is done manually to the ports to which the hosts are connected.

Kaur et al. [10] used POX as OpenFlow controller. Target server selection was made based on Round-Robin. They have demonstrated their work on a network topology consisting of 6 clients and 3 servers connected to 1 switch. Since there is only one switch in the network, routing is done manually to the ports to which the hosts are connected.

All of the above-mentioned studies were conducted using the Mininet network tool.

IV. MARKED PACKET FORWARDING WITH SDN RYU CONTROLLER

In this study, we first created a network with a specific network topology in the Mininet environment. After that, 4 hosts were connected to the created network. In theory, it is assumed that there are several IoT devices connected to each of the 4 hosts. Each IoT device writes the sensor data to the hosts it is connected to. Each host creates an IPv4 packet for each collected piece of data and sends the packet to the network with a predefined destination IP address. A predefined destination IP is specified because we assume the hosts do not have any information about the destination host. While data to be processed at the current host based on its type are not sent out, the data with the other types are sent to the network to be routed to the related host.

In order to carry out the above procedure, we created our control program using the Ryu SDN controller. The controller determines the destination host to which the packet will be transmitted, by looking at the TOS field of the IPv4 packets arriving in the network. After the destination host is determined, the destination IP address of the IPv4 packet is set to the determined IP address and the packet is sent out. Finally, the destination host receives the incoming IPv4 packet and stores the sensor data in a file.

A. Creation of Network Topology

We have designed a network topology based on Internet2 Network Infrastructure Topology [11]. Firstly, all non-branching intermediate nodes were removed from the topology. The updated form of the Internet2 Network Infrastructure Topology is shown in Figure 2.



Figure 2: Internet2 Network Infrastructure Topology [11]
(reorganized by the authors)

The network topology on this updated map was transferred to the Mininet environment with 15 switches and 24 links that connect these switches. All switches in this topology are configured to support the OpenFlow protocol version 1.3. We connected host h1 to node s1 (Seattle), host h2 to node s15 (Denver), host h3 to node s10 (Houston) and host h4 to node s6 (New York). The network topology including the connected hosts in the Mininet environment is shown in Figure 3.

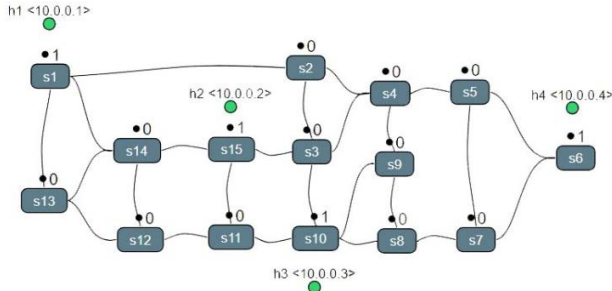


Figure 3: The network topology with 15 switches and 24 links

B. Marking and Transmitting IPv4 Packets

Creating Marked Packets

Each of the 4 hosts connected to the created network was designed as both client and server. In each host, we defined a file named "InData", which contains data coming from the other hosts and the "OutData", which contains the data to be sent to the other hosts. We considered the "OutData" file as a file in which different types of sensor data collected from IoT devices connected to the host are stored. Data format written in the "OutData" file is as follows: the first column is the data type, the second column is the data itself, and the third column is the destination port. Sample contents of these two files are shown in Figure 4.

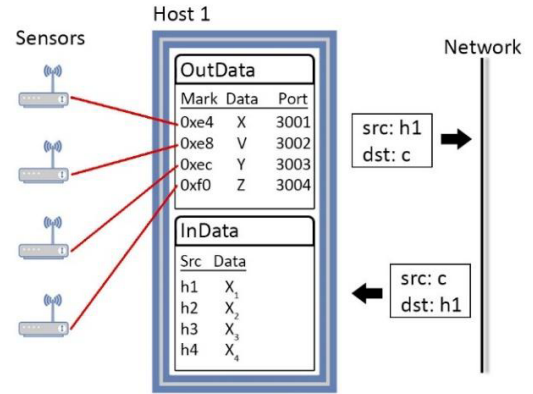


Figure 4: Creation of Marked Packets

The data type in the first column consists of an 8-bit hexadecimal number. In this study, we set the number of categories to 4 and set their hexadecimal equivalents to 0xe4, 0xe8, 0xec, 0xf0. The second column consists of raw sensor data and the last column consists of a destination host port number.

Transmitting and Receiving IPv4 Packets

We created a script to read the contents of the "OutData" files in each of the 4 hosts. This script reads the file and creates an IPv4 packet for each line of "OutData" with the Netcat tool. The flow diagram of these processes is shown in Figure 5.

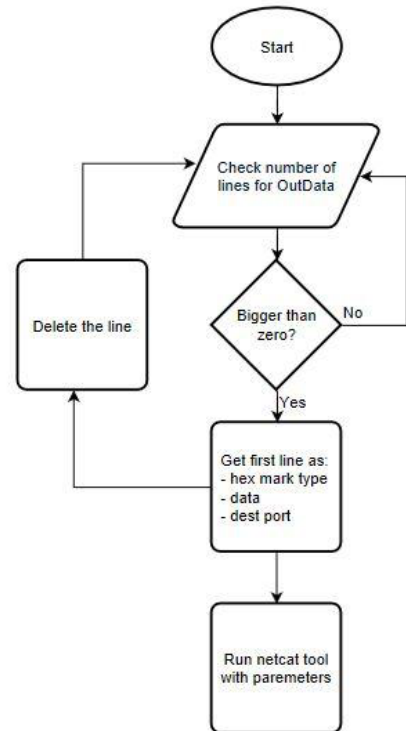


Figure 5: Sensor Data Transfer

The Netcat tool expects TOS bits, the data itself and the destination port information as the parameters. The script sets the TOS bits to the first column, the data to the second column, and the destination port to the third column of the "OutData"

file, and then runs the Netcat tool. Since the hosts do not have any information other than the data type contained in "OutData", a pre-defined IP address is entered in the Netcat as the destination IP address. All hosts send the packets to the predefined IP address via Netcat tool. This operations are applied to all lines in the "OutData" file. If data types to be processed on the current host are found in "OutData", this data is also written directly to the "InData" file on that host. The flow diagram where all operations are defined is shown in Figure 6.

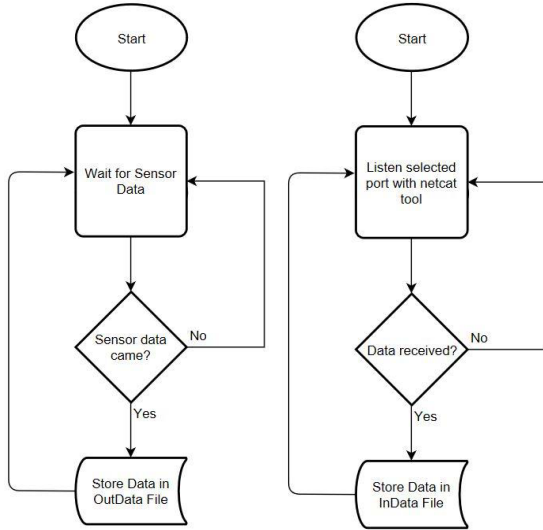


Figure 6: Sensor Data Storage

Figure 7: IP Packet Storage

In addition, we created another script that listens to the predefined ports of each host with the Netcat tool. This script constantly listens to the predefined port and saves it to the "InData" file when data arrives. The flow diagram of these operations is shown in Figure 7.

C. SDN Controller Operations

In order for a host to send an IPv4 packet prepared by the above method to a predefined IP address, this host first needs the MAC address of the destination host. For this purpose, the current host sends an ARP request for a predefined IP address to the network. This ARP request is replied by the SDN controller and the host sends the IPv4 packet to the network with a predefined MAC address which was sent by the controller as ARP reply. When the IPv4 packet arrives at a switch, the controller checks the TOS bits field of this IPv4 packet and determines the destination host to which the packet is to be sent. After the destination IP address has been determined, the controller modifies the destination IP address field of the packet and forwards the packet. At the same time, a rule is also created for the response packet that will be returned from the destination host. When the reply packet arrives at the current switch, the controller updates the source IP address of the reply packet with the predefined IP address and forwards the packet to the source host. In this way, the response packet that the source host expects from the pre-defined IP address is received. As a result, the trust relationship between source and destination host is ensured. The overall process specified above is shown in Figure 8.

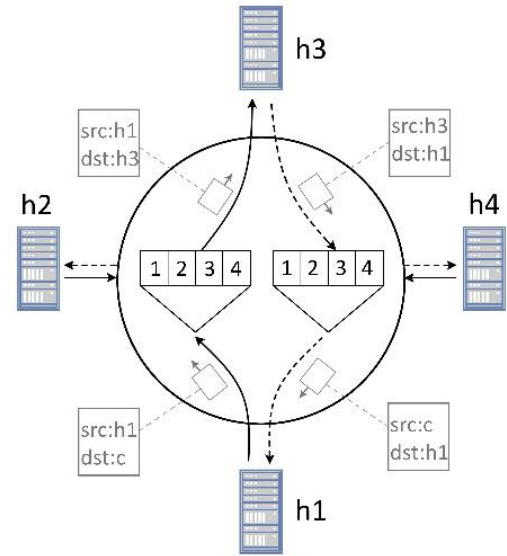


Figure 8: Transmitting and Receiving IPv4 Packets

D. Recommended Window Size Method and Test Operations

Since IoT devices can also collect real-time data, we have also carried out the above steps by UDP in addition to TCP. To this end, we updated the IPv4 packet sender and receiver scripts. In its new form, the Netcat tool generates IPv4 packets from the sensor data collected in the hosts and marks the TOS bit area of the packet according to the type of data, after which this packet is sent to the network via UDP. On the destination host side, these packets are received through UDP and stored on the host. The conducted experiments and their results are given below.

In this part, we proposed a method to increase the efficiency of the transmission of data coming from the sensors. According to our method, the host sends data to the network after the amount of sensor data with the same type has reached a certain number. If the amount of data for any data type reaches the number we set, IPv4 packets labeled with the appropriate TOS bits are generated for this data group and sent to the network sequentially by TCP or UDP. The purpose of this method is to shorten the transmission time by reducing the number of TCP connections established for the transmission of the data. On the other hand, this also aims to reduce the amount of lost packets sent by UDP. Table 1 summarizes the test results, which are explained further below.

Table 1: Packet transmission statistics with TCP and UDP

	TCP		UDP	
	With WS	Without WS	With WS	Without WS
Packet loss count per connection	0	0	1	1
Packet transmission per second	29.4	29.8	129.2	0

The test was primarily done with TCP. The data we sent via TCP did not exhibit any packet loss, regardless of whether we specified any window sizes (WS). It was also observed that the window size had no effect on the number of packets sent per second. In fact, there was a loss in yield due to the calculations of window size operations. We performed the same test steps with UDP. We observed that the first packet was lost, regardless of whether we determined the window size or not. In the scenario where we did not even specify a window size, if more than one data packet of the same type were available, only packets after the first could be transmitted. If there were not two consecutive packets of the same type, the packet loss reached 100%. One of the most important results of these tests was the one in which we set the window size with UDP. When we set a window size and send packets with UDP, the transmission speed of other packets in the window after the first packet loss increased substantially.

V. CONCLUSION

In this study, we have developed an SDN based method for routing large amounts of data collected by IoT devices to the corresponding distributed hosts where this data is processed. We began by simulating the Internet2 Network Infrastructure Topology in the Mininet environment. We then connected 4 hosts to this network and assumed that each host was collecting data from the sensors connected to it. After this assumption, we designed a method so that these hosts can send each type of data to the correct host in the network. In its essence, we did this by modifying the target information of IP packets according to the marked TOS bits with the Ryu SDN controller. When the number of packets of a particular type reaches a predefined window size, these packets are sent to the network. We observed a significant increase in the number of packets sent per second in the test we performed with UDP compared to TCP. The proposed method can be used for content based packet forwarding with SDN to enable efficient and real-time data processing in IoT systems.

REFERENCES

- [1] Available: <https://www.iotforall.com/what-is-iot-simple-explanation/> Accessed on 27 March 2019
- [2] F. Y. Okay, S. Ozdemir. "Routing in Fog-Enabled IoT Platforms: A Survey and an SDN-Based Solution." *IEEE Internet of Things Journal* 5.6: 4871-4889, 2018.
- [3] Mininet: An Instant Virtual Network on your Laptop (or other PC) Accessed on 27 March 2019, Available: <http://mininet.org/>.
- [4] The Internet2, Accessed on 27 March 2019, Available: <https://www.internet2.edu/about-us/>
- [5] Netcat, Accessed on 27 March 2019, Available: <https://en.wikipedia.org/wiki/Netcat>
- [6] Available: <https://www.slashroot.in/understanding-differentiated-services-tos-field-internet-protocol-header> Accessed on 27 March 2019
- [7] P. W. Chi, M. H. Wang, J. W. Guo, C. L. Lei. "SDN Migration: An Efficient Approach to Integrate OpenFlow Networks with STP-Enabled Networks", In 2016 International Computer Symposium (ICS) (pp. 148-153), IEEE, December, 2016.
- [8] S. W. Prakash, S. W., P. Deepalakshmi, "Server-based Dynamic Load Balancing", In 2017 International Conference on Networks & Advances in Computational Technologies (NetACT)(pp. 25-28), IEEE, July 2017.
- [9] M. I. Hamed, B. M. ElHalawany, M. M. Fouda, A.S.T. Eldien, "A new approach for server-based load balancing using software-defined networking", In 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS) (pp. 30-35). IEEE. December, 2017.
- [10] S. Kaur, K. Kumar, J. Singh, N. S. Ghumman, N. S., "Round-robin based load balancing in Software Defined Networking", In 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 2136-2139), IEEE. March 2015.
- [11] Available: <https://www.internet2.edu/media/medialibrary/2018/07/16/Internet2-Network-Infrastructure-Topology-All-legendtitle.pdf> Accessed on 27 March 2019

NetCar: A Testbed for Mobile Sensor Networks

R. B. KORKMAZ¹, K. KIZILIRMAK¹, and İ. F. ŞENTÜRK¹

¹ Bursa Technical University, Bursa/Turkey, ramazan.burak.korkmaz@gmail.com

¹ Bursa Technical University, Bursa/Turkey, kevserkzlrnk@gmail.com

¹ Bursa Technical University, Bursa/Turkey, izzet.senturk@btu.edu.tr

Abstract - Mobile sensor networks (MSNs) enable monitoring inhospitable areas without risking the human life in an autonomous manner. However, due to limited energy and exposure to harsh environmental conditions, the network can be subject to node failures. Depending on the network topology, location of the failure and its scope, the network can be partitioned into multiple isolated segments. Network partitioning can have catastrophic effects on the application due to the lack of data exchange and coordination among the partitions. The coverage declines and the fidelity of the collected data deteriorates. Considering the limited intervention to the application area, a self-configuring self-healing solution is required to restore connectivity in the network. An intuitive approach is exploiting mobility to restructure the network topology and link the partitions. In this paper, we present a mobile platform, named NetCar, which enables controlled mobility to facilitate MSNs. We discuss the design of NetCar and evaluate its performance in terms of energy consumption in various terrain features and levels. NetCar provides the basic building block to enable MSNs and can be used as a testbed to evaluate connectivity restoration algorithms for MSNs.

Keywords - Mobile sensor network, testbed, robot, connectivity restoration, energy consumption.

I. INTRODUCTION

THE recent advances in microelectromechanical systems (MEMs) has paved the way for low-cost low-power sensors coupled with wireless communication capabilities. By employing several sensor nodes forming a mobile sensor network (MSN), the phenomena can be monitored at a larger scale [1]. Despite availability of energy harvesting [2], most sensors are equipped with limited batteries that deem energy conservation. Therefore, typically, low-power wireless communication techniques are employed such as 6LoWPAN [3] that provides IPv6 connectivity over IEEE 802.15.4 based networks. Due to the limited transmission range of the facilitated radio technology, sensors form a multi-hop network. In the network, each sensor is regarded as a node. To connect the network with the rest of the world (e.g., cloud), a base station (BS) is employed. BS is less restricted in terms of energy and communication range and acts as a gateway between the network and the remote user. Since only a few nodes are expected to be located within the communication range of the BS, network-wide collaboration is required to deliver the sampled data.

Considering the limited on-board batteries and the tendency of deployment in areas with inhospitable surroundings [4], the network can be subject to node failures. Redundancy may compensate some node failures. However, network coverage will decline if the point of interest is monitored by the failed nodes exclusively. Furthermore, failure of the cut-vertex nodes will partition the network into segments isolated from each other. In such a scenario, only the nodes within the partition of the BS can deliver their data. The data collected by the rest of the nodes in other partitions cannot be delivered to the BS even if the nodes are still functional. Consequently, coverage declines drastically and the fidelity of the collected data degrades.

Depending on the application (e.g., battlefield surveillance, volcano monitoring, etc.), intervention to the application area may not be feasible. Thus, a self-healing solution is required. Mobility can be employed in various forms (e.g., random, controlled, etc.) to improve some of the network performance metrics including coverage [5] and network lifetime [6]. In this paper, we are particularly interested in controlled movement where the nodes can be relocated in a controlled manner. Controlled mobility can be exploited by attaching sensor nodes on mobile robots [7]. In a reactive manner, the node can control the mobile platform and move on-demand.

Mechanical movement incurs excessive energy cost [8] and therefore, mobility should be applied in a controlled manner and must be avoided unless it is required. In this paper, we assume the employment of controlled mobility to restore the network connectivity in partitioned mobile sensor networks. Various mobility-based connectivity restoration schemes are available [9]. One approach is moving the partition as a block towards another partition to link them. Note that the mobility of a single node may link the partitions. Thus, block movement is expected to cause redundant movement. Another approach is limiting the movement to a single leader node that sustains recovery with a single relocation at each step. If the relocation of the leader is not sufficient for recovery, a new leader is selected and the recovery proceeds in an iterative manner. Another approach is employing a mobile data collector (MDC) that visits partitions to collect their data and forward it to the BS. The mobile platform that we present in this paper is independent from the applied recovery scheme and can be used in sensor networks to enable controlled mobility. NetCar can be used as a testbed to evaluate the employed connectivity restoration algorithms. In this paper, we discuss the design details of the mobile platform and

present its energy consumption characteristics in various scenarios. Note that the application area where the nodes are deployed may be formed of varying terrain types and elevation levels. Depending on the terrain type, friction force applied to the mobile node will not be the same [10]. Therefore, energy consumption must be evaluated according to the considered terrain features. In this paper, we have considered three different terrain types, namely marble, grass, and concrete roads. Marble road represents an indoor environment while grassy and concrete roads are outdoor environments. In order to investigate the relationship between the movement direction and the mobility cost, we have also considered movement on a curved path. Finally, to assess the impact of the road gradient on the energy consumption, we have varied the gradient angle of the road and applied mobility both uphill and downhill.

The rest of the paper is organized as follows. The design of NetCar is discussed in Section II. Energy consumption model of NetCar is evaluated in Section III. The paper is concluded in Section IV.

II. DESIGN

NetCar is a custom made mobile platform to enable controlled mobility in sensor networks. NetCar can also be used as a testbed to evaluate mobility scenarios in mobile sensor networks. The current version of NetCar is the preliminary design and supports basic mobility features. The controller is based on Arduino that provides flexibility to enhance its features in the future.

A. Chassis

The chassis of NetCar is commercially available [11]. 4WD all-terrain mobile platform enables deployment in areas with harsh terrain types. We have performed experiments on various terrain types including marble, grass, concrete, and wood. We have also varied the road gradient up to 20 degrees. It should be noted that the gradient can be increased even further but we restricted the experiments due to the limited range of the road platform. Some of the specifications of the chassis can be found in Table 1.

Table 1: Specifications of the chassis [11].

Wheel diameter	120 mm
Wheel width	60 mm
Platform length	195 mm
Platform width	140 mm
Length of body	270 mm
Width of body	280 mm
Platform height	120 mm
Platform weight	1280 g
Ground clearance	26 mm

B. Controller

The control unit is connected with different modules, sensors, and actuators. Some of the key responsibilities of the

control unit are sensing the environment, performing calculations, and controlling the actuators (e.g., motors). We have employed Arduino [12], an open-source electronics platform, as the control unit. Arduino boards consist of an Atmel AVR microcontroller with various amounts of memory, pins, and features based on the model [13]. The boards have digital and analog pins to supply voltage and receive data. By changing the state of the pin, Arduino controls integrated circuits. Analog pins can convert analog voltage to digital data by using an ADC. The number of bits used in ADC determines the precision of the conversion. NetCar employs Arduino Mega 2560 as the control unit. This board has a 10-bit ADC that is sufficient to perform basic conversions. The board has 54 digital I/O pins. 14 of these pins are capable of generating PWM signals. The board also has 16 analog pins and a 16 MHz crystal oscillator with 4 KB EEPROM. The popularity of the Arduino platform with a wider community, which delivers several open-source projects every year, was the decisive factor in selecting the platform. The number of available pins and the cost of the board determined the Arduino model.

C. Peripheral Units

This section outlines the peripheral units on NetCar.

1. Motor Driver

NetCar has two BTS 7960 [14] half bridge motor drivers. Each motor driver has one p-channel high side MOSFET and one n-channel low side MOSFET with a microcontroller. The BTS7960 can operate up to 43 A. This maximum limit gives us a large margin of power to support mobility on terrains with rough conditions.

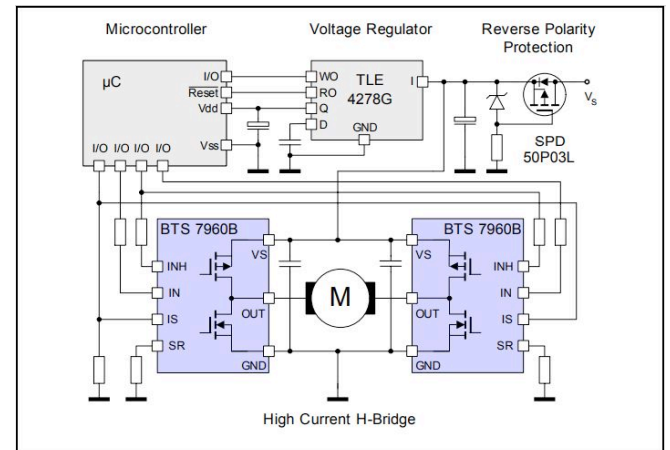


Figure 1: BTS7960 motor driver with a dc motor connection.

Motor drivers on NetCar control four geared dc motors. One of the motor drivers controls the right wheels and other one controls the left ones. By controlling different sides of NetCar, we can ensure the linear motion of the mobile or rotate it. Providing different levels of PWM (pulse width modulation) signals on each side enables controlling the direction of mobility. A detailed discussion on the PWM signal can be found in Section III.C. Figure 1 illustrates the

application of the motor driver with single motor. Note that, NetCar drives two motors in parallel simultaneously.

2. Gear Dc Motor

Gear dc motors are popular in mobile robot applications since motor drivers can control them easily. NetCar is equipped with a standard 12 V gear dc motor (Zhengke ZGA25RP 100RPM [15]). Employed dc motor supports up to 100 RPM. This rotational speed ensures mobility on all-terrain.

Table 2: Specifications of the gear dc motor [16].

Rated voltage	12 V
Load speed	100 RPM
Rated speed	90 RPM
Gearbox length	19 mm
Rated current	0.52 A
Rated torque	0.263 N.m
Maximum torque	0.597 N.m

III. EXPERIMENTAL EVALUATION

In this section, we evaluate the energy consumption characteristics of NetCar in different scenarios. First, we discuss the current sensor that we use in the rest of the experiments and clarify the measurement methodology. Then we explain motor control and duty-cycle. Lastly, we provide results after introducing scenarios.

A. Measuring the Power Consumption

In order to investigate the power consumption characteristics of NetCar, we have employed a bi-directional, high-side current/power monitor named INA219 [17]. INA219 reports current in amperes and calculates power in watts. INA219 can sense from 0 to 26 V which is sufficient considering the employed PWM signal levels. A high-side sensor is preferred to ensure improved dependability.

INA219 exploits 0.1 ohm resistance to measure voltage with 12-bit ADC. This enables improved precision on obtained current values compared to an Arduino board that has 10-bit ADC. By taking advantage of 12-bit ADC and low resistance and tolerance resistance, it is possible to obtain %1 precision. Note that, it also has an option to change resistance and get higher resolution. Despite availability of power monitor on BTS7960 motor drivers, we opt to use INA219 due to improved precision and reliability.

B. Scenarios

In the experiments, we have considered various scenarios considering different terrain features and mobility modes as listed below:

- Terrain type: Indoor (marble) and outdoor (grass and concrete) as illustrated in Figure 2.
- Mobility direction: Straight path and curved path.
- Road gradient: Flat road, uphill, and downhill with varying gradient levels.

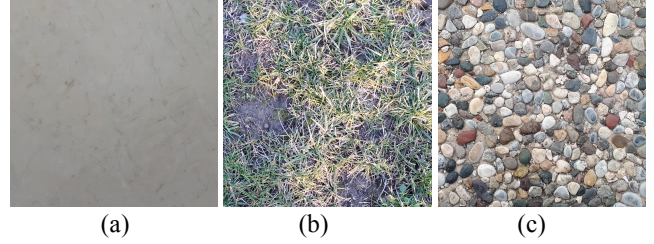


Figure 2: Considered terrain types for mobility; marble (a), grass (b), and concrete (c).

C. Motor Control

We employ pulse width modulation (PWM) to control the output torque of the DC motor. The main challenge of controlling a motor is the limitation of the states (e.g., on and off). When a dc motor is driven, the voltage that feeds motors shall change. Therefore, we must control DC voltage levels while our motors are operating. PWM offers a solution to this problem by providing the flexibility of changing the state of voltage supply that feeds the motor. If the state is changed faster than the controlled device can respond, device receives the average DC voltage. The percentage of the DC voltage is the proportion of time that spent as on and off. High PWM signal is regarded as the “on time”. The ratio of “on time” to “off time” is called the duty cycle. Duty cycle directly affects the average output voltage supplied to the motor. Low-frequency PWM signals can be generated through software. On the other hand, generating high-frequency PWM signals requires hardware support. Employed Arduino board provides 14 digital pins that support PWM output. PWM signal-duty cycle relationship is illustrated in Figure 3.

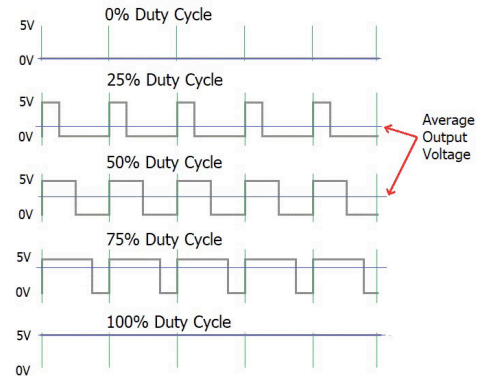


Figure 3: PWM signals and the duty cycle.

D. Results

This subsection discusses the results for evaluations. For each scenario, we repeated the experiment 10 times and reported the average for significance.

1. Straight path with different terrain types

In this subsection, we have evaluated energy consumption of NetCar while following a straight path on the road with different terrain types. The results are presented in Figures 3, 4, and 5 for marble, grass and concrete roads respectively. For each case, we have varied the duty-cycle of the motor and investigated the power consumption.

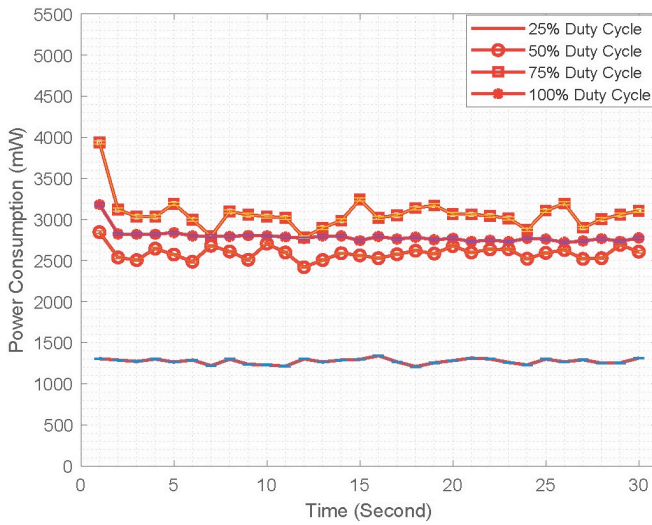


Figure 4: Power consumption on the marble road with respect to the duty cycle.

According to Figure 4, 25% duty cycle incurs the least cost as expected. An unexpected finding is the decline of the power consumption when the duty-cycle is increased from 75% to 100%. This can be attributed to the low-friction of the considered terrain type. We suspect the skidding of wheels due to the loss of road grip. Friction of the road is not sufficient to avoid the sliding of the tire rubber. Note that, NetCar is a 4WD all-terrain mobile platform that has tires more suitable to rough terrain conditions. As can be observed in the rest of the results, reduced power consumption is not a case for grassy roads when duty-cycle is set to 100%.

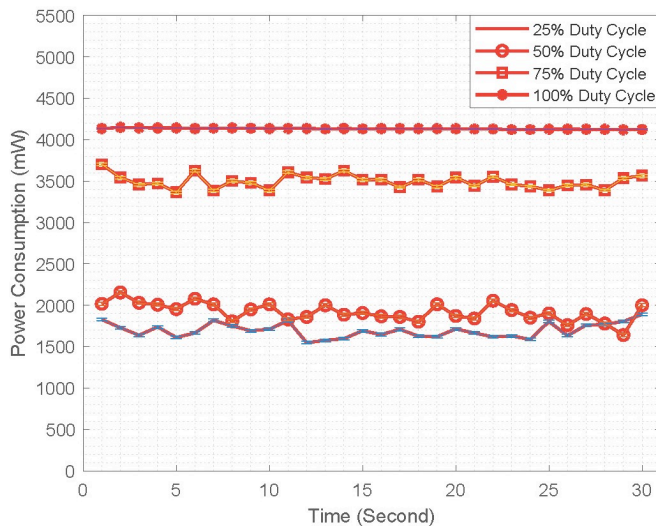


Figure 5: Power consumption on the grassy road with respect to the duty cycle.

Figure 5 signifies that the least power consumption can be attained with 25% duty-cycle as expected. When the duty-cycle is increased to 50%, power consumption increases slightly. Power consumption increases considerable upon increasing the duty-cycle to 75%. The highest cost is incurred

when the duty-cycle is set to 100%. Considering the increased friction force of the grassy road, the phenomenon that we observed on the marble terrain is not applicable anymore. Note that the average power consumption is higher on grassy road compared to the marble road.

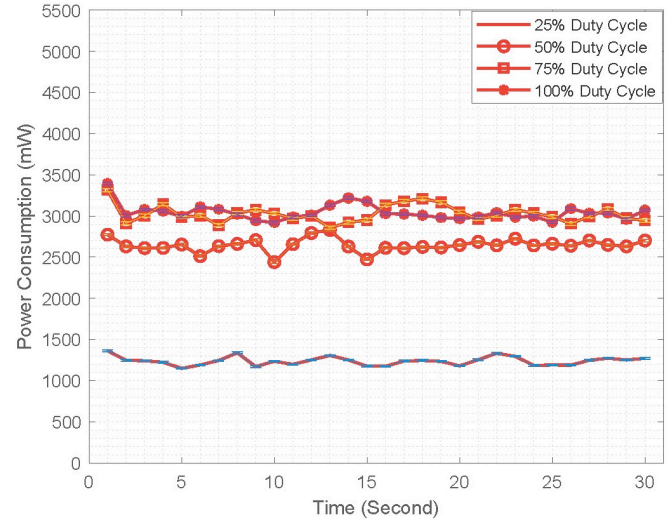


Figure 6: Power consumption on the concrete road with respect to the duty cycle.

The friction force of the concrete terrain is expected to be between marble and grassy roads. As Figure 6 indicates, obtained results show this relationship. As expected, 25% duty-cycle incurs the least cost. The cost increases considerably upon increasing the duty-cycle further. Despite increased cost with the increased duty-cycle, the cost difference is not as distinct as the grassy road. We have also observed declined power consumption on some of the results when the duty-cycle is increased from 75% to 100%. This cannot be generalized though. Results indicate possible skidding of the tires when the torque is increased on the concrete road occasionally.

2. Curved path on the marble road

In the second set of experiments, we have considered a curved path that NetCar makes a right turn while moving forwards. In this experiment, we set the duty-cycle to 100% and used the marble road. The results are depicted in Figure 7. It can be concluded from Figure 7 that power consumption increases considerably on the curved path compared to the straight-path movement. According to the obtained results, power consumption increases 46% upon following the curved path.

3. Straight path with various gradient levels

The last scenario investigates the impact of the gradient level on the power consumption. To implement this scenario, we used a custom wooden road platform to be able vary the gradient level as needed. Due to the limited length of the platform, we set the duty-cycle to 25% to control the velocity of NetCar especially when the gradient is negative (downhill). The results are illustrated in Figure 8. The gradient level is

varied between -20 degrees and +20 degrees. Negative gradient indicates downhill movement (descending) while positive gradient denotes uphill movement (ascending). Flat gradient is represented with 0 degrees. In order to represent the power consumption during ascending and descending separately, we indicate the results on the positive part of the x-axis with two distinct lines denoting uphill and downhill movement in Figure 8. Note that downhill movement actually represents negative gradient in Figure 8. According to results, gradient level is highly correlated with the power consumption. Gradient increases power consumption during uphill movement and alleviates the cost during downhill movement.

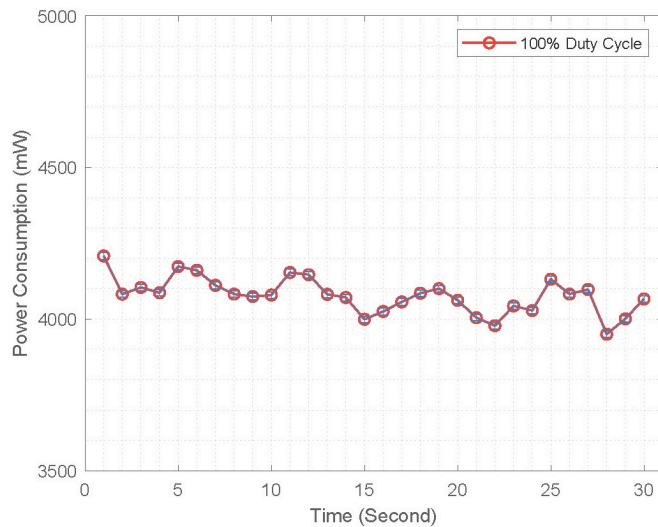


Figure 7: Power consumption upon following a curved path on the marble road with respect to the duty cycle.

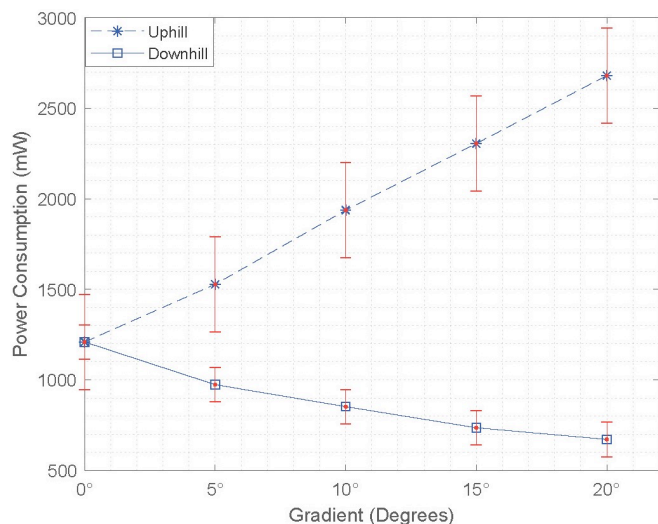


Figure 8: Power consumption when different gradient levels are applied on a wooden platform.

IV. CONCLUSION

NetCar is a mobile robot that can be employed in mobile sensor networks to enable controlled mobility. NetCar can be

used as an actuator connected to sensors that can provide mobility to sensors upon required. NetCar can also be used as a testbed to evaluate algorithms designed for mobile sensor networks. NetCar is a 4WD all-terrain mobile platform that enables deployment in areas with harsh terrain features. In this paper, we have evaluated power consumption characteristics of the platform by considering different scenarios including various terrain types.

ACKNOWLEDGMENT

This work was supported by the Scientific and Technical Research Council of Turkey (TUBITAK) under Grant No. EEEAG-117E050.

REFERENCES

- [1] I. F. Senturk and M. Bilgin, "Network connectivity and data quality in crowd-assisted networks," *Crowd Assisted Networking and Computing*, pp. 163-186. CRC Press, 2018.
- [2] F. K. Shaikh and S. Zeadally, "Energy harvesting in wireless sensor networks: A comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 5, pp. 1041-1054, 2016.
- [3] I. F. Senturk and G. Y. Kebe, "A New Approach to Simulating Node Deployment for Smart City Applications Using Geospatial Data," 2019 International Symposium on Networks, Computers and Communications (ISNCC'19), (to appear).
- [4] I. F. Senturk, "A prescient recovery approach for disjoint msns," 2017 IEEE International Conference on Communications, pp. 1-6, 2017.
- [5] M. Vecchio and R. López-Valcarce, "Improving area coverage of wireless sensor networks via controllable mobile nodes: A greedy approach," *Journal of network and computer applications*, vol. 48, pp. 1-13, 2015.
- [6] F. Tashtarian, M. H. Y. Moghaddam, K. Sohraby, and S. Effati, "On maximizing the lifetime of wireless sensor networks in event-driven applications with mobile sinks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 7, pp. 3177-3189, 2015.
- [7] S. Jananefat, K. Akkaya, I. F. Senturk, and M. Gloff, "Rethinking connectivity restoration in wsns using feedback from a low-cost mobile sensor network testbed," 38th Annual IEEE Conference on Local Computer Networks-Workshops, pp. 108-115, 2013.
- [8] G. Wang, C. Guohong, T. La Porta, and W. Zhang, "Sensor relocation in mobile sensor networks," *IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 4, pp. 2302-2312, 2005.
- [9] M. Younis, I. F. Senturk, K. Akkaya, S. Lee, and F. Senel, "Topology management techniques for tolerating node failures in wireless sensor networks: A survey," *Computer Networks*, vol. 58, pp. 254-283, 2014.
- [10] I. F. Senturk, K. Akkaya, and S. Jananefat, "Towards realistic connectivity restoration in partitioned mobile sensor networks," *International Journal of Communication Systems*, vol. 29, no. 2, pp. 230-250, 2016.
- [11] 4WD Aluminum Mobile Robot Platform, Available: http://www.alsrobot.com/index.php?route=product/product&product_id=535, Accessed: 03/04/2019.
- [12] Arduino, Available: <https://www.arduino.cc>, Accessed: 03/04/2019.
- [13] Wikipedia Article on Arduino, Available: <https://en.wikipedia.org/wiki/Arduino>, Accessed: 03/04/2019.
- [14] BTS7960 Motor Driver Data Sheet, Available: <https://www.mouser.com.tr/ProductDetail/Infineon-Technologies/BTS7960B?qs=wK%252BoHS4yu57Y%2Fa%252Bbuozvew==>, Accessed: 03/04/2019.
- [15] Zhengke DC Motor, Available: https://www.zhengkemotor.com/product/zhengkemotor_product_Dc_Gear_Motor_25mm_Centric_shaft_ZGA25RP_Eccentric_Shift_ZGB25RP.html, Accessed: 03/04/2019.
- [16] DC Motor Specifications, Available: <https://www.banggood.com/4WD-WIFI-Crosscountry-Offroad-Robot-Smart-Car-Kit-For-Arduino-p-927973.html>, Accessed: 03/04/2019.
- [17] INA219, Available: <http://www.ti.com/product/INA219>, Accessed: 03/04/2019.

Modeling and Simulation of a Reconfigurable Microstrip Antenna for Wireless Communication and Mobile

Ashraf Aoad

Istanbul Sabahattin Zaim University, Istanbul, Turkey, ashawad@hotmail.com

Abstract -This study presents a model of a reconfigurable microstrip antenna whose operating frequencies are between 1-10 GHz. The examined model comprised of an octagon and spiral patches to achieve antenna efficiency. The effective configuration of the proposed antenna, and hence its operating frequencies, can be tuned by adding or eliminating five resistors. The proposed antenna can be used and developed for different wireless communication systems. The layout model and simulation are performed using finite integration technique (FIT) and transmission line matrix (TLM) implementations.

Keywords - wideband; reconfigurable microstrip antenna; resistors

I. INTRODUCTION

Work on reconfigurable antennas started many years ago in wireless communication, radar systems, satellite, mobile and microwave link networks [1] [2]. It is still a need for developing new products of reconfigurable antennas to suit existing and emerging communication applications. They present important capabilities to adjust the performance in terms of s-parameters, directivity, radiation pattern and polarization based on communication system requirements [3] [4]. In addition, they have a light weight, a small dimension, easily fabricated, low cost and profile and no complexity to the design [5].

Ideally, to control the current distribution J and hence to obtain the desired results without increasing the physical size [6], it is necessary to add some mechanisms to the structure such as resistors, varactor diodes [7], PIN diodes [8], RF-MEMS [9] switches etc., or using metasurface [10]. In several recent studies, metallic strips are introduced to realize the ON state and removed to realize the OFF state [11].

This paper reports some of the growing needs in the continuing application of reconfigurable antenna designs. The design is comprised of a shape like an octagon patch in the center and spiral microstrips around [12]. They are integrated by five resistors which tune the current comes from the feeding system through octagon patch then to spiral. Combining three mentioned elements determine the results. The objective of this study is to model and simulate a wideband reconfigurable antenna to cover L band 1–2 GHz, S band 2–4 GHz, C band 4–8 GHz and X band 8–10 GHz. The dimensions should be small enough for a standard mobile application.

II. RECONFIGURABLE ANTENNA DESIGN

The studied antenna presents a novel reconfigurable microstrip antenna. This antenna consists of three layers and a feeding system at the center of the octagon patch. The

radiating conductors (first layer) consist of two parts: an octagon patch with different side-length and spiral patches which set to 0.55 cm, 1.2 cm, 1 cm, 1.4 cm and 1 cm respectively for each parameter of $L_1(0.025\lambda_0)$, $L_2(0.054\lambda_0)$, $L_3(0.045\lambda_0)$, $L_4(0.06\lambda_0)$, and $L_5(0.045\lambda_0)$ at a test frequency of $f_T = 1.36$ GHz, where λ_0 is the free-space wavelength. They are modeled on Rogers RT5880 substrate board (second layer) with a thickness of 0.3175 cm and a relative permittivity of 2.2. The ground plane (third layer) is printed on the back side of the substrate. The unfilled spaces of W_1, W_2, W_3, W_4 and W_5 between the spiral and the middle octagon patches set to 0.15 cm, 0.1 cm, 0.15 cm, 0.1 cm and 0.1 cm respectively, include five resistors (R_1, R_2, R_3, R_4 and R_5).

Resistors (except R_5) positioned in the middle of the sides of the octagon patch whereas R_5 is connected upper between L_1 and L_5 to distribute the current paths on the microstrips as shown in Figure 1. Note that the current distributions on both patches are not symmetrical [13]. To realize switching states (ON/OFF), five resistors are used. Each resistor has a resistance value of 5 Ohms for ON state, while removed for OFF state.

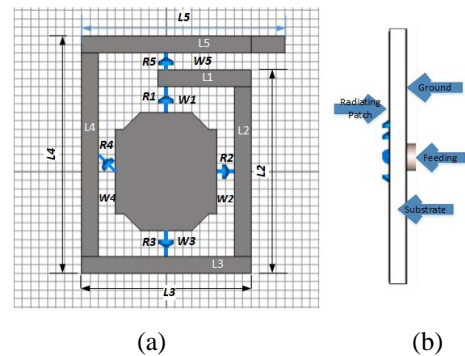


Figure 1: Reconfigurable antenna (a) Top view and (b) Side view.

III. RESULTS AND DISCUSSION

The proposed reconfigurable antenna includes five resistors connected sequentially and hence there are two cases of switching either ON or OFF. Therefore, the proposed antenna is repeatedly simulated 2^5 times without any changing in the structure dimensions. The accuracy of the models is presented by the S-parameter curves as shown in the following figures.

At first, Figure 2 shows the result of the simulation for the octagon patch only, which means all resistors are in OFF state and no physical or electrical connection occur between the center and spiral patches. Operating frequencies are 5.45 GHz

and 6.63 GHz with band width of 0.36 GHz & 0.43 GHz respectively, at a target of return loss of $S_{11} \leq -10$ dB.

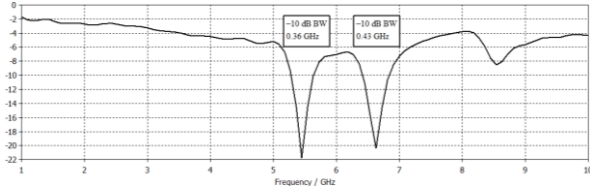


Figure 2: S-parameters, all switches is in OFF state.

Secondly, Figure 3 shows the verity in results depend on connecting R_1 as ON state which allows transition of the current to L_1 then to others sequentially. The current starts flowing via L_1 strongly, and slowly weaken. Operating frequencies are 1.36 GHz, 3.18 GHz, 5.18 GHz and 6.63 GHz with a band width of 0.32 GHz, 0.64 GHz, 0.8 GHz and 0.72 GHz respectively.

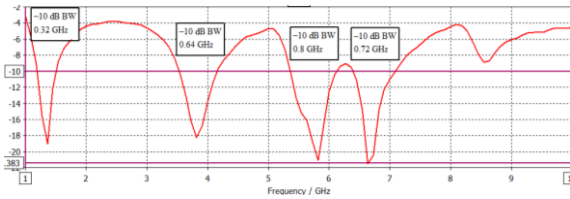


Figure 3: S-parameters, R_1 is only switched ON others OFF.

Thirdly, R_1 and R_2 are two different accesses for current flowing through L_1 and L_2 patches. Accordingly, the antenna resonates at 1.63 GHz, 5.45 GHz and 8.72 GHz with a band width of 0.42 GHz, 0.26 GHz and 0.17 GHz respectively (Figure 4).

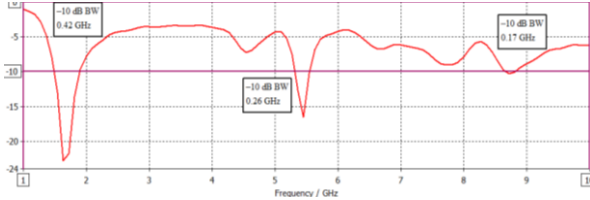


Figure 4: S-parameters, R_1 and R_2 are switched ON others OFF.

Fourthly, the current flowing through R_1 , R_2 and R_3 to the spiral patches affects the operating frequencies to be 2.18

GHz and 5.45 GHz with a band width of 0.48 GHz and 0.18 GHz respectively (Figure 5).

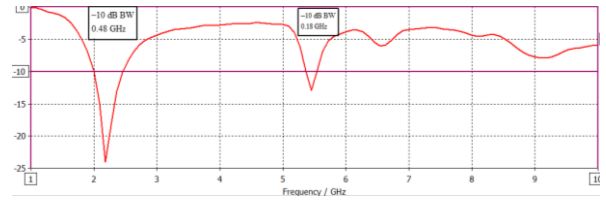


Figure 5: S-parameters, R_1 , R_2 and R_3 are switched ON others OFF.

Fifthly, R_1 , R_2 , R_3 and R_4 are switched ON, the current is transited from different accesses to the spiral. Therefore, the antenna resonates at 3.18 GHz and 5.45 GHz with a band width of 0.37 GHz and 0.13 GHz respectively (Figure 6).

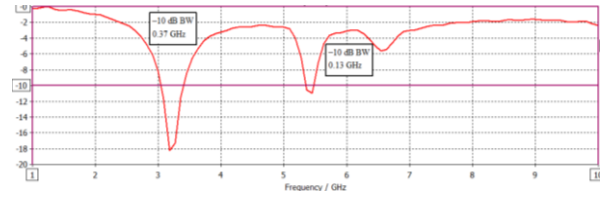


Figure 6: S-parameters, R_1 , R_2 , R_3 and R_4 are switched ON and R_5 is switched OFF.

Sixthly, even though all switches is in ON state, the antenna operates at a frequency of 5.45 GHz with a band width of 0.25 GHz (Figure 7).

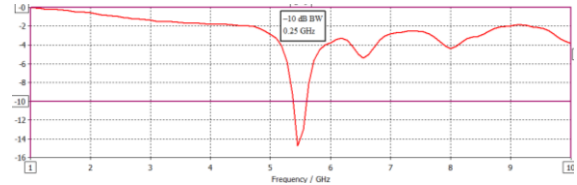


Figure 7: S-parameters, all switches is switched ON.

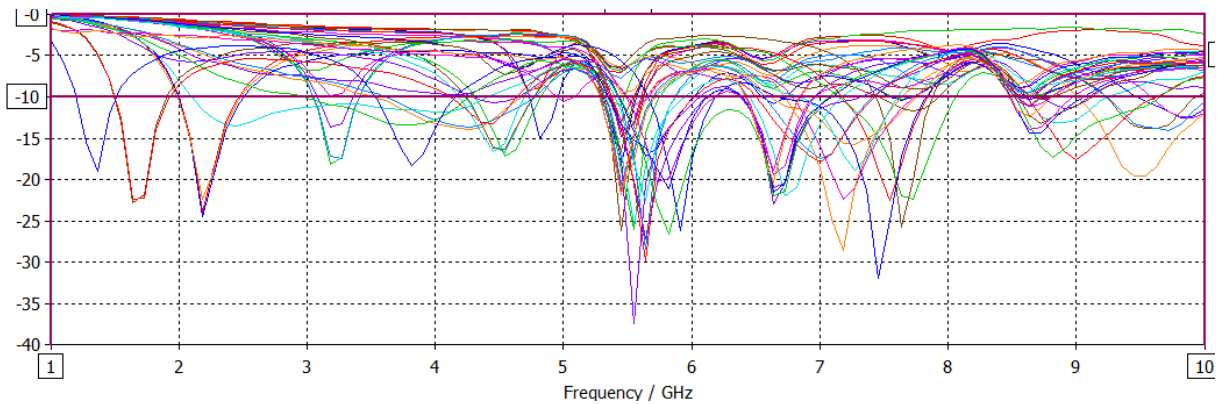


Figure 8: S-parameters of simulation for 32 switching cases.

Switching from a state to another, novel results are obtained as shown in Figure 8, which contains 32 modeling and simulation results. Note that the antenna covers the band

from 1.25 GHz to 8.1 GHz, weakening in resonating to 8.56 GHz and continued to 10 GHz. The current is also not

distributed symmetrically depending on the form of the antenna besides the state of resistors.

IV. CONCLUSION

A small electrical size and a wideband reconfigurable microstrip antenna was modeled and simulated to be suitable for wireless communication and mobile devices. Five resistors were added between the octagon and spiral patches leading to the creation of new resonant frequencies in each switching state. The proposed reconfigurable antenna provides a wide operating range of resonant frequencies between 1-10 GHz. Finally, this paper may be introduced as a handy tool for marketing the proposed antenna where the designer and/or user manages the switches (ON/OFF) for required frequency and bandwidth.

REFERENCES

- [1] J. T. Bernhard, *Reconfigurable antennas*, Champaign: Morgan & Claypool Publishers, 2007.
- [2] Y. Huang and K. Boyle, *Antennas: From Theory to Practice*, A John Wiley and Sons, Ltd, Publication, 2008.
- [3] J. Costantine, *Design, optimization and analysis of reconfigurable antennas*, Albuquerque, New Mexico: PhD, 2009.
- [4] T. Song, Y. Lee, D. Ga and J. Choi, "A Polarization Reconfigurable Microstrip Patch Antenna using PIN Diodes," in *Proceedings of APMC 2012*, Kaohsiung, 2012.
- [5] A. Aoad, Z. Aydin and E. Korkmaz, "Design of a Tri band 5-Fingers Shaped Microstrip Patch Antenna with an Adjustable Resistor," in *2014 IEEE CAMA*, Antibes Juan-Les-Pins, 2014.
- [6] D. B. Miron, *Small Antenna Design*, Elsevier Inc., 2006.
- [7] A. P. Deo, A. Sonker and R. Kumar, "Design of Reconfigurable Slot Antenna Using Varactor Diode," in *Computer, Communications and Electronics (Comptelix), 2017 International Conference*, Jaipur, India, 2017.
- [8] Y. Yashchyn, K. Derzakowski and G. Bogdan, "28 GHz Switched-Beam Antenna Based on S-PIN Diodes for 5G Mobile Communications," *IEEE Antennas and Wireless Propagation Letters*, vol. 17, no. 2, pp. 225 - 228, 2017.
- [9] M. D. Wright, W. Baron and J. Miller, "MEMS Reconfigurable Broadband Patch Antenna for Conformal Applications," *IEEE Transactions on Antennas and Propagation*, no. 99, pp. 1-1, 2017.
- [10] Z. Jiajie, W. Anguo and W. Peng, "A survey on Reconfigurable Antennas," in *IEEE ICMMT2008 Proceedings*, 2008.
- [11] J. S. Kulandai Raj, J. Bonney, P. Herrero and J. Schoebel, "A reconfigurable antenna for MIMO application," *IEEE Loughborough antenna & propagation conference*, 16-17 November 2009.
- [12] A. A. Omar, . O. A. Safia and M. Nedil, "UWB coplanar waveguide-fed coplanar strips rectangular spiral antenna," *Wiley Online Library*, 2017.
- [13] T. A. Milligan, *Modern Antenna Design*, A JOHN WILEY & SONS, INC., PUBLICATION, 2005.

Node Weighting Method in Centrality Measure of Complex Networks

O. FINDIK¹ and E. ÖZKAYNAK¹

¹Karabuk University, Karabuk/Turkey, oguzfindik@karabuk.edu.tr

¹Karabuk University, Karabuk/Turkey, eozykaynak@karabuk.edu.tr

Abstract—Complex networks are one of the tools commonly used in modeling all kinds of events that are related to each other. The identification of effective nodes in complex networks is an important issue needed for the analysis of complex networks. Degree, closeness and betweenness measures are the most important centrality measures commonly used to analyze networks. As a local metric, degree is relatively simple and less effective, although global measures such as the measure of closeness and betweenness can better define effective nodes. However, there are still some disadvantages and limitations of all of these measures. Degree, closeness and betweenness measures are only the data obtained from the topological structure of the network. However, in determining the centralization of the node, other than the topological structure of the network that affect the formation of the network, but not expressed in the topological structure of the network is also effective. In this study, the node weighting method was developed for the determination of node centralization in the network and compared with the current node centering criteria. The experimental study was conducted on a network of players and played competitions in the Australian Open Tennis Tournament held between 2000-2017. By using criteria such as time, experience and success, the weights of the nodes were calculated and compared with the node centering criteria. The results obtained from the experimental study show that the node weighting method gives successful results in the determination of active nodes in complex networks.

Keywords - Complex Networks, Node Weighting, Data Mining

I. INTRODUCTION

The communication and emerging interaction process of people with each other has been developing and growing for centuries. Together with technological developments, these communications and sharing methods are also developing. Everything that people do together is a natural result of communication and sharing. The fact that people connect and interact with each other directly or indirectly constitutes the elements of a complex network. Structures defined as complex networks are not limited to expressing the relationship between people and each other. Everything that has interactions or connections in various ways is actually a part of complex networks. Complex network science, gather and analyze all kinds of structure, system and situation that have the relationship, connection and sharing between them, direct and indirect ways, within the framework of certain rules and disciplines. [1]. In the analysis of complex networks, the position and importance of the nodes in the network is important to examine how the information flows in the network and the structure of the network [2]. Centrality is one of the

most important concepts that are considered when examining the structure of a complex network. The concept of centrality is a set of indicators that reveal the importance of the node in the network in a complex network. In most complex networks, some links or nodes are more central than others. To measure the centrality of the nodes, there are measures of centrality, such as degrees, interactivity, proximity and eigenvectors. These measures of centrality are obtained by using the topological information of the network. However, in complex networks, only topological consideration of the centrality of a node may not be sufficient to reveal some important nodes in the network. Increasing or decreasing the effect of the nodes on the neighbors with the expansion of the network over time can be important in determining the node center. According to the topological structure, the metrics that determine the node centrality do not take into account the factors that may affect the node centrality except the topological structure of the network. Depending on the time and the different conditions, changes in the nodes state in the network should be taken into account in revealing the node centrality. In this study, which emphasizes the importance of node weighting in uncovering node centrality, in order to calculate the weights of the nodes, the common criteria for all nodes in the network were determined and the weights of the nodes were calculated using the multi-criteria decision-making method. In the experimental study, node weighting method was used to reveal the node centrality in the network in a network composed of competitions played between 2000-2017 Australia in the Open Tennis Tournament. The experimental results show that the node weighting method gives successful results in order to reveal the node centrality.

II. METHOD

A. Network Centrality

The centrality factor is an important concept, particularly in complex network structures. Network centrality can be defined as the criterion of nodes in a complex network. [3]. This criterion measures the strength of a node's dependence on other nodes while on the other hand it measures its effectiveness on other nodes. In determining the important and active locations in the network, there are usually the most important or most known nodes. Different centrality measures have been proposed to determine network centrality. The proposed centrality criteria and the characteristics of the node position in social networks have been tried to be defined and measured. [4].

Centrality Criteria

A complex network analysis is the study of the interactions that are in interaction with each other, the types of relationships and the interaction between relationships. The concept of relationship can be defined as the link between social entities. [4]. The individuals, objects and communities that form the basis of relations are expressed by nodes in network analysis [5]. The relations established in the social networks due to the interaction between the nodes are the equivalent of the concept of links. The relationship between individuals can be expressed in different types such as friendship bond, kinship bond, competition tie, financial tie, belief tie [6-11]. The proposed criteria for examining the link structures between nodes in the networks formed contain information about the importance of the node in the network.

Degree Centrality

Degree centrality, which is the simplest of the criteria used to reveal network centrality, is calculated by the number of links to a node on the network. Although it is simple to calculate, it is an important criterion that can indicate the position of the node in the network. In most complex networks, the more links a node has, the more important and powerful it is. In fact, the node with the highest degree can be interpreted as the most active member of the network. In cases where links are direct, the number of in-degree links and the number of out-degree links are calculated separately [12,13].

Eigenvector Centrality

The Eigenvector Centrality criterion, another criterion used to elucidate network-centrality, is a criterion that indicates that the links owned by the node taken into account when calculating the rating are not equal. For a node on the network, the effect of links to key nodes may be more than any other ordinary links. The fact that the nodes on which it is connected are more centrally indicate that the node will be in a more central position. When calculating this criterion, the totality of the neighbors' centrality is taken into account. [14,15].

Betweenness Centrality

Betweenness centrality is one of the most complex criteria to be calculated in the centrality criteria, which is one of the criteria used to elucidate network-centrality. Betweenness centrality is found by the ratio of the most shortcuts that pass through a node in the network. First, the most shortcuts be found between all node pairs in the network. Ratio between this shortcuts (geodesics) and number of path that node is located on gives the betweenness centrality. Because it is a very costly measure that can be calculated on large networks, it can also be calculated by going down to certain level neighbors. Due to their location, the nodes with a high degree of gravity are more important than the other nodes and will be more aware of what is going on in the network [12,13].

Closeness Centrality

The Closeness centrality is another centrality criterion used to reveal network. It is found that any node in the network obtains the shortest mean distance to all other nodes (geodesic distance). If the links are directional, these aspects should be taken into consideration when finding the shortcuts. In this case, two different metric which are in-closeness and out-closeness are calculated [12, 14].

B. Gephi

Because of the large number of nodes and links due to the nature of complex networks, it is important to use appropriate analysis tools to ensure that the data obtained from these nodes and links are accurate, understandable and interpretable. In complex network analysis, it is important to detect network centrality, to find shortest paths between nodes, to obtain clustering coefficients and to calculate degree distributions. In addition, the visualization of the data obtained in an appropriate manner has an important place in the understanding and analysis of the structure of the network. In this study, Gephi software was used to obtain data showing the centrality of the network of tennis competitions and to visualize the network. Gephi software is preferred because of it is free and has been instrumental in choosing to provide detailed options during the performance of the analysis data and the visualization of the network[18].

C. Approach of Logarithmic Concept (APLOCO)

One of the multi-criteria decision-making methods, APLOCO, uses Multi-Layer Perceptron, which is one of the Deep Learning methods in the field of artificial intelligence in determining the weight of criteria. One of the important features of APLOCO is that it is not dependent on artificial intelligence at determining the weight of the criteria. Criterion weights can also be determined by any method. As a result, the determined weights can be used in this method if the criteria weights are determined by any method. The reason for using APLOCO method in this study is that the data used for experimental study does not show normal distribution and allows to calculate the node weight from different criteria. The fact that artificial intelligence does not exclude the data and have the ability to learn from the data constitutes one of the main reasons for using the criteria in the process of determining the weight [16].

D. Implementation Steps of APLOCO

The implementation of APLOCO is completed in 5 steps;

- Step 1: Building the decision matrix (DM)
- Step 2: Calculation of starting point criteria (SPC) values
- Step 3: Forming the logarithmic conversion (LC) matrix
- Step 4: Determining the weights of criteria (WC) and calculating the weighted logarithmic conversion (WLC) matrix
- Step 5: Determination of the best alternative (BA)

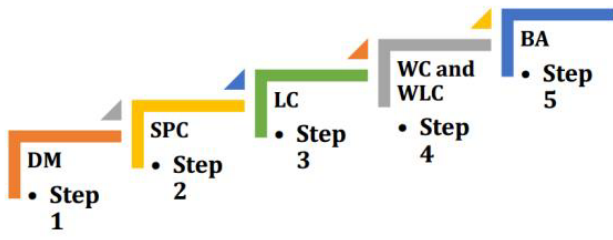


Figure 1: APLOCO Application Steps.

Step 1: Building the decision matrix (DM)

The generated decision matrix (X_{ij}) is a $C \times R$ dimensional matrix that contains the criteria in the rows and the alternatives in the columns. Here, respectively, the number of criteria and the number of alternatives are re-evaluated. X_j refers to decision variables and X_i refers to the values of decision variables. This matrix is shown in the equation in (1).

$$X_{ij} = \begin{matrix} \text{Criteria} \\ \begin{bmatrix} C_1 \\ C_2 \\ \dots \\ C_c \end{bmatrix} \end{matrix} \begin{matrix} \text{Alternatives} \\ \begin{bmatrix} A_1 & A_2 & \dots & A_r \end{bmatrix} \end{matrix} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1r} \\ X_{21} & X_{22} & \dots & X_{2r} \\ \dots & \dots & \dots & \dots \\ X_{c1} & X_{c2} & \dots & X_{cr} \end{bmatrix} \quad (1)$$

Step 2: Calculation of starting point criteria (SPC) values

At this stage, if the criterion value needs to be maximum, the maximum value between the values of the corresponding criterion in that line is determined as the maximum value. If the criterion value is to be minimum, the minimum value between the values in the corresponding row is set to the minimum value. In this case, if the desired criteria is the maximum, the criteria values in the row to which the maximum value belongs are subtracted from the maximum value. On the other hand, if the required criterion is minimum, the minimum value is subtracted from the criterion values in which it belongs. In order to perform said operations, the P_{ij} values containing the maximum and minimum starting point criterion values are obtained using the equations (2), (3) and (4). Here, $i = 1, 1 \dots c$; $j = 1, \dots, r$.

$$P_{ij} = \begin{cases} \max x_{ij} - x_{ij} & \text{if } P_{ij} \text{ is the maximum starting point criterion.} \\ x_{ij} - \min x_{ij} & \text{if } P_{ij} \text{ is the minimum starting point criterion.} \end{cases} \quad (2)$$

(3)

$$X_{ij} = \begin{bmatrix} X_{11} - \min x_{1j} & X_{12} - \min x_{1j} & \dots & X_{1r} - \min x_{1j} \\ X_{21} - \min x_{2j} & X_{22} - \min x_{2j} & \dots & X_{2r} - \min x_{2j} \\ \dots & \dots & \dots & \dots \\ X_{c1} - \min x_{cj} & X_{c2} - \min x_{cj} & \dots & X_{cr} - \min x_{cj} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1r} \\ X_{21} & X_{22} & \dots & X_{2r} \\ \dots & \dots & \dots & \dots \\ X_{c1} & X_{c2} & \dots & X_{cr} \end{bmatrix} \quad (4)$$

Step 3: Forming the logarithmic conversion (LC) matrix

At this stage, +2 integer value is added to each of the P_{11} , P_{12} , P_{13} , \dots , P_{1r} values in the rows. The LC values are then calculated by calculating the cyclic natural logarithm as opposed to the results obtained. This is done by the equation (5) and the normalization process is completed. Logarithmic transformation matrix is then obtained in equation (6). Here, \ln is undefined when 0 and negative values occur and $\ln = 0$. For

these reasons, +2, greater than 1 in the equation (4), is considered to be an integer value. Another reason to add number 2 as an integer value to the logarithm value at that time is to avoid excessive values and negative values and to ensure that the values are positive.

$$\ln(x) = \log_e(x) \text{ and } L_{ij} = \frac{1}{\ln(p_{ij} + 2)} \text{ for } (i = 1 \dots c \text{ and } j = 1 \dots r) \quad (5)$$

$$X_{ij} = \begin{bmatrix} \frac{1}{\ln(p_{11} + 2)} & \frac{1}{\ln(p_{12} + 2)} & \dots & \frac{1}{\ln(p_{1r} + 2)} \\ \frac{1}{\ln(p_{21} + 2)} & \frac{1}{\ln(p_{22} + 2)} & \dots & \frac{1}{\ln(p_{2r} + 2)} \\ \dots & \dots & \dots & \dots \\ \frac{1}{\ln(p_{c1} + 2)} & \frac{1}{\ln(p_{c2} + 2)} & \dots & \frac{1}{\ln(p_{cr} + 2)} \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1r} \\ l_{21} & l_{22} & \dots & l_{2r} \\ \dots & \dots & \dots & \dots \\ l_{c1} & l_{c2} & \dots & l_{cr} \end{bmatrix} \quad (6)$$

Step 4: Determining the weights of criteria (WC) and calculating the weighted logarithmic conversion (WLC) matrix

W is a weight coefficient. Here,

$w_1 \in \mathbb{R} \text{ ve } \sum_{i=1}^n w_i w_1 \in \mathbb{R} \text{ ve } \sum_{i=1}^n w_i = 1$ The equation of the WLC matrix (7) is obtained by multiplying the logarithmic transformed l_{ij} values by the weight levels (w_i) of the criteria determined by any method.

$$T_{ij} = \begin{bmatrix} l_{11}w_1 & l_{12}w_1 & \dots & l_{1r}w_1 \\ l_{21}w_2 & l_{22}w_2 & \dots & l_{2r}w_2 \\ \dots & \dots & \dots & \dots \\ l_{c1}w_n & l_{c2}w_n & \dots & l_{cr}w_n \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1r} \\ t_{21} & t_{22} & \dots & t_{2r} \\ \dots & \dots & \dots & \dots \\ t_{c1} & t_{c2} & \dots & t_{cr} \end{bmatrix} \quad (7)$$

Step 5: Determination of the best alternative (BA)

The maximum values of the criteria in each order are determined as optimal solution values (β_j) and after the total score is obtained as β_{sj} . This process is done by the equations (8) and (9). The final scores (j_i) of each alternative are calculated by proportioning the total scores (α_{si}) of the criterion values of the alternatives to the optimum solution values (β_{sj}) collected. This process is done by (10) and (11) equations, respectively. The scores obtained from Equation (11) are between 0 and 1, and the scores are allowed to be evaluated within this range. Then the θ_i values are sorted from large to small and the first order alternative is considered the most suitable alternative [17].

$$B_{sj} = \{ \max t_{ij} \} \text{ and } B_{ij} = \{ t_{11}, t_{12}, t_{13}, \dots, t_{1r} \} \quad (8)$$

$$B_{si} = \sum_{i=1}^n (l_1, l_2, l_3, \dots, l_n) \quad (9)$$

$$a_{sj} = \sum_{j=1}^n (t_1, t_2, t_3, \dots, t_n) \quad (10)$$

$$0 \leq \theta_i \leq 1 \text{ and } j = 1, 2, \dots, r. \theta_i = \frac{\alpha_i}{\beta_i} \quad (11)$$

E. Node Weighting Method

Node weighting is the process of calculating the weights of the nodes in the network according to the determined criteria by APLOCO method which is one of the multi criteria decision making methods. In this approach, the calculated node weights indicate the importance of the node in the network. As the weight is reduced, the position of the nodes is away from the centrality of the network as the weight of the node is at the center of the network.

III. EXPERIMENTAL STUDY

In the experimental study, the network in Figure 2 was formed using the athletes who competed in the Australian Open [17] tennis tournaments between 2000-2017 and the competitions taking place between them.

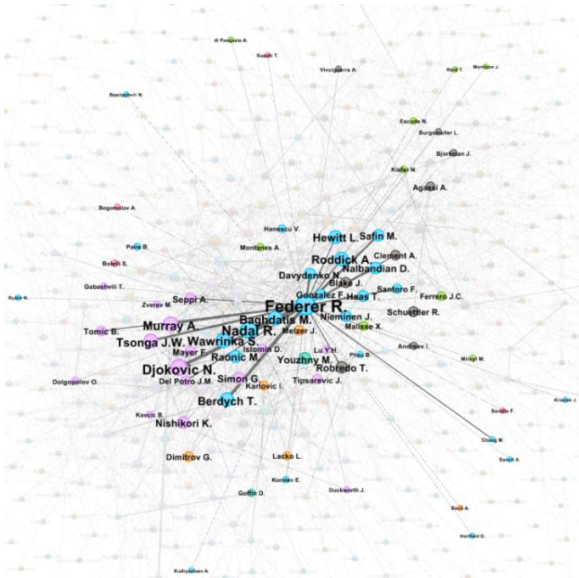


Figure 2: Network Created for Australian Open Tennis Tournament

In the created network, the nodes are composed of athletes and the links are composed of the matches between athletes. The links between the nodes are constructed as unidirectional. The node and link information for the created networks is shown in Table 1.

Table 1: Nodes and Links for the Tennis Network

	Australian Open
Nodes	553
Edges	2156

Centrality measures were calculated by modeling the network through the Gephi program to reveal the nodes in the center of the network.

A. Criteria

APLOCO method was used for the proposed node weighting method and node weights were calculated. For data to be used in APLOCO, time, experience, success and tour criteria are determined according to the structure and characteristics of the network. The criteria in the generated network are shown in (12), (13), (14) and (15).

Time Criteria

In time-dependent expanding networks, the position of the nodes over time is an effective element in determining their position in the network. In the tennis network, the effectiveness of an athlete who has decreased or ended his / her participation in tennis competitions must decrease in the network. Therefore, the time criterion should be used as an important parameter in determining the effectiveness and importance of nodes in the network. The time criterion used in this study was calculated as shown in equation (12) as the ratio of the last year of the tournament year played to the time interval of the netting weighted network.

$$Time = \frac{\text{The Last Competition Year} - \text{The First Tournament Date}}{\text{The Last Year Calculated} - \text{The First Year Calculated} + 1} \quad (12)$$

Here, the last competition year refers to the year of the last tournament where the athlete participated, the first competition year refers to the year of the first tournament where the athlete participated, the last year calculated refers the date of the last tournament in the network that will be weighted, the calculated year refers to the date of the first tournament that took place in the network to be weighted. A value of 1 in Equation 2.30 is used to avoid the result of 0 when the desired range corresponds to the same year.

Experience Criteria

In a network of sporting events, such as a tennis tournament, the athlete's experience in the tournament is effective being in the center of the network. The more matches an athlete has, the more likely he is to link with so many athletes. The criterion of experience used in this study was calculated as shown in (13) as the ratio of the number of matches played in tournaments to the total number of matches in tournaments.

$$Experience = \frac{\text{The Number Of Matches Played}}{\text{The Total Number Of Matches}} \quad (13)$$

The number of matches played here is the total number of matches of the athlete, while the total number of matches played represents the total number of competitions in tournaments.

Success Criteria

The expansion of the network in sports tournament networks is not just about joining new athletes to the network. Athletes with high success in the network also have effects on the expansion of the network. It is possible for the athlete with high success to be positioned in the center of the network due to the fact that he competes with many athletes. The success in the tournaments as well as the number of matches played by the athlete is a criterion that increases the effectiveness and importance of the network. The success criterion used in this study was calculated as the ratio of the number of matches won by the athlete in the tournaments to the total number of matches played in the tournaments as shown in the equality (14).

$$Success = \frac{\text{The Number Of Matches Won}}{\text{The Number Of Matches Played}} \quad (14)$$

Here, the number of matches won by the athlete refers to matches won by athlete in the matches played, the number of matches played refers to the total number of competitions the athlete has in tournaments.

Tour Criteria

Considering the fact that the tennis tournaments are composed of tours within themselves, as in most sports events, the number of rounds played by an athlete in each tournament is one of the factors in the position of the athlete in the network. The fact that the athlete remains in the rounds in the tournament is an important criterion in revealing the role of the athlete in the network. The tour criterion used in this study as shown in (15) is obtained by summing the ratio of the number of rounds played by the athlete per year between the years determined and the total number of rounds in the tournament.

$$Tour = \frac{\sum_{\text{Year}=\text{The First Tournament Date}}^{\text{The Last Competition Year}} \frac{\text{The Number Of Rounds Played}_{\text{year}}}{\text{The Total Number Of Rounds}}}{\text{The Total Number Of Rounds}} \quad (15)$$

Here, the first competition year refers to the date of the first tournament on the network that will be weighted by the node, the last competition year refers to the date of the last tournament on the network that will be weighted by the node, the number of rounds played by the athlete refers to the number of rounds played by the athlete on a yearly basis, the total number of rounds refers to the total number of rounds held on a year. Although the number of criteria for calculating the weights of the nodes is completely independent, it is important to construct the parameters with distinctive features according to the structure and characteristics of the network.

After determining the criteria for the weights of the nodes, multi-criteria decision-making method APLOCO was carried out and the weights of the nodes were calculated based on the determined criteria. These weights could also be used to evaluate the importance and effectiveness of nodes in the network. After calculating the weights of the nodes, the nodes

in the center of the network have been revealed according to the proposed node weighting method.

IV. RESULTS AND FINDINGS

In the experimental study, a network of competitions between the years 2000-2017 of the Australian Open Tennis Tournament was established, and the network centrality criteria and the proposed node weighted centrality in this study were calculated. Calculated degree of centrality is shown in Table 1, Closeness centrality in Table 2, eigenvector centrality in Table 3, betweenness centrality in Table 4, and node-weighted centrality criteria are shown in Table 5.

Table 2: Degree Centrality

	Athlete	Degree
1	Federer R.	71
2	Ferrer D.	51
3	Djokovic N.	48
4	Murray A.	46
5	Berdych T.	45
6	Nadal R.	45
7	Roddick A.	43
8	Wawrinka S.	42
9	Hewitt L.	40
10	Tsonga J.W.	39

Table 2: Closeness Centrality

	Athlete	Closeness Centrality
1	Federer R.	0,460
2	Nadal R.	0,429
3	Ferrer D.	0,425
4	Djokovic N.	0,421
5	Roddick A.	0,420
6	Murray A.	0,418
7	Baghdatis M.	0,415
8	Berdych T.	0,413
9	Hewitt L.	0,410
10	Wawrinka S.	0,409

Table 3: Eigenvector Centrality

	Athlete	Eigenvector Centrality
1	Federer R.	1,000
2	Nadal R.	0,762
3	Djokovic N.	0,750
4	Ferrer D.	0,692

5	Murray A.	0,640
6	Wawrinka S.	0,636
7	Tsonga J.W.	0,602
8	Roddick A.	0,598
9	Baghdatis M.	0,598
10	Berdych T.	0,580

Table 4: Betweenness Centrality

	Athlete	Betweenness Centrality
1	Federer R.	0,1195
2	Ferrer D.	0,0547
3	Berdych T.	0,0462
4	Murray A.	0,0441
5	Hewitt L.	0,0424
6	Roddick A.	0,0414
7	Djokovic N.	0,0377
8	Nadal R.	0,0372
9	Lopez F.	0,0369
10	Tsonga J.W.	0,0342

Table 5: Node Weighting Centrality

	Athlete	Node Weighting Centrality
1	Federer R.	0,975
2	Djokovic N.	0,889
3	Nadal R.	0,869
4	Murray A.	0,857
5	Wawrinka S.	0,844
6	Berdych T.	0,839
7	Ferrer D.	0,837
8	Tsonga J.W.	0,837
9	Raonic M.	0,835
10	Agassi A.	0,833

Calculates network centrality according to the topological structure of the network; According to the results of Table 1, Table 2, Table 3 and Table 4, it is seen that the nodes, which have lost their importance over time, are in the centrality of the network. It has been observed that sportsmen such as Ferrer, Hewitt and Roddick, who have not played for a long time, are in the center of the network due to the high number of competitions in the past and still take place as popular nodes. However, it is important that these athletes, who have not played for a long time, have reduced their effectiveness and importance in the network and these situations can be identified when analyzing the network.

In the present study, it is seen that the weighted centrality results of the proposed node consist of the athletes who continue to play tennis as seen in Table 5. This indicates that effective and important nodes in the network can be detected. It is seen that the athletes who stopped playing tennis or who have lost their success in the early rounds, have moved away from the centrality of the network.

V. CONCLUSION

In this study, node weighting was performed according to the criteria determined by multi-criteria decision-making method in a network composed of athletes in Australian Open tennis tournaments played between 2000-2017 and their competitions. The results of the network centrality calculated according to the obtained node weights were compared with the results of the network centrality criteria used in the complex networks. As a result of the experimental study, it was observed that the proposed node-weighted centrality criterion yielded more successful results compared to the network centrality measures that were calculated by topological measurements. Particularly in dynamic networks and time-varying conditions are included, more effective results have been obtained in determining the effective and important nodes in the network with the node-weighted centrality criterion. Especially in networks which are formed by different factors and which continue to grow, calculating network centrality with the topological structure of the network as well as the factors affecting the growth of the network, the network can be analyzed better.

REFERENCES

- [1] S. H. Strogatz, *Exploring Complex Networks*, Nature, 410, (6825), 268-276 (2001).
- [2] Ş. Akal, "Gerçek ve model ağların karakteristik özelliklerinin karşılaştırılması", Marmara Üni. Öneri Dergisi., 11 (41): 251-272 (2014).
- [3] H.Kuduğ., "Sosyal Ağ Analizi Ölçütlerinin İş Ağlarına Uyarlanması", Yüksek Lisans Tezi, Ege Üniversitesi Fen Bilimleri Enstitüsü, İzmir, (2011).
- [4] C.Salman, "Yeni Bir Ağ Merkezilik Ölçütü: Göreceli Kenar Önemi Metodu", Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, (2018).
- [5] Lee, C. Y., "Correlations Among Centrality Measures In Complex Networks". arXiv preprint physics/0605220, (2006).
- [6] Wasserman, S., & Faust, K., "Social Network Analysis: Methods And Applications (Vol. 8)". Cambridge university press., (1994).
- [7] Bolland, J. M. . "Sorting Out Centrality: An Analysis Of The Performance Of Four Centrality Models In Real And Simulated Networks". Social networks, 10(3), 233-253, (1988).
- [8] Borgatti, S. P." Centrality And Network Flow". Social networks, 27(1), 55- 71, (2005).
- [9] McCulloh, I. "Detecting Changes In A Dynamic Social Network". Doctoral dissertation, Carnegie Mellon University, Pittsburgh (2009).
- [10] Rothenberg, R. B., Poterat, J. J., Woodhouse, D. E., Darrow, W. W., Muth, S. Q., & Klov Dahl, A. S. "Choosing A Centrality Measure: Epidemiologic Correlates In The Colorado Springs Study Of Social Networks". Social Networks, 17(3-4), 273-297, (1995).
- [11] Faust, K. (1997). "Centrality In Affiliation Networks". Social networks, 19(2), 157- 191, (1997).
- [12] Scott, J., "Social Network Analysis: A Handbook", London, Sage Publications, 209p., (2000).
- [13] Mika, P., "Social Networks And The Semantic Web, Semantic Web And Beyond Computing For Human Experience", Jain, R. and Sheth, A. (Eds.), New York, Springer Science and Business Media, 234p., (2007).
- [14] Newman, M.E.J., "Mathematics Of Networks", 12pp., The New Palgrave Dictionary of Economics, Durlauf, S.N. and Blume, L.E. (Eds.), Palgrave Macmillan, (2008).

- [15] Ruhnau, B., “*Eigenvector Centrality: A Node-Centrality?*”, Social Networks, 22 (4): 357-365,(2000).
- [16] T. Bulut, A New Multi Criteria Decision Making Method : Approach of Logarithmic Concept (APLOCO), Int. J. Artif. Intell. Appl. 9 (2018) 15–33. doi:10.5121/ijaia.2018.9102.
- [17] Tennis Results and Tennis Betting Data. ATP Men's Tour, Available: <http://www.tennis-data.co.uk/alldata.php>, (March, 2019)
- [18] Bastian M., Heymann S., Jacomy M. (2009). “*Gephi: an open source software for exploring and manipulating networks*”.International AAAI Conference on Weblogs and Social Media,(2009).

Deep Learning Based Web Application Security

M. SEVRİ¹ and H. KARACAN²

¹ Gazi University, Ankara/Turkey, mehmetsevri@gmail.com

² Gazi University, Ankara/Turkey, hkaracan@gazi.edu.tr

Abstract - With the widespread use of web applications, there has been an increase in the number of attacks against these applications and ensuring web application security has become an important problem. Nowadays, the number of tools used in web attacks is increasing, and more sophisticated attacks are supported. However, the need for technical information to attack web application with using these attack tools decreases. In this study, a web application security model based on Long Short Term Memory (LSTM) has been developed for web application security. As part of the study, CSIC-2010 dataset, which is known and up-to-date in web security field, is used. The text-based data for web traffic in the dataset was converted to sequential binary vectors using one hot encoding technique, and a new approach has been introduced with the deep learning-based LSTM method, simulating the web security problem as a text-processing problem. As a result of the study, a model with a high accuracy rate of 94.41% has been developed for classifying web traffic as normal or attack. This study distinguishes from other web security studies in terms of its high classification success and applied technique.

Keywords - web application security, waf, deep learning, lstm, one hot encoding.

I. INTRODUCTION

WEB applications have an indispensable place in people's daily lives and they are seen as important services/applications in private and public areas with their features such as making life easier, being economical, saving time and being space-independent. The widespread use of web applications renders them to be very favorable targets by hackers. Nowadays, the number of tools used in attacks has increased and these tools are becoming able to support attacks that are more sophisticated. Figure 1 shows the increase in the development of attack tools and the corresponding decrease in the required knowledge / competence level [1].

According to Symantec's Internet security threat report [2] at the end of 2017, web applications continue to be one of the biggest targets in the area of cyber security, with an average of 229,000 daily attacks. 76% of the scanned websites contain vulnerabilities and 9% contain critical (high-risk) vulnerabilities (see Figure 2).

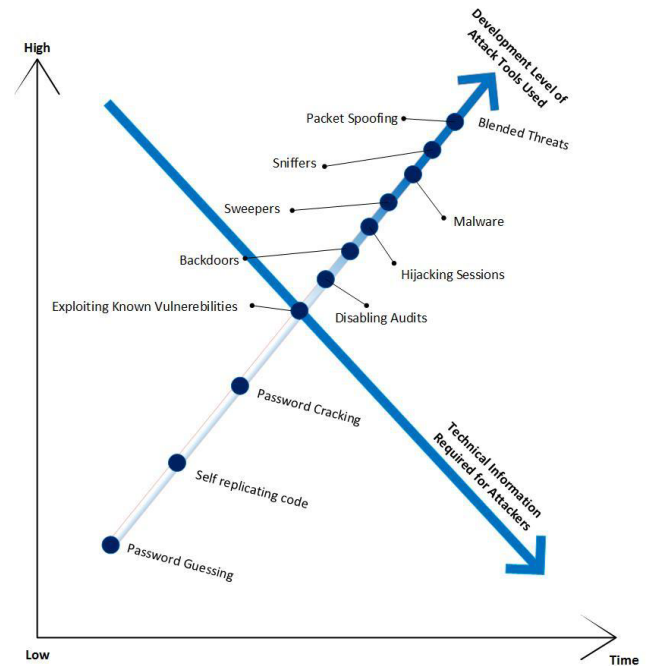


Figure 1: The relationship between the development level of sophisticated attack tools and the technical attacking knowledge needed [1]

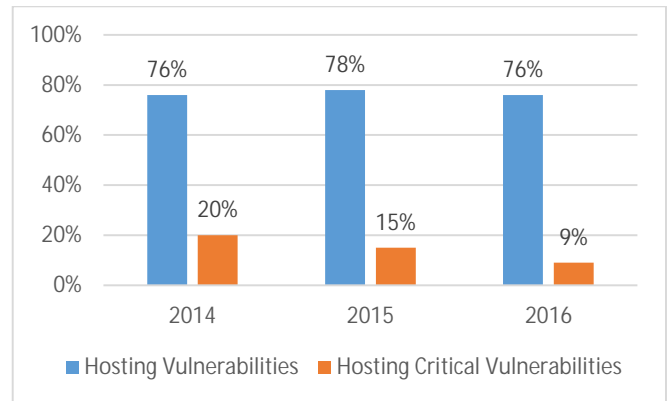


Figure 2: Percentage of the scanned websites hosting vulnerabilities and critical vulnerabilities

When the cyber security researches are examined, it is seen that the most studies are on network security, while the number of studies conducted in the field of web security is limited.

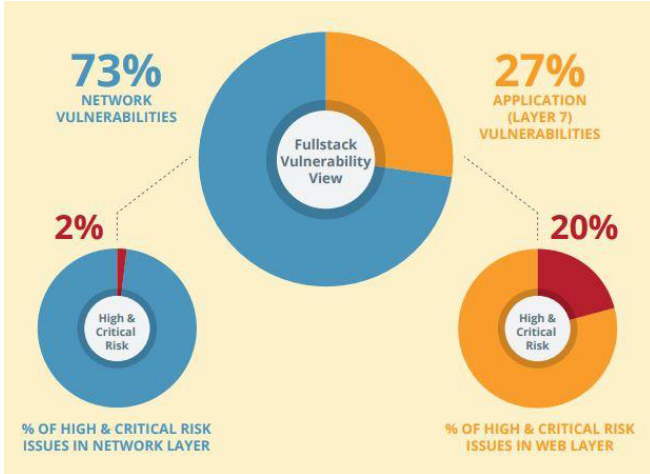


Figure 3: Network and application layer vulnerabilities and critical vulnerabilities percentages [3]

However, according to the 2018 vulnerability statistics report published by the cyber security firm Edgescan, 73% of the detected violations are network-based with only 2% critical vulnerabilities, whereas 27% of the detected violations concern application level (vulnerabilities are all related to web applications) and 20% of these are critical vulnerabilities [3]. According to this, the rate of critical vulnerabilities at the application level is 5.7%, while this rate is only 1.46% for the network-based violations (see Figure 3).

As the number of attacks on web applications has increased and become more sophisticated, web security has become an important problem. There are two main ways to ensure web security; signature based and anomaly based methods. Recently, with the progress in computational power, deep learning methods have begun to stand out in the field of artificial intelligence with their outstanding success and performances. Within the scope of this study, an artificial intelligence-based web application firewall (WAF) has been designed which can detect and prevent attacks on web applications based on deep learning methods with LSTM. The proposed WAF system is unique in terms of the method used and stands out with its success in detecting web attacks. In the rest of the study, web vulnerabilities and prominent studies in the literature have been discussed and the proposed system has been presented with the method and results, and with the result and discussion section, it has been concluded.

II. WEB VULNERABILITIES AND WAF SYSTEMS

The Open Web Application Security Project (OWASP) periodically determines the most critical 10 web application security risks for web vulnerabilities. In the study, firstly it is planned to address some of the vulnerabilities in OWASP Top Ten. The most critical 10 web application security risks listed by OWASP lastly in 2017 [4] respectively, can be seen in Table 1, and the source of these risks and the affected layers are available below.

Table 1: The most critical 10 web application security risks, OWASP - 2017

No	Security Risk	Source	Destination
1	Injection	Application	Server / Database
2	Broken Authentication	Application	Clients
3	Sensitive Data Exposure	Application	Clients
4	XML External Entities - XXE	Application	Server / Network
5	Broken Access Control	Application	Clients
6	Security Misconfiguration	Server	Server
7	Cross-Site Scripting, XSS	Application	Clients
8	Insecure Deserialization	Application	Server
9	Using Components with Known Vulnerabilities	Application	Server / Clients
10	Insufficient Logging & Monitoring	Server / Network	Server / Clients

Firewalls are placed at the in / out point of the network when designing internal networks. Firewalls which can work on different network layers, even on the 7th layer (application layer), will not be able to adequately protect the internal network from outside attacks, if the configuration settings are not right and adequate. In general, organizations write security rules on the classic firewalls, creating security rules at 4th layer in OSI by using the source IP, destination IP, port number and packet type (TCP or UDP). Since a web application will communicate with end users via TCP 80 (http) or 443 (https) ports, traffic from these ports must be allowed. As can be seen in Figure 4, in this case a vulnerability in the web application can be easily exploited without being noticed by the firewall. It is not possible to prevent attacks against web applications with traditional firewalls.

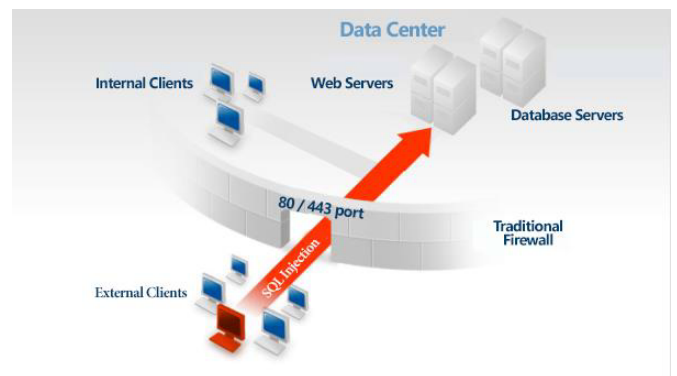


Figure 4: Bypassing the traditional network firewall with web attack [5]

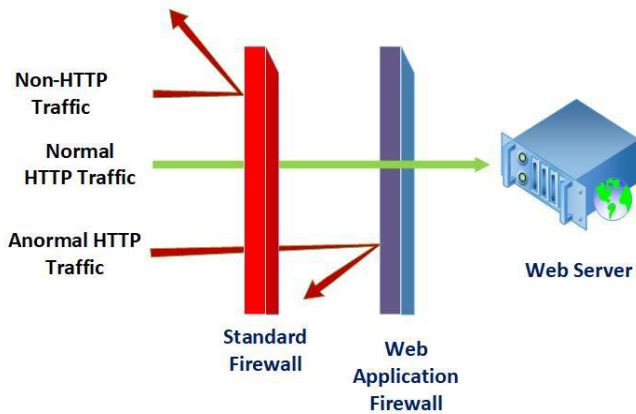


Figure 5: Difference between Standard Firewall and WAF

Today, the next-generation firewall, which serves in the application layer and enables to create protocol-based (http, ftp, snmp, telnet, etc.) rules instead of port numbers, are also insufficient when faced with the vulnerabilities of web applications. In order to prevent the discovery and exploitation of vulnerabilities in the web applications by attackers, it is necessary to use Web Application Firewalls which can analyze web requests deeply and in more detail, detect web attacks, prevent them and report them to system administrators, as shown in Figure 5 [6]. WAFs are able to perform signature-based or anomaly-based intrusion detection. Anomaly-based intrusion detection systems use artificial intelligence methods to create a model from previous attack and normal web traffic and detect whether new incoming web requests carry attack patterns.

III. LITERATURE REVIEW

Within the scope of this study, cyber security area and more specifically, web security studies based on machine learning

and deep learning in the last 5 years have been examined. In this scope, articles, especially indexed, have been included in the review. When the studies on the web security in the literature are examined, there are many current studies using traditional machine learning algorithms [7-15, 17]. Studies using traditional machine learning algorithms related to web security generally show that SQL Injection [7, 9-13, 15] and malicious code analysis [8, 14, 17] have been conducted. While using traditional machine learning algorithms firstly such as SVM, decision tree, NB, KNN in the studies for web security, hidden markov chains, graphs, random trees and static analysis methods are also applied successfully.

Based on the literature review, studies on the network-based intrusion detection systems [18-20] and malware [23-25] are the majority of studies based on deep learning. Although, there are several studies in areas such as web security [21, 26], malicious javascript code analysis [16].

In deep learning-based studies, Auto Encoder (AE) [18, 19, 22, 23], Convolutional Neural Network (CNN) [24-26], Recurrent Neural Network (RNN) [20, 26], Deep Belief Network (DBN) [22], Restricted Boltzmann Machines (RBM) [23], and many deep learning methods have been used successfully, respectively.

Most of the studies (indexed publications) where deep learning methods are applied to cyber security area are composed of 2018 and 2017 studies. Studies that use deep learning approaches in the field of cyber security have been started to be published since 2015 and there is a very new and up-to-date study area. Table 2 shows the comparison of studies based on traditional machine learning and deep learning methods in the field of web security since last 5 years, in terms of algorithm, method and result.

When we look at the above studies, it is seen that they usually deal with one or two types of attacks (Sql injection and/or XSS attack). Our study stands out from the other studies by considering multiple types of attacks, data pre-processing method, developed model and classification success.

Table 2: Some indexed publications on web application security (last 5 years)

Author	Scope	Methodology	Algorithm	Results	Year
Pinzon, C. I. et al. [7]	SQL Injection	Machine Learning	CBR, SVM, Neural networks	99.01%	2013
Goseva-Popstojanova, K. et al. [8]	Malicious Web Traffic	Honeypot, Machine Learning	SVM, J48, PART	98.63%	2014
Oza, A. et al.[9]	HTTP attacks	Static Analyze and Machine Learning	n-gram Hidden Markov Model	99%	2014
Kim, M. Y., & Lee, D. H. [10]	SQL Injection	Machine Learning	SVM	99.6%	2014
Shar, L. K. et al. [11]	SQL injection, XSS, remote code execution and file inclusion	Machine Learning	Hybrid model: logistic regression (LR) and RandomForest (RF)	77%	2015
Choraś, M., & Kozik, R. [12]	detect cyber attacks on web applications	Machine Learning	Graph-based approach	94.46%	2015

Kar, D. et al. [13]	SQL Injection	Machine Learning	Graph of tokens, Support vector machine	99.47%	2016
Medeiros, I. et al. [14]	Detecting and Removing Web Application Vulnerabilities	Machine Learning	C4.5/J48, NB, LR, KNN, MLP, SVM	92%	2016
Kozik, R. [15]	anomaly detection in web layer	Machine Learning	AdaBoost, Random Forest	91.9%	2016
Wang, Y. et al. [16]	Detecting malicious JavaScript	Deep Learning	SAE, LR	95%	2016
Amrutkar, C. et al. [17]	Detecting Mobile Malicious Webpages	Machine Learning	SVM, NB, LR	90%	2017
S. Mohammadi, A. Namadchian [18]	IDS for Network	Deep Learning	AE, Memetic Classifier	%92.72	2017
Yu, Y. et al. [19]	IDS for Network	Deep Learning	Stacking dilated convolutional autoencoders	%96.96	2017
Yin, C. et al. [20]	IDS for Network	Deep Learning	RNN	%83.28	2017
Yan, R. et al. [21]	Detecting code injection attacks	Deep Learning	Abstract Syntax Tree(AST), n-gram, CNN, LSTM	97.55%	2018
Li, Y. et al. [22]	Web Spam	Deep Learning	DAE, DBN	91.62%	2018
Ye, Y. et al. [23]	Malware detection	Deep Learning	Auto Encoder (AE) Restricted Boltzmann Machines (RBM)	%98.82	2018
Nguyen, M. H. et al. [24]	Malware detection	Deep Learning	Control flow graph (CFG), CNN	%95.12	2018
Ni, S. et al. [25]	Malware detection	Deep Learning	SimHash, CNN (image based)	%98.86	2018
Liu, H. et al. [26]	Web payload detection	Deep Learning	CNN, RNN	%96.13	2019

IV. DATASET AND CHARACTERISTICS

The HTTP DATASET CSIC 2010 [27] dataset was used for learning and testing. This dataset was produced by the National Research Council of Spain (CSIC), the Information Security Institute, via recording normal and attack traffic for an e-commerce application developed by themselves. There are not too many up-to-date datasets available for using to test WAFs. Nowadays, DARPA KDD 99 data set is the most known and most used dataset in the test of Intrusion Detection Systems (IDS) [28]. However, this dataset is criticized because it is outdated and does not contain current attack types [29]. Therefore, this dataset is preferred because it is more appropriate to use datasets containing more current types of attacks and traffic data in the studies aimed at detecting web attacks.

The data set contains 44717 http data in total, which is distributed as 20729 normal and 23988 attack traffic. Http requests are labelled as normal or anormal. Anormal labelled http traffic in the data set contains web attack types such as SQL injection, buffer overflow, information gathering, file disclosure, CRLF injection, cross site scripting (XSS), Server Side Include (SSI), parameter tampering and so on [27].

A. Dataset Characteristics

- The original data set was recorded in the extended HTTP / 1.1 (RFC 2616) protocol format.
- Two columns are added to the data set. First of all, it is aimed to follow http packages with the numeric index field. In this way, different records are created by giving the same index value to the packages which have multiple values such as key = value. Secondly, labels are used for classification.

The columns in the data set are: "index", "method", "url", "protocol", "userAgent", "pragma", "cacheControl", "accept", "acceptEncoding", "acceptCharset", "acceptLanguage", "host", "connection", "contentType", "cookie", "payload", "label"

- "index" field is not unique.
- "cookie" and "payload" fields are formatted as KEY = VALUE.
- Empty fields are filled with "null".

V. LSTM BASED WAF MODEL AND RESULTS

Information gain was used for feature selection, so that the most significant 6 properties (method, url, host, content Length, contentType and payload) were used in classification of web traffic.

In this study, a model that can classify attacks on web applications based on LSTM was created. The LSTM architecture is a method located under the RNN architecture that solves the problem of long-term connections in RNN architectures, which is the result of long and short-term memory [30]. In this method, the architecture is the same as the RNN architecture but the structure of the neurons varies. Three gates in each neuron allow for long-term connections with input, output and forget gates [31]. The LSTM architecture is seen in Figure 6.

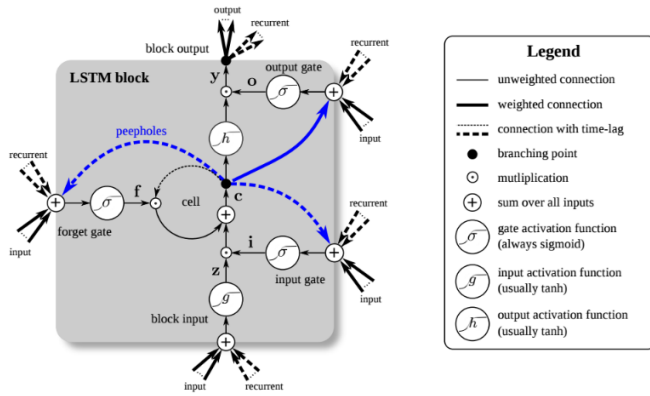


Figure 6: LSTM architecture [31]

The equations of the LSTM units used in the study are given below [30].

Definition of symbols in equations

x_t : input vector
 f_t : forget gate activation vector
 i_t : input gate activation vector
 o_t : output gate activation vector
 h_t : output vector of lstm unit
 c_t : cell state vector
 W, U, b : weight matrices and bias vector parameters

$$f_t = \sigma_c(W_f x_t + U_f h_{t-1} + b_f) \quad (1)$$

$$i_t = \sigma_c(W_i x_t + U_i h_{t-1} + b_i) \quad (2)$$

$$o_t = \sigma_c(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

$$c_t = f_t o_{t-1} + i_t o \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \quad (4)$$

$$h_t = o_t o \sigma_h(c_t) \quad (5)$$

Some artificial neural network architectures, such as LSTM, do not take textual or categorical data as input directly. The data to be input to the LSTM network must be converted to numerical vectors. There are many different word embedding techniques (Word2vec, GloVe, fastText, One hot encoding, etc.) used to convert text-based data into sequential vectors. One hot encoding technique was used in this study. One hot

encoding is a simple and efficient technique to convert word vectors into binary vectors (consisting of 0 and 1) [32].

The features in the dataset were first converted to discrete categorical numbers and then converted into binary sequential vectors with one hot encoding. Numpy, Panda, Scikitlearn libraries and Keras, Tensorflow platforms were used for pre-processing of the dataset, training of the LSTM model and testing of the classification model. 75% of the data set was used for training and the remaining 25% for verification. The model consists of two successive LSTM layers and an output layer that ends with the softmax function. After both LSTM layers, a dropout of 0.3 ratio was applied to prevent over-fitting. The hyperbolic tangent (tanh) activation function was used as an activation function in the LSTM layers. In the optimization of the model, Adam optimization algorithm was used with 0.001 learning rate. Categorical cross entropy loss function was used as loss function. In the training of the model, the batch size to be used for each iteration was set to 16.

The training of the LSTM model, which can classify web requests as attacks or normal, was quickly completed at 5 epoch. Figure 7 shows the classification success of the model on the training and test data. According to this, 99% and 94.41% classification success was achieved in the training and verification set, respectively.

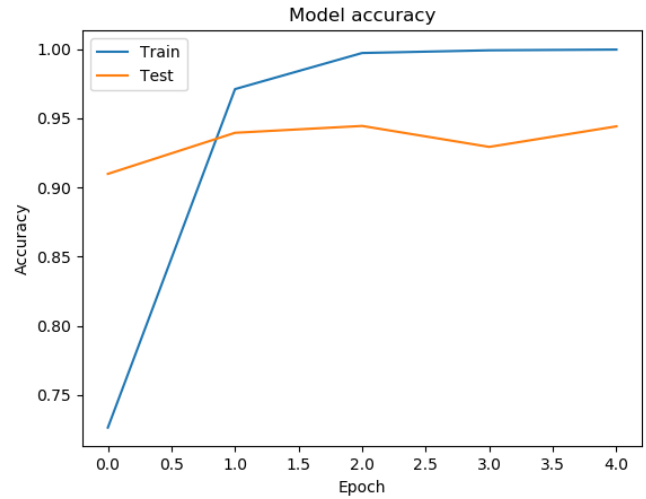


Figure 7: Model classification success

The change in the loss value of the model is shown in Figure 8 and the training process is completed with the loss value of 0.0023.

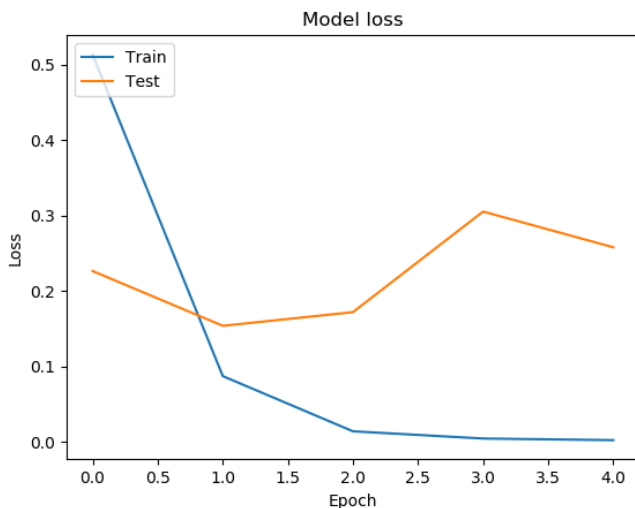


Figure 8: Change in the loss value of the model

VI. CONCLUSION AND DISCUSSION

Within the scope of this study, an anomaly based attack detection model based on LSTM architecture has been developed. With this study, it has been shown that LSTM architecture, which is one of the deep learning methods, will work successfully in web application security area. In future studies, different word embedding techniques and different deep learning architectures can be used to develop models that can better detect anomalies on web traffics. In addition to HTTP requests, it is evaluated that similar studies can be performed on log files related to web applications.

REFERENCES

- [1] Cisco, "IP Next-Generation Network Security for Service Providers", Cisco Public Information White Paper, 2006. Available: https://www.cisco.com/c/dam/en/us/solutions/collateral/service-provider/secure-infrastructure/net_implementation_white_paper0900aecd803fcbbe.pdf [Accessed: 03.02.2019].
- [2] Symantec, "Internet Security Threat Report", 2017. Available: <https://www.symantec.com/content/dam/symantec/docs/reports/istr-22-2017-en.pdf> [Accessed: 04.02.2019].
- [3] EdgeScan, "Vulnerability Statistics Report", 2018. Available: <https://www.edgescan.com/wp-content/uploads/2018/05/edgescan-stats-report-2018.pdf> [Accessed: 04.02.2019].
- [4] OWASP, "OWASP Top 10 Application Security Risks – 2017", 2017. Available: https://www.owasp.org/images/7/72/OWASP_Top_10-2017_%28en%29.pdf.pdf [Accessed: 05.02.2019].
- [5] BGA, "Web Application Firewall Tercih Rehberi", 2013. Available: <https://www.bgasecurity.com/makale/web-application-firewall-tercih-rehberi/> [Accessed: 05.02.2019].
- [6] Tectonic Security, "Web Application Firewall", Available: <http://tectonicsecurity.com/managed-security-services/web-application-firewall/> [Accessed: 04.02.2019].
- [7] Pinzon, C. I., De Paz, J. F., Herrero, A., Corchado, E., Bajo, J., & Corchado, J. M. (2013). "idMAS-SQL: intrusion detection based on MAS to detect and block SQL injection through data mining", *Information Sciences*, 231, 15-31.
- [8] Goseva-Popstojanova, K., Anastasovski, G., Dimitrijević, A., Pantev, R., & Miller, B. (2014). "Characterization and classification of malicious Web traffic", *Computers & Security*, 42, 92-115.
- [9] Oza, A., Ross, K., Low, R. M., & Stamp, M. (2014). "HTTP attack detection using n-gram analysis", *Computers & Security*, 45, 242-254.
- [10] Kim, M. Y., & Lee, D. H. (2014). "Data-mining based SQL injection attack detection using internal query trees" *Expert Systems with Applications*, 41(11), 5416-5430.
- [11] Shar, L. K., Briand, L. C., & Tan, H. B. K. (2015). "Web application vulnerability prediction using hybrid program analysis and machine learning", *IEEE Transactions on Dependable and Secure Computing*, 12(6), 688-707.
- [12] Choraś, M., & Kozik, R. (2015). "Machine learning techniques applied to detect cyber attacks on web applications", *Logic Journal of the IGPL*, 23(1), 45-56.
- [13] Kar, D., Panigrahi, S., & Sundararajan, S. (2016). "SQLiGoT: Detecting SQL injection attacks using graph of tokens and SVM" *Computers & Security*, 60, 206-225.
- [14] Medeiros, I., Neves, N., & Correia, M. (2016). "Detecting and removing web application vulnerabilities with static analysis and data mining", *IEEE Transactions on Reliability*, 65(1), 54-69.
- [15] Kozik, R., Choraś, M., & Hołubowicz, W. (2016). "Evolutionary-based packets classification for anomaly detection in web layer", *Security and Communication Networks*, 9(15), 2901-2910.
- [16] Wang, Y., Cai, W. D., & Wei, P. C. (2016). "A deep learning approach for detecting malicious JavaScript code", *Security and Communication Networks*, 9(11), 1520-1534.
- [17] Amrutkar, C., Kim, Y. S., & Traynor, P. (2017). "Detecting mobile malicious webpages in real time", *IEEE Transactions on Mobile Computing*, 16(8), 2184-2197.
- [18] S. Mohammadi, A. Namadchian, (2017). "A New Deep Learning Approach for Anomaly Base IDS using Memetic Classifier", *International Journal Of Computers Communications & Control*, 12(5), 677-688.
- [19] Yu, Y., Long, J., & Cai, Z. (2017). "Network Intrusion Detection through Stacking Dilated Convolutional Autoencoders", *Security and Communication Networks*.
- [20] Yin, C., Zhu, Y., Fei, J., & He, X. (2017). "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks", *IEEE Access*, 5, 21954-21961.
- [21] Yan, R., Xiao, X., Hu, G., Peng, S., & Jiang, Y. (2018). "New deep learning method to detect code injection attacks on hybrid applications", *Journal of Systems and Software*, 137, 67-77.
- [22] Li, Y., Nie, X., & Huang, R. (2018). "Web spam classification method based on deep belief networks", *Expert Systems with Applications*, 96, 261-270.
- [23] Ye, Y., Chen, L., Hou, S., Hardy, W., & Li, X. (2018). "DeepAM: a heterogeneous deep learning framework for intelligent malware detection", *Knowledge and Information Systems*, 54(2), 265-285.
- [24] Nguyen, M. H., Le Nguyen, D., Nguyen, X. M., & Quan, T. T. (2018). "Auto-detection of sophisticated malware using lazy-binding control flow graph and deep learning", *Computers & Security*, 76, 128-155.
- [25] Ni, S., Qian, Q., & Zhang, R. (2018). "Malware Identification Using Visualization Images and Deep Learning", *Computers & Security*.
- [26] Liu, H., Lang, B., Liu, M., & Yan, H. (2019). "CNN and RNN based payload classification methods for attack detection", *Knowledge-Based Systems*, 163, 332-341.
- [27] Torrano-Giménez, C., Pérez-Villegas, A. and Álvarez Maraño, G. (2010). "An Anomaly-Based Approach for Intrusion Detection in Web Traffic", *Journal of Information Assurance and Security (JIAS)* 5(4): 446-454.
- [28] Tavallaei, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009, July). "A detailed analysis of the KDD CUP 99 data set" In 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications (pp. 1-6).
- [29] McHugh, J. (2000). "Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory", *ACM Transactions on Information and System Security (TISSEC)*, 3(4), 262-294.
- [30] Hochreiter, S. and Schmidhuber, J. (1997). "Long short-term memory". *Neural Computation*, 9(8): 1735-1780.
- [31] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J. (2017). "LSTM: A search space odyssey." *IEEE Transactions on Neural Networks and Learning Systems*, 28(10):2222-2232.
- [32] Brownlee J., "How to One Hot Encode Sequence Data in Python", Available: <https://machinelearningmastery.com/how-to-one-hot-encode-sequence-data-in-python/> [Accessed: 04.02.2019].

Evaluation of Student Academics Performance via Machine Learning Algorithms

A.ÖZTÜRK^{1,2} and M. ÇOBAN¹

¹KTO Karatay University, Computer Engineering Dept., Konya/Turkey, ali.ozturk@karatay.edu.tr

¹ KTO Karatay University, Computer Engineering Dept., Konya/Turkey, mustafacoban994@gmail.com

²Havelsan Inc., Ankara/Turkey, aliozturk@havelsan.com.tr

Abstract – In this study, the factors affecting the academic performance of the students were evaluated by means of different machine learning algorithms. The data set which consists of 131 students was divided into three subsets by grouping the 22 features as educational, economic and social environment. The effects of these feature subsets on exam grades were evaluated via Random Forest (RF), k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), Multilayer Perceptron Neural Network (MLPNN) and Radial Basis Function Network (RBFN) algorithms using 10-fold cross validation. According to the experimental results, using educational features provides highest accuracy on exam results prediction. The algorithm which has the best performance changes depending on the data set and the type output feature. The performance results for the training data set are also given for comparison purposes.

Keywords – Academic Performance Evaluation, Student Data, Machine Learning

I. INTRODUCTION

Prediction of academic performance of the students is important for educational institutions for gaining a better understanding of the factors affecting the success of the students and make meaningful assessments in managerial decisions. Machine learning algorithms are being used by educational institutions to improve the services they provide to the students in order to increase their success, measure their retention, manage fund grant [1] or predict students' adequacy for admission [2]. On the other hand, student profiling [3] deals with grouping students into categories using specific features like exam grades, project scores, attendance or quiz results. The institutions can use the grouping analysis to predict the academic trends and patterns by helping to gain better insight on learning manners of the students [4].

There are various studies in the literature on academic performance prediction with machine learning algorithms. Ahmad et al. compared Decision Tree (DT), Naïve Bayes (NB) and Partial Tree (PART) algorithms to obtain best prediction model [5]. They collected the data during 8 year period for the first year bachelor students taking Computer Science course. The data set consisted of demographic, previous academic records and social information of the students. They found that the Partial Tree algorithm has the highest accuracy value of %71.3. Khasanah and Harwati [6] conducted feature selection to select high influence features in

determining the student academic performance by comparing two classification algorithms, namely Bayesian Network and Decision Tree. They found that student attendance and GPA were the most influential features and Bayesian Network had higher accuracy than Decision Tree. Almarabeh [7] analyzed and evaluated the performance of university students using DT, MLPNN, NB and Bayesian Network algorithms within Weka tool [8]. According to his results, Bayesian Network had the highest accuracy of %92. The dataset used in his study included 225 instances with 10 features. Hussain et al. [9] collected data from three different colleges consisting of 22 features of 300 students. They evaluated J48, PART, RF and Bayesian Network classifiers within Weka. They applied feature selection on the data set beforehand to find the most relevant features. The RF algorithm had the highest accuracy of %99 when they used 12 features after feature selection. The accuracy dropped to %84.33 when all the features were used. The Bayes Net had the worst performance of %65.33. However, they did not use k-fold cross validation on the dataset during their evaluation. They obtained their results using the whole data set with only training data set option in Weka without using any separate test set.

In our study, the dataset used in [9] was divided into 3 different subsets depending on social, economic and educational levels of students. According to our analysis, the most influential factor on the academic performance of the students was found as the social features. Five different machine learning algorithms, e.g. RF, k-NN, SVM, MLPNN and RBFN, were used for the evaluation. We obtained our results by applying 10-fold cross validation on the data set. We also evaluated the algorithms using the whole data set for comparison purposes.

II. MATERIAL AND METHODS

A. Data Set

The features of the dataset used in this study and their corresponding values are given in Table 1. The explanations of the features which are used for classification of the students are given in the following.

- **TNP:** Nationwide examination in Assam, India [9] for determining the academic level of the Class 10 students. The possible values can be "Best" for above %80, "Very Good" for between %80 and %60,

“Good” for between %60 and %45, “Pass” for between %45 and %30, “Fail” for below %30.

- **TWP:** Same as TNP but it is applied to the students of Class 12.
- **IAP:** Internal Assessment Percentage which consists of the evaluation driven by term examinations, quizzes, homework assignments etc.
- **ESP:** End of Semester Percentage which determines the success of the student at the end of each semester. The output value for TWP, IAP and ESP is calculated as in TNP.

Table 1: The input features of the data set.

Feature	Description	Possible Values
GE	Gender	{M,F}
CST	Social Caste	{G,ST,SC,OBC,MOBC}
ARR	Arrear (Failed) Paper	{Y,N}
MS	Marital Status	{Married,Unmarried}
LS	Lived in Town or Village	{T,V}
AS	Admission Status	{Free,Paid}
FMI	Family Income	{Vh,High,Am,Medium,Low}
FS	Family Size	{Large,Average,Small}
FQ	Father Qualification	{II,Um,10,12,Degree,Pg}
MQ	Mother Qualification	{II,Um,10,12,Degree,Pg}
FO	Father Occupation	{Service,Business,Retired,Others}
MO	Mother Occupation	{Service,Business,Retired,Others}
NF	Number of Friends	{Large,Average,Small}
SH	Study Hours	{Good,Average,Poor}
SS	Student Shool	{Govt,Private}
ME	Education Language	{Eng,Asm,Hin,Ben}
TT	Travel Time	{Large,Average,Small}
ATD	Attandance Percentage	{Good,Average,Poor}

We divided the whole dataset into 3 different subsets, e.g. Economic, Social and Educational, depending on the category of the input features as given in Table 2.

Table 2: The categories of the input features.

Category	Input Features
Economic	AS, FMI, FO, MO, SS, TT
Social	CST, MS, LS, FS, FQ, MQ, FO, MO, NF, SH, ME, TT
Educational	TNP, TWP, IAP, ESP, ARR, AS, ME

The classification performances of the algorithms were evaluated according to the output features TNP, TWP, IAP and ESP. These features were also used as input features for the Educational data subset evaluation.

III. THE CLASSIFICATION ALGORITHMS

A. Random Forests

Random Forests is an integrated classification algorithm which is formed by a decision tree classifiers set $\{h(x, \Theta(k)), k = 1, 2, 3 \dots K\}$. Here, x is the input vector and $\Theta(k)$ is a random vector having equal distribution which independently determines the growing process of a single tree. If A is continuous attribute and m is the sample subsets on a node, then the random forest algorithm for A is

- The samples are sorted in ascending order on continuous attribute A using corresponding discrete sequence $\{A_1, A_2, \dots, A_m\}$.
- On the sequence, $m-1$ division points are generated. The division point j ($0 < j < m$) is adjusted by the

formula $\frac{w_1 - (A_j + A_{j+1})}{2}$ which divides the sample set on node A into subsets $\{s \mid s \in S, A(s) \leq W_j\}$ and $\{s \mid s \in S, A(s) > W_j\}$.

- The Gini coefficients of $m-1$ divided points are calculated as in (1) and the points having minimum Gini coefficients are selected to divide the sample set.

$$Gini(S) = 1 - \sum_{n=1}^n p_i^2 \quad (1)$$

Here S is the sample set and $|S|$ is the number of samples. The number of samples belonging to class C_i is $|C_i|$ and the probability p_i is $\frac{|C_i|}{|S|}$ [10].

B. k Nearest Neighbors (kNN)

kNN is an instance based learning algorithm [11] which defines each sample x as an n -dimensional feature vector as $\langle a_1(x), a_2(x), \dots, a_n(x) \rangle$ where n is the number of features. The distance between X_i and X_j is calculated as in (2).

$$d(x_i, x_j) \equiv \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (2)$$

When a new instance is to be classified, the nearest k samples are selected from the data set as $\langle x, f(x) \rangle$ and the equation in (3) is used for classification.

$$f(y) \leftarrow \arg \max_{v \in V} \sum_{i=1}^k \delta(v, f(x_i)) \quad (3)$$

Where, V is the set of nearest neighbors. The k which is the number of nearest neighbors is specified by the user.

C. Support Vector Machines (SVM)

SVMs map input dataset to a high dimensional feature space using a nonlinear mapping function. The estimation function is

$$f(x) = (w \times \phi(x)) + b \quad (4)$$

Where, w and b are the coefficients which are estimated from data set and $\phi(x)$ is the non-linear function in feature space.

The risk function to be minimized is

$$R(w, \xi^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (5)$$

$d_i - w\phi(x_i) - b_i \leq \xi_i$, $(w\phi(x)) + b - d_i \leq \xi_i^*$, where $\xi_i, \xi_i^* > 0$.

The Support Vector estimation function is

$$f(x) = \sum_{i=1}^{NSV} (\alpha_i - \alpha_i^*) K(X, X_i) + b \quad (6)$$

Where α_i and α_i^* are Lagrange coefficients and NSV is the number of support vectors.

The kernel function $K(X_i, X_j) = \phi(X_i)\phi(X_j)$ is selected as the Pearson VII function-based kernel (PUK) [12] in this study.

The C , ϵ and kernel parameters σ and ω are specified by the user.

D. Multilayer Perceptron Neural Network (MLPNN)

Multilayer Perceptron Neural Networks with back propagation are feed-forward networks where modification on the weights is based on the difference between the computed and actual values of the output nodes [13]. The error is computed as in the following

$$E = \frac{1}{2} \sum_j (x_o - x_t)^2 \quad (7)$$

Where x_o is the output value and x_t is the target value.

The output value for neuron x_j is calculated as

$$X_j = \theta(\sum_i^n x_i w_{ij}) \quad (8)$$

Where θ is the transfer function for which sigmoid (9) is used in this study.

$$\theta(x) = 1 / (1 + e^{-x}) \quad (9)$$

The back propagation algorithm uses gradient descent in which the weights are updated according to the following formula

$$\Delta W_{ji}(n+1) = \eta \delta_{pj} O_{pi} + \alpha \Delta W_{ji}(n) \quad (10)$$

Where η is the learning rate, α is the momentum coefficient and δ_{pj} is the error value for the neuron on L^{th} layer.

The learning rate, momentum coefficient and number of hidden neurons are specified by the user.

E. Radial Basis Function (RBF) Network

RBF networks are feed-forward neural networks which have a single hidden layer of nonlinear units [14]. The activation function of the hidden neurons was Gaussian in this study.

The radial distance d_i between the input vector u and the center of the basis function c_i is computed for hidden unit i as in the following

$$d_i = \|u - c_i\| \quad (11)$$

The output h_i of each hidden unit i is then computed by applying the basis function G to this distance

$$h_i = G(d_i, \sigma_i) \quad (12)$$

The output for the j^{th} unit in the output layer is computed as,

$$o_j = f_j(u) = w_{0j} + \sum_{i=1}^L w_{ij} h_i, \quad j = 1, 2, \dots, M \quad (13)$$

Until minimizing the following function

$$J(w, c) = \sum_{k=1}^K \|y^k - f(u^k)\|^2 \quad (14)$$

Where u^k and y^k are input and output training samples, respectively.

The number of hidden neurons is specified by the user.

IV. EXPERIMENTAL RESULTS

Weka tool [8] was used for the implementation of the classifiers. The overall dataset was divided into 3 subsets as given in Table 2. First, the algorithms were trained by using these subsets and compared according to the training results. Then, they were evaluated using 10-fold cross validation.

The training performance results for the dataset with social input features are given in Table 3. According to the results, RF and SVM algorithms outperformed the others except for TWP output which is the same with MLPNN. The k-NN is the worst one among the algorithms.

Table 3: Training performance of algorithms on social input features.

Algorithm	TNP	TWP	IAP	ESP
RF	100%	99,24%	100%	100%
kNN	66,41%	74,80%	69,47%	67,94%
SVM	100%	99,24%	100%	100%
MLPNN	99,24%	99,24%	99,24%	96,18%
RBFN	88,55%	84,73%	87,03%	95,41%

The training performances of the algorithms on social input features are given in Figure 1.

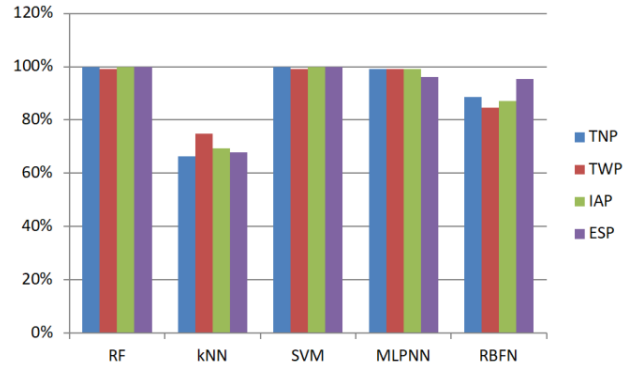


Figure 1: The performance of the algorithms on training data set with social input features.

Table 4 summarizes the training performances of the algorithms for the data set with economic input features. The RF algorithm outperformed the other algorithms in terms of TNP and TWP outputs. The performances of RF and MLPNN are the same for IAP and ESP outputs. The RBFN is the worst one among the algorithms.

Table 4: Training performance of algorithms on economic input features.

Algorithms	TNP	TWP	IAP	ESP
RF	80,92%	83,97%	80,92%	79,39%
kNN	63,36%	58,78%	68,70%	64,89%
SVM	57,25%	54,96%	61,83%	56,49%
MLPNN	80,15%	82,44%	80,92%	79,39%
RBFN	51,15%	53,43%	54,97%	54,20%

In Figure 2, the training performances of the algorithms on the dataset with economic input features are demonstrated.

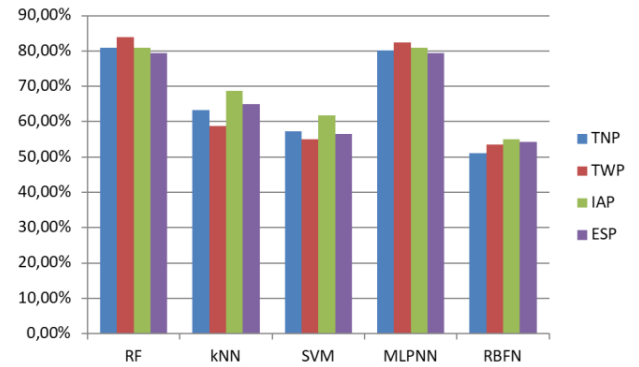


Figure 2: The performance of the algorithms on training data set with economic input features.

The training performances of the algorithms on the dataset with educational input features are given in Table 5. According to the results, the performances of RF, SVM and MLPNN are the same for TNP and ESP outputs. The RF algorithm outperformed the others for TWP output and has the same performance with SVM for IAP output. The RBFN is the worst one for TNP output and k-NN is the worst for all the other outputs.

Table 5: Training performance of algorithms on educational input features.

Algorithms	TNP	TWP	IAP	ESP
RF	87,02%	83,97%	81,68%	87,02%
kNN	72,52%	74,81%	68,70%	74,05%
SVM	87,02%	83,21%	81,68%	87,02%
MLPNN	87,02%	83,21%	80,92%	87,02%
RBFN	64,89%	78,63%	74,05%	83,97%

Figure 3 demonstrates the training performances of the algorithms on the dataset with educational input features.

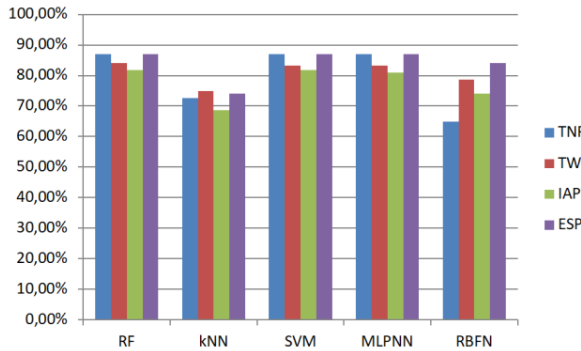


Figure 3: The performance of the algorithms on training data set with educational input features.

The evaluations of the algorithms on 3 different data subsets using 10-fold cross validation were also performed. The accuracy of the algorithms degraded significantly as expected which is due to the small data set size. In Table 6, the classification results for the dataset with social input features are given. The RF gave the best accuracy with %60,31 for TWP output. The worst performance belongs to MLPNN with %40,46 for TNP output.

Table 6: The performance of algorithms on social input features with 10-fold cross validation.

Algorithm	TNP	TWP	IAP	ESP
RF	47,33%	60,31%	48,86%	46,57%
kNN	46,57%	54,97%	53,44%	49,62%
SVM	45,04%	48,86%	48,86%	43,51%
MLPNN	40,46%	45,80%	53,44%	52,67%
RBFN	48,85%	51,14%	49,62%	46,56%

Figure 4 gives the performances of the algorithms on the dataset with social input features using 10-fold cross validation.

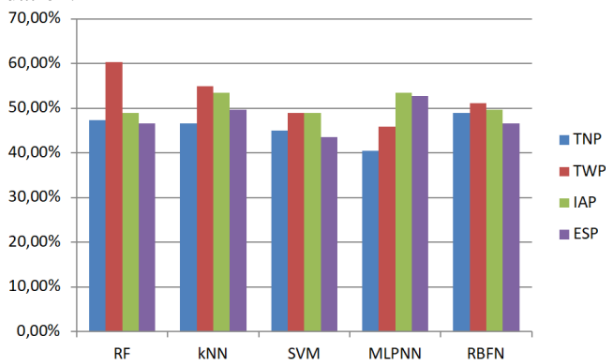


Figure 4: The performance of the algorithms on data set with social input features using 10-fold cross validation.

The classification results for the dataset with economic input features with 10-fold cross validation are given in Table

7. The best performance belongs to kNN with %51,15 for TWP output. The MLPNN has the worst performance with %40,46 for ESP output.

Table 7: The performance of algorithms on economic input features with 10-fold cross validation.

Algorithm	TNP	TWP	IAP	ESP
RF	47,33%	48,85%	46,56%	45,04%
kNN	48,86%	51,15%	45,04%	45,80%
SVM	45,04%	50,38%	49,62%	41,22%
MLPNN	45,80%	43,51%	44,27%	40,46%
RBFN	45,04%	49,62%	51,14%	45,80%

The performances of the algorithms on the dataset with economic input features using 10-fold cross validation are given in Figure 5.

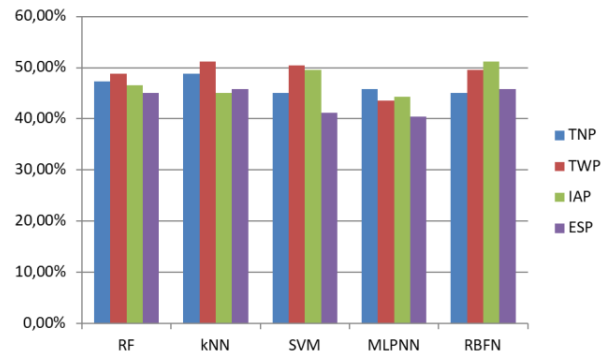


Figure 5: The performance of the algorithms on data set with economic input features using 10-fold cross validation.

Table 8 includes the classification results for the dataset with educational input features with 10-fold cross validation. The RF gave the best performance with %71,76 for TWP output. The SVM has the worst performance with %48,09 IAP output.

Table 8: The performance of algorithms on educational input features with 10-fold cross validation.

Algorithm	TNP	TWP	IAP	ESP
RF	58,78%	71,76%	53,44%	70,99%
kNN	61,01%	64,89%	51,91%	68,70%
SVM	62,60%	63,36%	48,09%	69,47%
MLPNN	55,73%	61,07%	50,38%	64,89%
RBFN	58,01%	61,06%	57,25%	68,70%

The performances of the algorithms on the dataset with educational input features using 10-fold cross validation are given in Figure 6.

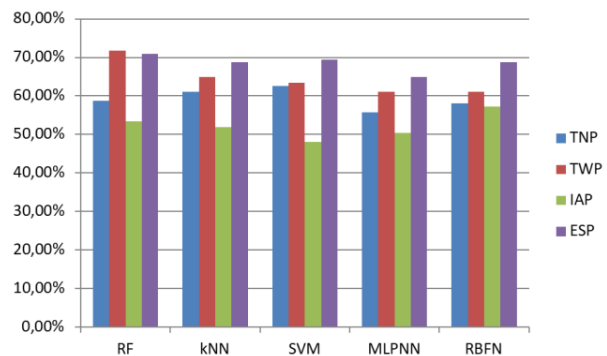


Figure 5: The performance of the algorithms on data set with educational input features using 10-fold cross validation.

The MLPNN implementations used in this study had a single layer with different number of hidden neurons for different data subsets. The number of hidden neurons ranges from 3 to 12 which gave the best results. The presented results above were written for the best accuracies. The learning rate and momentum were 0.3 and 0.2, respectively. For the SVM algorithm, we used PUK as the kernel function where σ and ω were set between 0.5 and 1.0, and the C and ϵ were set 1.0 and $1.0e^{-3}$, respectively. We tried various k values for k-NN and different k values gave the best result for different data subsets. The bag size for RF algorithm was empirically set to a value between %10 and %90 of the training set size in order to find the bag size which gave the best accuracy for RF. Different numbers of hidden neurons were tried for the RBFN for different data subsets. For example, we obtained the best result when we set the number of hidden neurons to 2 for TWP output on economic inputs dataset. The best result for IAP output was obtained with 5 hidden neurons on social and economic inputs dataset. However, the highest accuracy for IAP output on educational inputs dataset was obtained with 2 hidden neurons.

V. CONCLUSION

In this study, the classification performances of different machine learning algorithms were evaluated on the dataset which was used in [7]. The dataset was shared on [15] by the authors. In their study, the authors obtained their classification results using training session only. We made further analysis on the dataset using 10-fold cross validation and observed that the classification results severely degrade which is due to the small dataset size. We also divided the dataset into three subsets by grouping related features as social, economic and educational. We obtained the classification results using these feature subsets and identified their relation to the student academic performance on four different examinations.

REFERENCES

- [1] Y. Zhang, S. Oussena, T. Clark, and H. Kim, "Use Data Mining to Improve Student Retention in Higher Education – A Case Study." in Proc. of the 12th International Conference on Enterprise Information Systems, 2010, pp. 190-197.
- [2] E. P. I. García, and P. M. Mora, "Model Prediction of Academic Performance for First Year Students", 10th Mexican International Conference on Artificial Intelligence, 2011, pp. 169-174.
- [3] P. Nithya, B. Umamaheswari and A. Umadevi, "A Survey on Educational Data Mining in Field of Education", International Journal of Advanced Research in Computer Engineering & Technology, vol. 5, no.1, pp.69-78, 2016.
- [4] S. Parack, Z. Zahid, and F. Merchant, "Application of data mining in educational databases for predicting academic trends and patterns", IEEE International Conference on Technology Enhanced Education (ICTEE), 2012, pp. 1-4.
- [5] F. Ahmad, N.H. Ismail and A.A. Aziz, "The prediction of students' academic performance using classification data mining techniques," *Applied Mathematical Sciences*, vol. 9, no. 129, pp. 6415–6426, 2015.
- [6] A. U. Khasanah and Harwati, "A comparative study to predict student's performance using educational data mining techniques," in *IOP Conference Series: Materials Science and Engineering*, 2017, pp. 1-7.
- [7] H. Almarabeh, "Analysis of students' performance by using different data mining classifiers," *International Journal of Modern Education and Computer Science*, vol. 9, no. 8, p. 9-15, 2017.
- [8] Weka: Data Mining Software in Java, University of Waikato [Online]. Available: <http://www.cs.waikato.ac.nz/ml/index.html>.
- [9] S. Hussain, N. A. Dahan, F. M. Ba-Alwi, and N. Ribata, "Educational data mining and analysis of students' academic performance using weka," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 9, no. 2, pp. 447–459, 2018.
- [10] Y. Xu, "Research and Implementation of Improved Random Forest Algorithm Based on Spark", IEEE 2nd International Conference on Big Data Analysis, pp. 499–503, 2017.
- [11] D. W. Aha, D. Kibler and M. K. Albert, "Instance-based learning algorithms", *Machine Learning*, 6: 37- 66, 1991.
- [12] B. Üstün, W.J. Melssen and L.M.C. Buydens, "Facilitating the application of support vector regression by using a universal Pearson VII function-based kernel", *Chemometrics and Intelligent Laboratory Systems*, 2006; 81: 29-20.
- [13] E. Alpaydm, "Introduction to Machine Learning. 2nd Ed", Cambridge, MA, USA: MIT Press, 2010.
- [14] U. Halici, "Artificial Neural Networks – Lecture Notes", Ankara, 2004.
- [15] M. Lichman, UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science. 2013.

Criminological Evaluation of Cyber Attacks on Information and Network Security

Halil Yalcin AKDENIZ

Eskisehir Osmangazi University, Eskisehir/Turkey, hyakdeniz26@gmail.com

Abstract – With the advancing technology, the use of internet and connected virtual networks has become indispensable. However, the internet is not as innocent as when it first appeared. At this point, information technologies bring along some important risks and threats besides making the benefits and life easier. As a result of cyber attacks, private institutions, government agencies and enterprises can face very heavy costs in terms of economic, strategic and political aspects. In this study, cybercrime, attacks and living under the relevant information and network security threats and future living to the world regarding the possible criminological danger and Turkey have also discussed the measures to be taken is made a general assessment of the situation.

Keywords – Cyber Security, Information Technology Management and Control, Information and Network Security, Cryptology

I. INTRODUCTION

Computer, communication, transport and financial systems all depend on a few degree of security for their operation. Internet with billions of devices as a network of networks where millions of networks are connected information on the framework of specific protocols it allows exchange [1]. Internet communication, information exchange and computer the connection between the use in the world is rapidly spreading

45% of the world's population are now social media users: a whopping 3.5 Billion people. Number of people connected to the Internet: 4.38 Billion, Number of people actively using social media: 3.484 Billion, Number of mobile device users: 5.112 Billion and Number of users using social media via mobile device: 3.256 Billion. [2]



Figure 1: Current Status of Internet Use in 2019 [2]

Recent developments in information systems technologies automation in many applications in areas. Due to the fact that the data in the enterprises become a very important source, data access, data sharing, data creation and The use of information is an important requirement. Data and In addition to increasing demand for information management, the security of applications and information systems is critical. Data and information against unauthorized access as well as minimal corruption need to be protected. As a result of the rapid dissemination of the Internet, many people have access to the data. Hence the data and the need for effective mechanisms for the protection of applications are heard [3]. Increased use of the Internet through the rapid development of information technologies; it has become an indispensable element of life in the public, private sector and even on a personal scale. Information, network and the Internet to become widespread in the world is a while creating unlimited freedom on the other hand; abuse of information, network and system technologies due to security vulnerabilities and in this sense a crime mechanism it causes it to become [4-5].

In this work, cybercrime, attacks and living under the relevant information and network security threats and future living to the world regarding the possible criminological danger and Turkey have also discussed the measures to be taken is made a general assessment of the situation.

II. NETWORK AND CYBER SECURITY

Historically, The term cyber comes from its cybernetic origin. It was first used in 1958 by Louis Couffignal, the father of the science of cybernetics, who studied the communication discipline between living things and / or machines. "Computer network", "Internet belonging", "Virtual Reality" means the meaning. The term cyber security or information security was first introduced by engineers in the 1990s. This has been suggested in networked computers.

Cyber space word interconnected system, software, hardware and people communication and / or interaction with the abstract or concrete used to describe the area. Cyber Attack target selected persons, companies, institutions, organizations and states information and transmission systems and critical infrastructures of buildings planned and coordinated attacks [6].

Information security management system in the manner specified in integrity and accessibility needs to be provided. Information access only The use by the persons given is his privacy. This access to written information or information systems information stored by authorized persons only access. Even if sensitive information is present The authorities should be informed. The integrity of the information is unmodified, partial or complete. Accessibility is the authority of stored information when necessary This is because it is accessible to people [7].

Cyber threats now only to computer systems damage to systems (infiltration into systems, information from systems) and to provide false information to systems. One critical communication of the country systems, computer systems, energy and transportation networks, to the extent necessary to damage military command and control systems, It appears as an asymmetric war pattern. Cyberspace one of the major threats in the coming years the idea that; all started [8]

For this reason, it is very important to construct more effective defense systems and to prepare emergency preparedness. Immediate detection of attacks, virtual or physical barrier building, development of regional and national cyber security policies it has become mandatory [9-10].

III. CYBER CRIME AND CYBER ATTACKS

The aim is to attack individuals or individual groups in order to damage the dignity of the victim or to cause the victim to be physically or mentally harmed, directly or indirectly, in the Internet (interview room, emails, advertisements and communities) and mobile phones (SMS / MMS). [11]. Such offenses may also pose a threat to the security and economic integrity of a nation. The outlook for this type of crime is the desire to gain high returns, especially in the breakdown of software CDs, copyright violations, child abuse, and other forms of child abuse. In the event of loss of confidential information or illegitimate information, there is a crime of violation of privacy.

In the international arena, both state and non-state players engage in cybercrime and other cross-border crimes in areas such as intelligence, financial thefts. Acts that transcend international borders and concern the interests of at least one nation-state are considered cyber warfare. International legal systems try to make these players accountable in the international crime court for their actions. [12]

Cyber crimes include a wide range of acts. Usually, although, the crimes directly targeting the computer; Although the crimes committed by using computer networks and devices are divided into two basic categories, the primary target is the freedom of the computer network or device.

i. First offenses targeting computer network and devices: Computer viruses, Denial of service attacks, Malware

ii. Offenses committed using computer networks and devices: Cyber monitoring, Fraud and identity theft
Information playback, Playing personal network information (User name, login password, credit card information, etc.)

- Sending unwanted messages

Either sending unintended bulk of messages for commercial purposes is considered an illegal act in the legal system of some states. While legal regulations against mass messages are new, they limit the existence of unwanted electronic communications. [13]

- Fraud

Informatics and virtual fraud is misrepresenting the fact that the truth is the result of any dishonest act that causes another person to do or not to do an action. Fraud is defined in Article 157 of the Turkish Criminal Code as the fraudulent act of cheating a person and providing a benefit to him or to others as to his or others' detriment. In Article 158 of the Turkish Penal Code, information systems are considered to be in the scope of Qualified Fraud if processed easily and it is stated that the penal sanction is between two years and seven years in prison. In this context, fraud will result in the following:

Replacing the computer input with an unauthorized path. This act requires a certain amount of technical knowledge and is not an ordinary form of theft because the operators change the information before the entry or enter incorrect information or enter the instructions that they have not been authorized or do not perform.

Replacing, destroying, decelerating or stealing the printout with often unauthorized transactions.

Changing or deleting stored information.

Replacement or misuse of existing system tools or software packages, or writing or replacing passwords for fraud.

Other forms of fraud, such as bank fraud, identity theft, extortion, the stealing of customized information, can be achieved through the use of information systems. [14]

- Obscene or offensive content

Network pages and other electronic communication tools may contain unpleasant, obscene, or many offensive content. In some cases, these communications may be illegal. There are many jurisdictions stating the situations in which freedom of expression has been exceeded and prohibiting racism, profanity and insult, politically damaging, defamatory or defamatory, provocative or hateful content of hate crimes.

The content of such illegal communication may vary according to states and nations. In internet pornography, the most severe sanctions are aimed at preventing sexual abuse of minors. [14]

- Cyber Terrorism

State resources and Information and Virtual Technologies security experts have documented a significant increase in Internet issues and server scans since 2001. However, the vast majority of these violations were part of the actions of organized intelligence, foreign intelligence agencies or other groups that wanted to mapping the existing security gaps of key systems. [15]

- Social Media Networks

Services that allow you to connect with other people of similar interests and background. Usually these consist of a profile, various ways to interact with other users, ability to setup groups, etc. The most popular amongst them are facebook, whatsapp, instagram, linked-in, via Skype, Yahoo Messenger, Google Talk and so on. A computer can prove the crime again by another computer. In case of any offense, the user who uses the computer in addition to the crime he committed will also have an IT crime. Because it used a public infrastructure to commit crimes. [16]

Cybercrime crimes in New York City of the United States are classified as C class first-class crimes and punishments of imprisonment ranging from three to 15 years are imposed on short-term prison sentences and fines. [17]

In Turkey, combating such crimes enacted in the year 2007 5651 titled "Regulation of publications on the internet and the law on combating crimes committed by these publications" in accordance maintained, thereby blocking access, providing monitoring content provider, obligations of the location providers and access providers is held. [18]

The new Turkish Criminal Law No. 5237 also regulates the crimes related to cybercrime, for example, the crime of defamation offense regulated in Article 125 is punishable by imprisonment from three months to two years. less than a penalty can be given. [19]

A similar situation is theft, fraud etc. For the offenses it has been introduced as aggravating cause. Furthermore, Article 243 of the Turkish Penal Code contains sanctions for cyber crimes. Accordingly, the penalties vary between 6 and 5 years. [20]

IV. CONCLUSION

In this academic study, a general evaluation has been made with the examples of criminological information and network security and cyber attacks and related punishments in Turkey and in the world. Several studies have been conducted in the general literature on cyber security and it is foreseen to be made in the future. Joint efforts are needed to raise awareness of the current danger of persons, institutions and nations, and the greater danger that is approaching in the future. The sanctions against cyber insecurity in the virtual environment and the resulting unlawful actions must be more deterrent. In Turkey and in the world Companies and nations must do more financial and administrative spending on cyber defense technology and security strategy. They should not forget that in the future such attacks and illegal transactions may face more of their current financial and administrative expenses.

REFERENCES

- [1] R. A. Clarke, R.K. Knake, "Cyber War-The Next Threat to National Security and What to Do About It", New York: HarperCollins
- [2] <https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates>
- [3] Kumar, V., Srivastava, J., ve Lazarevic, A. (2005). Managing cyber threats: issues, approaches, and challenges. Springer.
- [4] N. F. Doherty, L. Anastakis, H. Fulford, "The Information Security Policy Unpacked: A Critical Study of the Content of University Policies", International Journal of Information Management, 29(6), 449-457, 2009.
- [5] F. Aslay "Cyber Attack Methods and Turkey's Cyber Security Situational Analysis", p. 5, 2017
- [6] B. Alaca, (2008) "Cyber Crimes in Turkey and the Impact of the Internet on Crime (with Anthropological and Legal Aspects)", Ankara University Institute of Social Sciences, Department of Anthropology, Master Thesis. Ankara.
- [7] F. Aslay "Cyber Attack Methods and Turkey's Cyber Security Situational Analysis", p. 5, 2017
- [8] A. Aytekin. "Evaluation of Turkey's Cyber Security Strategy and Action Plan", Unpublished Master Thesis, Information Systems Department, Gazi University, 2015.
- [9] S. E. Goodman, "Critical Information Infrastructure Protection", Terrorism (Ed.), Responses to Cyber Terrorism NATO Science for Piece and Security, IOS Press (Volume 34), Ankara, 25, 2008.
- [10] F. Aslay "Cyber Attack Methods and Turkey's Cyber Security Situational Analysis", p. 5, 2017
- [11] Moore, R. (2005) "Cyber crime: Investigating High-Technology Computer Crime," Cleveland, Mississippi: Anderson Publishing.
- [12] [Cyber Warfare And The Crime Of Aggression: The Need For Individual Accountability On Tomorrow's Battlefield](#)". Law.duke.edu.
- [13] Telephone Consumer Protection Act of 1991, Do-Not-Call Implementation Act of 2003, CAN-SPAM Act of 2003.
- [14] Aydin, E. D. (1992b). Introduction to cybercrime and law. Ankara: Doruk Publications.
- [15] Informatics Association of Turkey (2010). "The final report of ethics of information", http://www.tbd.org.tr/usr_img/cd/kamubib14/raporlarPDF/RP2-2011.pdf
- [16] D. R. Pahuja, "IMPACT OF SOCIAL NETWORKING ON CYBER CRIMES : A STUDY", s. 6.
- [17] Computer fraud charges in New York. May 2011. Bukh Law Firm, PC - 14 Wall St, New York NY 10005 - (212) 729-1632
- [18] Turkish Penal Code, "EN LAW ON THE ORGANIZATION OF PUBLICATIONS IN INTERNET ENVIRONMENT AND COMBATING WITH THE CRIMES PROTECTED BY THESE PUBLICATIONS" <http://www.mevzuat.gov.tr/MevzuatMetin/1.5.5651.pdf>
- [19] Turkish Penal Code, Article: 5237, <https://www.tbmm.gov.tr/kanunlar/k5237.html>
- [20] Turkish Penal Code, Article: 243 <https://www.tbmm.gov.tr/kanunlar/k5237.html>

Digital Forensics in Social Media; Evaluation in the Context of System, Informatics, Network and Cyber Security

Halil Yalcin AKDENIZ

Eskisehir Osmangazi University, Eskisehir/Turkey, hyakdeniz26@gmail.com

Abstract – Thanks to the globalization of technology, the number of social media users will continue to increase. Through social media, people have been able to communicate with each other, with people they do not know, with celebrities or organizations in a shorter time than in the past. Social media is a more free, faster and more open environment than other media. In recent years, social media platforms such as Facebook, Twitter, Youtube can be used for managing the masses and creating a perception and creating a tendency for this perception. However, this is sometimes not a freedom, but it can become a social crime mechanism. In this study; based on statistics and examples in the world and Turkey, If it is not managed well, how dangerous and criminal forensic cases will be dangerous are evaluated in terms of system, informatics, network and cyber security.

Keywords – Cyber Security, Social Media Crimes, Information Security and Cryptology, Digital Forensics, Information and Network Security

I. INTRODUCTION

Recent developments in information systems technologies automation in many applications in areas. Due to the fact that the data in the enterprises become a very important source, data access, data sharing, data creation and The use of information is an important requirement. Data and In addition to increasing demand for information management, the security of applications and information systems is critical. Data and information against unauthorized access as well as minimal corruption need to be protected. As a result of the rapid dissemination of the Internet, many people have access to the data. Hence the data and the need for effective mechanisms for the protection of applications are heard [1]. Increased use of the Internet through the rapid development of information technologies; it has become an indispensable element of life in the public, private sector and even on a personal scale. Information, network and the Internet to become widespread in the world is a while creating unlimited freedom on the other hand; abuse of information, network and system technologies due to security vulnerabilities and in this sense a crime mechanism it causes it to become [2-3].

In this work, based on statistics and examples in the world and Turkey, If it is not managed well, how dangerous and criminal forensic cases will be dangerous are Digital Forensics in Social Media; Evaluation in the Context of System, Informatics, Network and Cyber Security

II. SOCIAL MEDIA AND CRIME

In 2019 social media statistics, this number reached 3.48 billion. 45% of the world's social media users and mobile social media users reached 3.2 billion. [4]

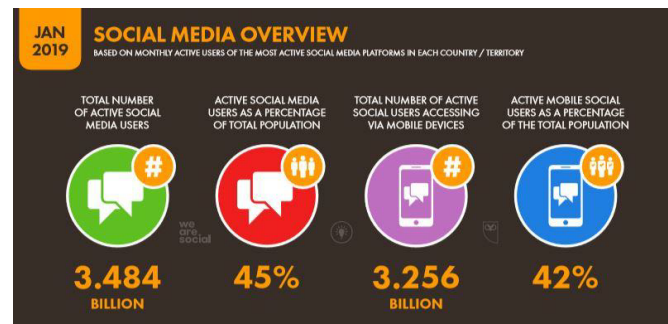


Figure 1: Using Social Media in the World, in 2019 [4]

According to 2019 social media usage statistics, Facebook ranks first with 2.27 billion users. The second most used platform is Youtube, followed by Instagram. [4]

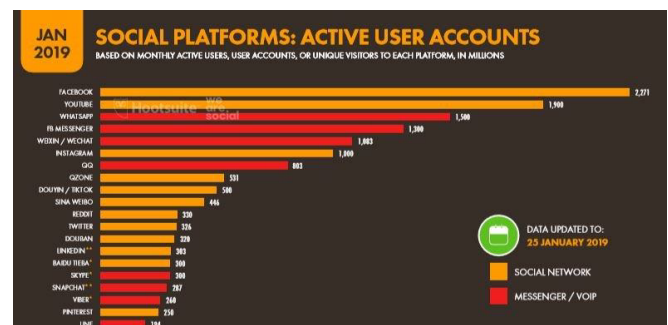


Figure 2: Using Social Media Platforms in the World, 2019 [4]

Turkey has a population of 82.4 million, There are 59.36 million Internet users, 72% of the population, There are 52 million active social media users, 63% of the population, There are 44 million active mobile social media users who make up 53% of the population. 98% of adult humans in Turkey to use their mobile phone and 77% use smartphones. The proportion of those using a desktop computer or laptop is 48%, while the proportion of tablet users is 25%. [4]

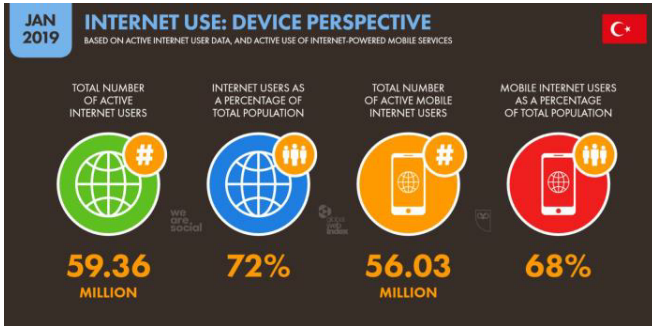


Figure 3: Current Status of Internet Use in Turkey [4]

People spend an average of 7 hours a day on the internet in Turkey. They spend an average of 2 hours and 46 minutes a day on social media. On average, they watch 3 hours and 9 minutes of TV per day and listen to music on average 1 hour and 15 minutes a day. There are a total of 52 million social media users in Turkey with 44 million mobile users and devices are connecting to social media. YouTube is the most active social media platforms in Turkey. Instagram and Facebook are used respectively. [4]

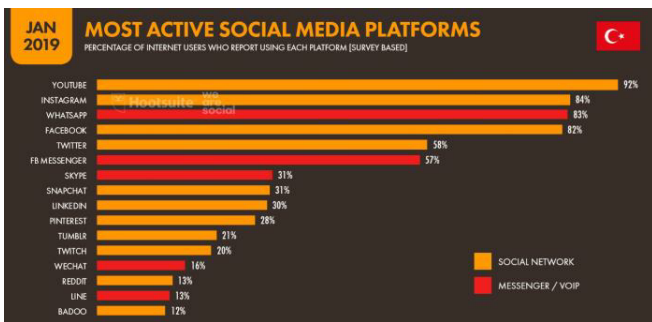


Figure 4: Using Social Media Platforms in Turkey [4]

There are 43 million Facebook users in Turkey is higher than the number of men and women. There Instagram'in 38 million users in Turkey and is a platform used by more men than women, such as Facebook. Twitter has 9 million users in Turkey and 8 of every 10 men from Twitter users. Snapchat, which is the platform that women use more, has 6.3 million users in Turkey. [4]

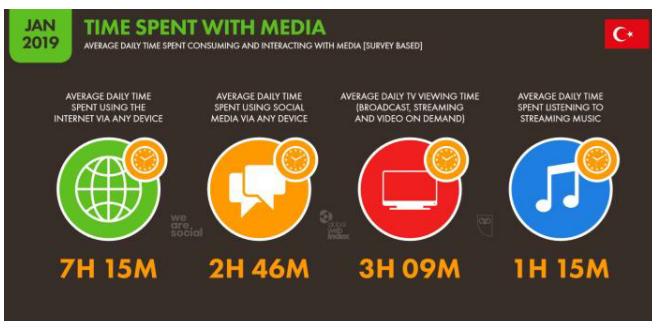


Figure 5: Time spent on social media in Turkey, in 2019[4]

Turkey social media men according to age and gender distribution of social media users use social media for each age

group are more likely than women. In general, 1/3 of the social media users are in the 25-34 age group. [4]



Figure 6: Using Social Media in the Turkey[4]

Information security management system in the manner specified in integrity and accessibility needs to be provided. Information access only The use by the persons given is his privacy. This access to written information or information systems information stored by authorized persons only access. Even if sensitive information is present The authorities should be informed. The integrity of the information is unmodified, partial or complete. Accessibility is the authority of stored information when necessary This is because it is accessible to people [5].

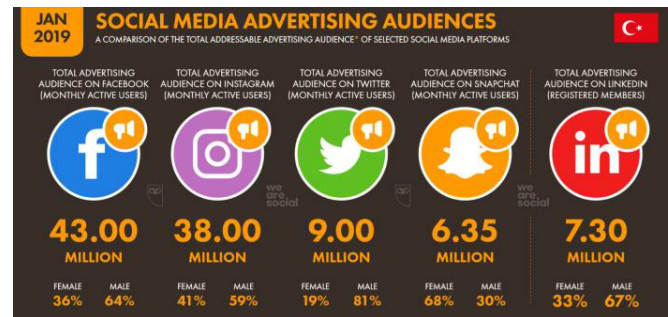


Figure 7: Social Media Platforms in Turkey [4]

Cyber threats now only to computer systems damage to systems (infiltration into systems, information from systems) and to provide false information to systems. One critical communication of the country systems, computer systems, energy and transportation networks, to the extent necessary to damage military command and control systems, It appears as an asymmetric war pattern. Cyberspace one of the major threats in the coming years the idea that; all started [6]

Earlier to the boom of individual computers, conventional computer examinations were of the burglary of computer components. In any case, computers presently play either a target (e.g., hacking), device (e.g., cyberbullying), or coincidental (e.g., capacity gadget for evidence) part in criminal examinations [7]. In expansion, conventional crimes committed through the utilize of computer innovation are presently alluded to as "old violations with unused tricks," such as cyberbullying, Web child explicit entertainment utilize, and cyberterrorism. laid out the distinctive ways in which social media may be

utilized by hoodlums: burglary, social designing, phishing, malware, character burglary, cyberbullying, cyberstalking, misuse, sexual attack, prostitution, organized wrongdoing, and cyberterrorism. The diverse ways in which social media may be utilized by offenders: burglary, social building, phishing, malware, personality robbery, cyberbullying, cyberstalking, abuse, sexual attack, prostitution, organized wrongdoing, and cyberterrorism. Not as it were are social media locales utilized to commit violations, hoodlums may take off a path of computerized prove, counting posts, brief message benefit (sms) communications, photographs, or geotaging, to title a few. [8-9]

For this reason, it is very important to construct more effective defense systems and to prepare emergency preparedness. Immediate detection of attacks, virtual or physical barrier building, development of regional and national cyber security policies it has become mandatory [10-11].

III. DIGITAL FORENSICS

For illustration, the kidnap, rape, and kill of Kimberly Proctor in March 2010 was solved much obliged to a digital path of evidence cleared out behind by the two young boys. [12], "police investigating this case ... accumulated the digital equivalent of 1.4 billion pages of paper evidence, counting Facebook and MSN messages, content messages and chat histories" (para. 7). In expansion, one of the young boys confessed to killing Kimberly Delegate whereas chatting with his online gamer girlfriend on World of Warcraft [13]. The murder of Kimberly Delegate was solved through the path of digital prove cleared out on social media (e.g., social networking destinations, virtual game worlds).

In expansion, there are a number of cases where people have confessed to their crimes on social media; "it shows up that the require for bravado is much more prominent than any concerns around getting caught" [14-15]. Hannah Sabata was captured after she posted a video on YouTube bragging that she had stolen a car and victimized a bank at gunpoint; she moreover flashed a huge wad of cash within the video [16-17]. In another example, a indicted savage criminal, and Midwest Lord pack part, nicknamed "P Smurf," posted a video of himself on Facebook shooting an AK-47. This video was forensically protected by neighborhood law authorization, and "P Smurf" was federally prosecuted in 2012 for weapon charges. [17]

Conventional computer forensics advanced into the field of digital forensics in response to the assortment of advanced gadgets, other than conventional computer hardware, that housed computerized evidence. Computer forensics was not a term that accurately spoken to the different shapes of digital evidence. [17]

Digital evidence is characterized as information that's either transferred or put away through a computer and digital evidence may be found on mobile gadgets, global situating frameworks (GPS) devices, gaming systems, and networks, to title some. In this way, digital forensics is an umbrella term that refers to the analysis of advanced evidence, which incorporates network forensics (Web activity), computer forensics, mobile-device

forensics (e.g., cell phone), and malware forensics (e.g., infections. [18-19]

IV. CRYPTOLOGY AND INFORMATION SECURITY

Cryptology, science concerned with data communication and storage in secure and usually secret form. It encompasses both cryptography and cryptanalysis. The term cryptology is derived from the Greek *kryptós* ("hidden") and *lógos* ("word"). Security obtains from legitimate users being able to transform information by virtue of a secret key or keys—i.e., information known only to them. The resulting cipher, although generally inscrutable and not forgeable without the secret key, can be decrypted by anyone knowing the key either to recover the hidden information or to authenticate the source. Secrecy, though still an important function in cryptology, is often no longer the main purpose of using a transformation, and the resulting transformation may be only loosely considered a cipher. Cryptology is the branch of mathematics that encompasses both password science (cryptography) and password analysis (cryptanalysis). The purpose of password science is to ensure message security. The purpose of the password analysis is to solve the existing passwords.

Encryption is the process of making the content of a message (plain text) unreadable without obtaining the appropriate information (key information). The purpose of the encryption is to prevent the message from being read by unwanted individuals. Decryption (decoding) is the process of converting encrypted text into plain text. Today's cryptography includes more than encryption and decryption. Authentication is as important as privacy. When we add a name to any message and send it over the network, we need electronic methods to prove our identity. The solution provided by cryptography is the digital signature. It can be secret keyed and open key cryptography. [20]

For information security, data is protected against unauthorized alteration, use, disclosure, review, recording, or damage. In network security, methods used to monitor and prevent unauthorized access to networked systems and resources (use of cryptographic algorithms in network protocols, etc.) are examined. To achieve Information Security goals: Confidentiality, Integrity, Availability, Identity Authentication, Non-repudiation and Accountability. Cryptographic algorithms are used to prevent malicious effects that communications may be exposed to as components of security protocols. Information Security Goals;

- Privacy: grant access to information
- Integrity: protection of information against modification or destruction (verifying the source of knowledge, etc.)
- Continuity: Use of information in a timely and reliable manner providing.
- Authentication: that users are the users they claim and check that the data accessing a system comes from a trusted source.
- Undeniable: sending data with the help of generated evidence or denying that the field is doing data transfer or reception to prevent.
- Traceability: to be responsible for a security breach ensuring the follow-up of user transactions to remove

and denial, discouragement it supports. [21]

V. CONCLUSION

In this study, a general evaluation has been made with the examples and statistics of digital forensics in social media and System, Informatics, Network and Cyber Security. According to this, Families should not allow their children to spend much time on social media. The relevant units of the state should provide training to parents on this issue. Users should be complained about Social Media Platforms that are engaged in terrorist activities, sharing information on illegal sites, sharing porn-related content, encouraging drugs and illicit substances, sharing illegal information, and disclosing state secrets. The penalties for those who share these shares in a virtual environment must be deterrent and the penalties identified should be announced on social media platforms.

REFERENCES

- [1] Kumar, V., Srivastava, J., ve Lazarevic, A. (2005). Managing cyber threats: issues, approaches, and challenges. Springer.
- [2] N. F. Doherty, L. Anastasakis, H. Fulford, "The Information Security Policy Unpacked: A Critical Study of the Content of University Policies", International Journal of Information Management, 29(6), 449–457, 2009.
- [3] F. Aslay "Cyber Attack Methods and Turkey's Cyber Security Situational Analysis", p. 5, 2017
- [4] <https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates>
- [5] F. Aslay "Cyber Attack Methods and Turkey's Cyber Security Situational Analysis", p. 5, 2017
- [6] A. Aytekin. "Evaluation of Turkey's Cyber Security Strategy and Action Plan", Unpublished Master Thesis, Information Systems Department, Gazi University, 2015.
- [7] Maras, M. (2012). Computer forensics: Cybercriminals, laws, and evidence. Sudbury, MA: Jones and Bartlett Learning.
- [8] Brunty, J., & Helenek, K. (2013). Social media investigation for law enforcement. Waltham, MA: Elsevier Inc..
- [9] Kathryn C. Seigfried-Spellar et al "The intersection between social media, crime, and digital forensics: #WhoDunIt?" , 2016
- [10] S. E. Goodman, "Critical Information Infrastructure Protection", Terrorism (Ed.), Responses to Cyber Terrorism NATO Science for Piece and Security, IOS Press (Volume 34), Ankara, 25, 2008.
- [11] F. Aslay "Cyber Attack Methods and Turkey's Cyber Security Situational Analysis", p. 5, 2017
- [12] Roberts, H. (2011). Teenage killer who tortured and suffocated classmate, 18, had left digital trail of sick plot and confessed on World of Warcraft. Retrieved from dailymail.co.uk.
- [13] Zetter, K. (2011). Teen Murderer Undone by World of Warcraft Confession and Trail of Digital Evidence. Retrieved from wired.com.
- [14] Gross, D. (2013). Why people share murder, rape on Facebook. CNN. Retrieved from www.cnn.com.
- [15] Kathryn C. Seigfried-Spellar et al "The intersection between social media, crime, and digital forensics: #WhoDunIt?" , 2016
- [16] Locker, M. (2012). Watch: Woman brags about bank robbery on YouTube, Gets Arrested. Time. Retrieved from www.newsfeed.time.com.
- [17] Kathryn C. Seigfried-Spellar et al "The intersection between social media, crime, and digital forensics: #WhoDunIt?" , 2016
- [18] Casey, E. (2011). Digital evidence and computer crime: Forensic science, computers, and the internet (3rd ed.). Waltham, MA: Academic Press.
- [19] Kathryn C. Seigfried-Spellar et al "The intersection between social media, crime, and digital forensics: #WhoDunIt?" , 2016
- [20] Encyclopedia of Britannica "Cryptology" <https://www.britannica.com/topic/cryptology/Cryptography> (2019)
- [21] Yildirim, H. "Information Security and Cryptology" , International Forensic Informatics Symposium, 2014

Gyroscope-Accelerometer Controlled Smart Disabled Wheelchair

T. ÖZSOY¹ and F. ATASOY¹

¹ Karabuk University, Karabuk/Turkey, tayfur.ozsoy@gmail.com

¹ Karabuk University, Karabuk/Turkey, ferhatatasoy@karabuk.edu.tr

Abstract – In the modern world, there is a great deal of effort to bring disabled people into social life. Disabled people want to be in all areas of the real life. However, this can be sometimes dangerous. Life and communication style have been changed and communication has moved in cyber world. People in the real world are not aware of each other. Therefore, we developed a system for disabled people in order to make their life easier. The developed system informs disabled person's relatives immediately when unexpected situations occurred. For this purpose, Arduino, SIM801 GSM module, NRF24L01 wireless transceiver, MPU6050 gyroscope-accelerometer and L298N motor drive are integrated into the disabled wheelchair. If the disabled wheelchair tilts over or drops, the system detects this situation and inform disabled person's relatives. Additionally, to facilitate control of the disabled wheelchair, gyroscope-accelerometer is used. Thus, if disabled person does not have arms and feet, he can control the wheelchair by gyroscope-accelerometer which can be mounted on glove, helmet or moving limb.

Keywords – Arduino, disabled wheelchair, gyroscope-accelerometer, global positioning system, smart disabled wheelchair.

I. INTRODUCTION

ACTUALLY when the first wheelchair was invented and used is unknown. In 1750s batch chair was introduced by English inventor James heath. In 19th century widely used by veterans in United States. After 1932, Hebert A. Everest and Harry C. Jennings introduced the standard design of cross-frame wheelchair. The first motorized wheelchair was appeared in the early 1900s. After World War II, demand for electric wheelchairs increased. In the 20th century wheelchairs were improved considerably [1].

User-friendly design and making control easy with new technology devices are grown up with improvements of the technology and manufacturing techniques [2, 3, 4, 5, 6]. Ergonomic designs are important, additionally software for control and user interface, devices and other environments are important, too. Especially using non-invasive and contactless devices are the most comfortable way for users. Another important thing is easy to control mechanism. In [3] electromyography (EMG) signals are used to control electric wheelchair. However, the developed system is not suitable in daily usage. There are many cables and disturbing sensors. The system can be produced in a compact structure, but it is a necessity to use contact sensors to detect EMG signals. As a

result, the developed system may cause skin sensitization in long-term usage. In [5] trackball is used to control electric wheelchair. Also, the trackball is not contactless, but it is more comfortable and better than attached sensors. In [2] and [4], they used contactless sensors to control electric wheelchair and hospital bed. The contactless sensors provide the most comfortable usage. In [6] face behaviors are used to control electric wheelchair, but in daily usage it may be hard to use. In outdoor usage, a person cannot move along with open mouth. In [4] the system is developed for indoor usage and eye-movements are used to change hospital bed position. As the authors mentioned, training time for users depends on their age. It means that people who are used to looking screens are more successful to use the system.

In this study, we developed a gyroscope-accelerometer controlled smart disabled wheelchair as a prototype. Disabled person can control the wheelchair with gyroscope-accelerometer that can be located on glove, helmet or moving limb. The system informs the disabled user's relatives when an undesirable situation occurred, such as a vehicle tilt over or crash, is encountered.

II. COMPONENTS AND INTEGRATION

In this study, user-friendly components were preferred. Since rapid prototype development can be suitable with commonly used components. After test are passed, the developed system can be optimized.

Arduino Pro mini is used as a microcontroller. Since Arduino platform provides open-hardware and software development environment. Additionally, many developers and companies work on the platform and introduce compatible components in to the market. Two Arduino are used in the project. One of them communicates with Global Service Mobile (GSM) module, global positioning system (GPS) module, gyroscope-accelerometer and other Arduino via wireless with nRF24L01 and controls electric motors by motor driver integrated circuit (IC) L298N. The other Arduino gets moving directions from gyroscope-accelerometer which is located on glove and sends via nRF24L01. Lithium ion batteries are used for power supply. In power stages voltage regulators and diodes are used to protect circuits.

MPU6050 is an integrated 6-axis motion tracking device. It uses I²C protocol and working voltage range is between 2.37V and 3.46V. It consists of 3-axis gyroscope and 3-axis

accelerometer [6]. According to information from it, Arduino which is located on glove sends moving directions to another Arduino. Other MPU6050 is located on disabled wheelchair to detect sudden acceleration. If sudden acceleration happens, that means wheelchair crashes or falls. Also, gyro position is considered, whether the wheelchair tilts over slowly.

nRF24L01 is a low power radio transceiver module. Its working frequency is 2.4 – 2.5 GHz. It is suitable for wireless data communication applications [7]. In this study two nRF24L01 are used. One of them is on glove and another is on disabled wheelchair. They provide wireless communication between other.

Additionally, two LM1117 adjustable voltage regulators are used. Even if battery voltage is high from 4.75V up to 10V, it guarantees to provide 3.3V [8]. Thus, supply voltage of Arduino Pro mini microcontrollers is provided by LM1117.

A. Structure of Glove

Connection diagram of Arduino on glove is given in Figure 1. A5 and A4 are used for communication with MPU6050. A0 and 10-13 pins are used for wireless communication with nRF24L01.

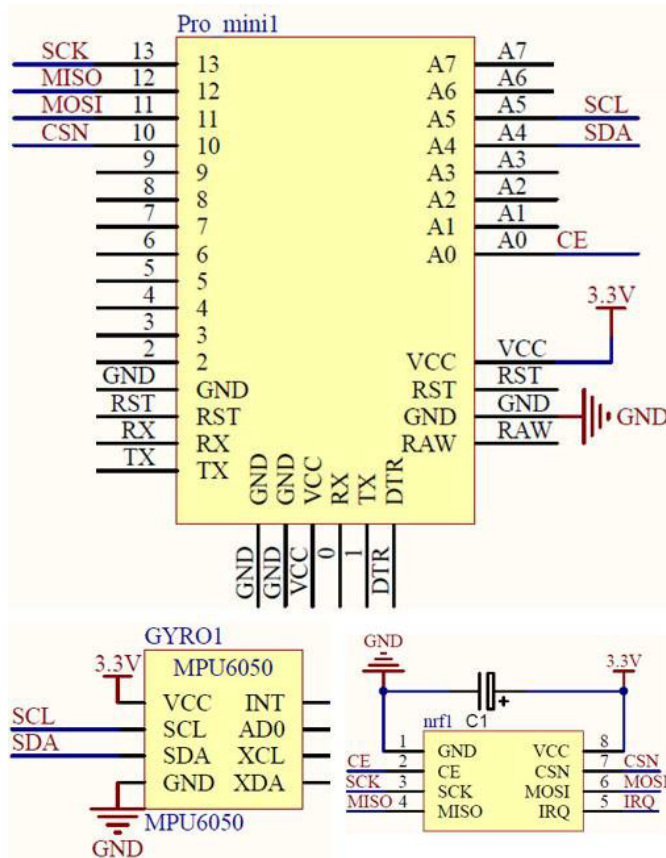


Figure 1: Connection diagram of Arduino on glove.

Power stage of glove is given in Figure 2. In power stage LL4148 is a diode and its breakdown voltage is 100V [9]. It is used to protect the circuit, if a battery is reverse connect. Remaining part of the power stage is connected according to datasheet of LM1117.

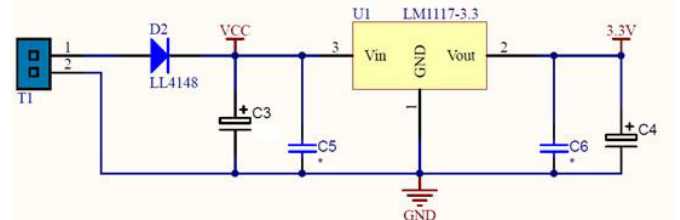


Figure 2: Power stage of glove.

B. Structure of Disabled Wheelchair

In addition to glove circuit, disabled wheelchair circuit consists of SIM800L GSM module, a new power stage for SIM800L, Neo-6m u blox GPS module, a relay control unit, a led as a warning light and L298n motor driver. Connection diagram of Arduino on disabled wheelchair is given in Figure 3.

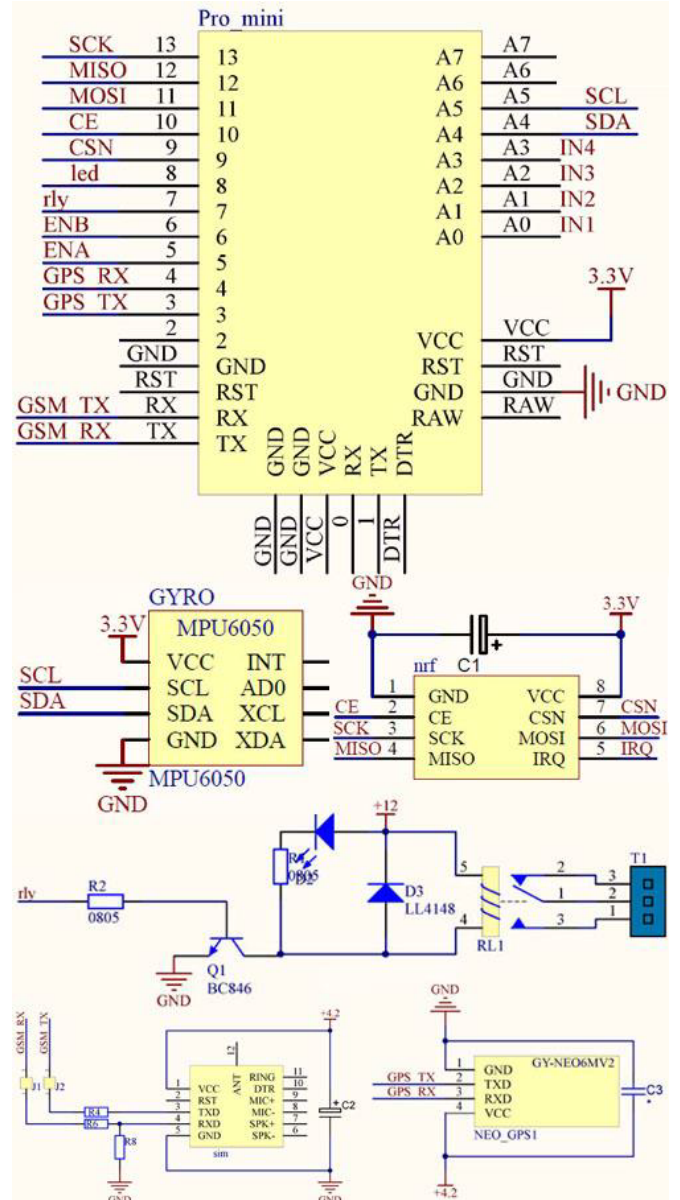


Figure 3: Connection diagram of Arduino on disabled wheelchair.

As seen in Figure 3, MPU6050 gyroscope-accelerometer and nRF24L01 connections are same as in glove circuit.

In Figure 3, the relay control unit is seen. BC846 is a npn transistor and used as switch. Its working region depends on 7th pin of Arduino. If 7th pin is High, the transistor is in saturation. That means the transistor and relay act close switches. Otherwise the transistor is in cut-off region and both the transistor and the relay act as open switches. High voltage loads can be driven by relay. In this study we used the relay unit to inform disabled user when the system receives short message via GSM module or unusual situation happens on the smart system.

SIM800L GSM module is a quad band GSM/GPRS (Global Packet Radio Service) module [10]. It is commonly preferred because of price and size of it. Although it supports from 3.4V to 4.4V, manufacturer recommends 4V. Thus, LM2596 adjustable step-down voltage regulator, which is recommended for SIM800L voltage value, is used powerful communication. Input voltage range of the voltage regulator is from 7V to 35V [12]. Output voltage can be adjusted from 1.25V to 30V via trimpot on the circuit board. The voltage regulator circuit is shown in Figure 4.

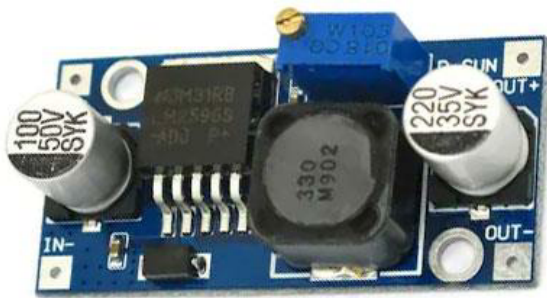


Figure 4: LM2596 adjustable step-down voltage regulator.

Neo-6m u blox GPS module is cost-effective, high performance GPS module [11]. It provides geo-location, when unusual situation happened.

L298N is used to drive motors. It is dual full-bridge motor driver. It is suitable for hobby electronics and robotic projects [14]. The motor driver circuit is given in Figure 5.

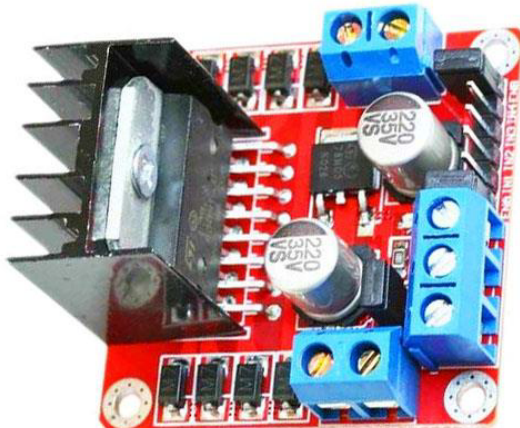


Figure 5: L298N motor driver circuit.

III. DEVELOPED PROTOTYPE

In Figure 6, developed prototype of gyroscope-accelerometer controlled smart disabled wheelchair is shown. In real applications, motors, motor driver should be changed. Besides, glove used as a remote controller unit and smart disabled wheelchair are in modular structure. Thus, another device can be used to control the smart disabled wheelchair.

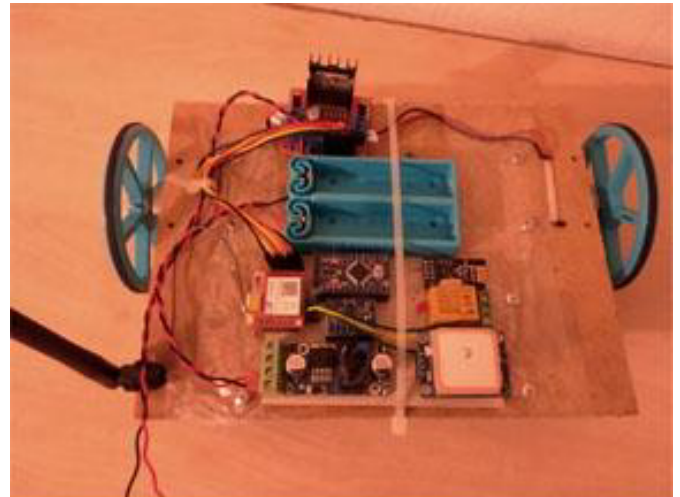


Figure 6: Developed prototype. (a) Internal structure of prototype. (b) Final prototype.

Gyroscope-accelerometer is located on glove. In developed prototype, circuit and battery are located on glove and each part of the system can be seen in Figure 7. Disabled person can control disabled wheelchair by using the glove. The circuit and the battery can be located on helmet or moving limb, whether disabled person is not able to use his hands and arms.



Figure 7: Remote controller of smart disabled wheelchair.

Adding and removing phone numbers to the system is done via SMS. Example SMS for adding relative is following: ">kyt1- 5XXXXXX". The fifth character of SMS describes the record number and 5XXXXXX part means the phone number. To delete registered phone number, an SMS must be sent as "kyt1- 0000000000". Similarly, the fourth character describes record number.

IV. CONCLUSION

In this study, a system was designed to make it safer for disabled people to travel outside alone. When unusual situation occurs, the disabled person's relatives are informed in this version. The system is not enough for commonly use in daily life. Tracking on web, number management interface and security sides should be improved. System can be improved and transformed into a structure that can be coordinated with municipalities or by state-owned disability assistance centers.

REFERENCES

- [1] N. Watson and B. Woods, "History of the wheelchair," 11 02 2015. [Online]. Available: <https://www.britannica.com/topic/history-of-the-wheelchair-1971423>.
- [2] A. Ghorbel, N. B. Amor and M. Jallouli, "An embedded real-time hands free control of an electrical wheelchair," in IEEE VCIP'14, Dec. 7 - Dec. 10, 2014, Valletta, Malta., 2014.
- [3] R. Hardiansyah and A. Ainurrohman, "The Electric Wheelchair Control Using Electromyography Sensor of Arm Muscle," IEEE, 2016.
- [4] N. A. Atasoy, A. Çavuşoğlu and F. Atasoy, "Real-Time motorized electrical hospital bed control with eye-gaze tracking," Turkish Journal of Electrical Engineering and Computer Sciences, vol. 24, pp. 5162-5172, 2016.
- [5] T. Sreejith, J. Vishnu and G. Vijayan, "Trackball Controlled Novel, Cost Effective Electric Wheelchair," in 2018 International Conference on Control, Power, Communication and Computing Technologies (ICCPCT), 2018.
- [6] N. B. Viet, N. T. Rai, N. V. Thuyen, Member and IEEE, "Hands-Free Control of an Electric Wheelchair Using Face Behaviors," in 2017 International Conference on System Science and Engineering (ICSSE), 2017.
- [7] "Sparkfun," 2013 08 19. [Online]. Available: <http://43zrtwysvxb2gf29r5o0athu.wpengine.netdna-cdn.com/wp-content/uploads/2015/02/MPU-6000-Datasheet1.pdf>. [Accessed 10 10 2018].
- [8] "Sparkfun," 03 2006. [Online]. Available: https://www.sparkfun.com/datasheets/Components/nRF24L01_prelim_p rod_spec_1_2.pdf. [Accessed 10 10 2018].
- [9] "Texas Instruments," 01 2016. [Online]. Available: <http://www.ti.com/lit/ds/symlink/lm1117.pdf>. [Accessed 10 10 2018].
- [10] "ON Semiconductor," 01 2016. [Online]. Available: <https://www.onsemi.com/pub/Collateral/LL4148-D.pdf>. [Accessed 10 10 2018].
- [11] "SIMCom," 20 08 2013. [Online]. Available: https://img.filipeflop.com/files/download/Datasheet_SIM800L.pdf. [Accessed 10 10 2018].
- [12] "Texas Instruments," 05 2016. [Online]. Available: <http://www.ti.com/lit/ds/symlink/lm2596.pdf>. [Accessed 10 10 2018].
- [13] "u-blox," 05 12 2011. [Online]. Available: https://www.u-blox.com/sites/default/files/products/documents/NEO-6_DataSheet_%28GPS.G6-HW-09005%29.pdf. [Accessed 10 10 2018].
- [14] "Sparkfun," 01 2000. [Online]. Available: https://www.sparkfun.com/datasheets/Robotics/L298_H_Bridge.pdf. [Accessed 10 10 2018].

The Optimization of The Process of Recognition Fingerprint Through The Minutiae Technique

E.GAZİOĞULLARI¹ and E.DUMAN²

¹ Firat University, Elazığ/Turkey, gaziogullariesra@gmail.com

² Firat University, Elazığ/Turkey, erkanduman@firat.edu.tr

Abstract - Biometric data science can be referred to identifying people uniquely and specifically, and in this field one of the most commonly used parameter is “fingerprint”. Fingerprints are peculiar to each person, and thus using fingerprints has become very important especially in identifying people. Identifying fingerprints specific to each person can be achieved by applying some algorithms in correct order.

Our main research subject in this study is to suggest an approach that will use this commonly used identifying system in the most efficient way and that will perform it much faster, too.

In this article the optimization of identifying fingerprint process has been examined by using minutiae technique, which is preferred mainly after the image processing algorithms, described as pre-process.

Keywords – Fingerprint, Minutiae, Image Processing, Crossing number, Biometric science.

I. INTRODUCTION

Using fingerprint as a biometric system is generally available in more specific areas compared to other systems such as had hand-held scanner, retinal scanning and voice recognition. In Forensic Area fingerprints have been used regularly and commonly for more than a hundred year and also the automatic fingerprint defining systems were founded approximately 50 years ago [1]. The fingerprint, which is used to ensure security in military and forensic areas for long years, is also important and used for datas security in private sectors in the last years [2].

There are many different techniques to attain fingerprints as it is commonly used in various areas.

There are some difficulties to define fingerprints and one of the most difficult one is related to the quality of pictures regarding with fingerprints [3]. In the defining systems the processes, which are defined as pre-processing, are performed to make the image enhancement. In this process, the aim is to get better and clear results from data.

The researches among the various fingerprint matching algorithms in the literature are mostly based on Minutiae Matching. Minutiae techniques is the most preferred one along with the hybrid feature matching and local methods [3].

In the process of defining fingerprints, it is possible to perform the process in two ways: filtering and minutiae.

In the filtering technique, there are some filters to make the images seem better, some of these filters are Contrast Enhancement, Mean Value Method, Median Filtering, Laplacian Filter.

The noises on the images can be removed by using some kinds of filters on the images, which is possible by applying the filters successively. Due to the fact that this process requires being serially applied, much more complex algorithms are generally preferred [4].

In the filtering technique, it is benefited from the datas like fingerprint direction and frequency of the improved and cleaned images [2]. The properties like the quality, magnitude and clearness of the images gained from the fingerprint device have a very important place in the filtering technique.

Moreover, if there is not a clear image, it becomes more difficult to achieve true a results, in other techniques, but the minutiae technique is the least affected technique from these negative factors.

In the Minutiae technique, the points which differ from the others on the fingerprint image are determined, and then they are named according to their shapes suitable to their positions. The main line and opposite lines determine the minutiae classifications, which are ridge ending, bifurcation, dot, island, lake, hook, bridge, double bifurcation, trifurcation, opposed bifurcation, ridge crossing, opposed bifurcation/ridge ending [5]. With the determination of these specific points, it is aimed to match local points and it requires matching the identified minutiae points instead of matching the fingerprints as image-based. The storage of these points as data provides much efficiency and less cost compared to the image based studies.

Through the determination of the minutiae points, it is possible to identify the patterns of a person's fingerprint as a string, in which the standard properties of formerly determined specific points are found.

II. METHOD

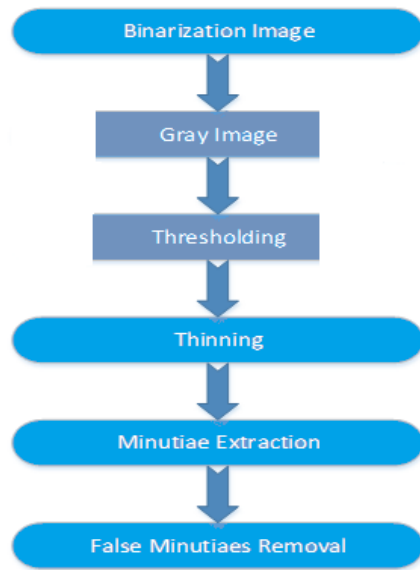


Figure 1: Minutiae Extraction Steps [1].

As in the figure 1, the first step of the Minutiae extraction process is to transform the available image into the grayscale image. Following this step, the process of threshold value is performed, which is used to attain the black-white images.

2.1 The Step of Converting an Image into Grayscale Image

Grayscale image is the image which has only gray tones colors. In fact, a gray color is the color in which red, green and blue components all have equal density in RGB space. Therefore, it requires stating only a density value for the three density values of each pixel [6,7].

The usage of the grayscale images is very common as most of our present time's screen and image catching hardware can support only 8 bit images. Moreover, most of the grayscale images are sufficient in many areas in terms of usage, and thus there is no need to use much more complex and difficult colorful images [7].

In general, gray toning density is stored as an 8 bits integer that gives possible 256 different gray tones from black to white. The "0" value represents the black color while the "255" value represents the White color [7].

In fact, the converting process of a colorful image to a gray toned color through the grayscale method is attaining gray toned images that have their equivalents in each color in the RGB model [6,7].

The gray toned image can be gained by computing the arithmetic mean that corresponds to the colors in RGB model. However, this computation doesn't precisely reflect the sensitivity of detecting each color differently. For this reason, to reach the most realistic and concrete result, the formula as shown in (1) that is performed by scaling must be used [7,8].

$$\text{Gray-Scale} = 0,299 * R + 0,587 * G + 0,114 * B \quad (1)$$

2.2 The Step of Finding Threshold Value

During the binarization process, a thresholding process must be performed after getting gray color.

This process is fundamental to determine whether black or white color must be used for each pixel value of an image. According to the determined thresholding value, one can decide to use which color, black or white [9].

In the global thresholding, a random value is used for the threshold value. However, in this process, it may not be possible to get the same influence and result for the images that have different color densities. Therefore, determining a suitable dynamic threshold value peculiar to the image you have and using that value is much more accurate [9,10].

There is another technique to determine threshold value: Otsu threshold. This technique provides the most optimum threshold value by converting the image on gray level to dual level. First of all, color histogram is computed as it is important to look at the numbers of each color density on images and the process goes on according to this color histogram [10].

The hypothesis that the image consists of two color class as background image and foreground of the image is created and then the intra-class variance value of this two colored class is calculated for the all threshold values. The threshold value that makes this value the lowest is called as the optimum threshold value [10,11].

After applying these steps, A grayscale image is turned into binarization image. A binarization images has only two different values, the "0" value for black color and the "1" value for white one [6,7].

2.3. The Step of Thinning

Thinning can be used for various applications, but it is more beneficial to use it especially for Extracting Skeletons and Medial Axis Transform [12]. In this mod, all the lines are turned into only one pixel and it is necessary to have binary image first, which we have attained in the first step of our study.

In our study, The Zhang-Suen Thinning algorithm which is based on the logic of extracting skeletons has been used [12].

Zhang-Suen Thinning Algorithm:

The application of this algorithm is not complex and slow, also the iterations in this algorithm consists of two subordinate parts. In the first one if the four conditions stated below are carried out, one pixel $I(i,j)$ is deleted [13]. These conditions are:

- 1.) The connection number has the value of "1".
- 2.) There can be at least two black neighbors and it is also not more than 6.
- 3.) At least one of the $I(i,j+1)$, $I(i-1,j)$ and $I(i,j-1)$ is white.
- 4.) At least one of the $I(i-1,j)$, $I(i+1,j)$ and $I(i,j-1)$ is white.

In the second part of the iteration, the third and forth conditions, belonging the first part, are changed as started below:

- 1.) The connection number has the value of "1".
- 2.) There can be at least two black neighbour and it is not more than six.

- 3.) At least one of the $I(i-1,j), I(i,j+1)$ and $I(i+1,j)$ is white.
- 4.) At least one of the $I(i-1,j), I(i,j+1)$ and $I(i,j-1)$ is white.

While applying the algorithm, if the repetitions used are completed, the pixels carrying the necessary conditions are deleted. If there are not any pixels to be deleted, then the algorithm is ended [14].

2.4. Minutiae Extraction

In a fingerprint defining system, it is necessary to match the each user in database with the stored fingerprint template. This process includes many computations and searching, which results in additional load. Therefore, we need to a fingerprint classification system that helps us limit the dimension of database template dramatically. In order to achieve this goal, minutiae properties are extracted and then they are matched with the fingerprints. The dimension of template that bases on the minutiae technique is small and most of the fingerprint defining systems depend on this minutiae technique, too [15].

Minutiae points are the specific points that belong to a fingerprint image and are used to make the image unique. Minutiae can be defined as the points on which ridge lines end or bifurcate and there are many minutiae types [15].

Minutiae types:

- If the ridge ends suddenly, that points is called as 'ridge ending'.
- If one ridge branches out into one or more ridges, then it is 'ridge bifurcation'.
- Too small ridges are called as 'ridge dots'.
- If the ridges are a bit longer than the dots and are found between two diverging ridges, they are called as 'ridge islands'.
- If there is nothing between two diverging ridges then they are lakes.
- If small ridges participate in two longer ridges, then they are 'bridges'.
- If two ridges cross one another, then they are called as 'crossovers' [15].

Among these minutiae types, ridge ending and ridge bifurcation are the most common minutiae types.

In the minutiae technique, after our first step pre-processing are completed, then it is time to extract minutiae. In this process, the most widespread minutiae extraction technique is the "crossing number", in which minutiae points are identified. This technique is generally preferred to extract minutiae points due to its pureness, simplicity and performance efficiency.

Cross Number Technique:

This is the most common minutiae extraction method and it is shown as CN referring to crossing number.

P4	P3	P2
P5	P	P1
P6	P7	P8

Figure2: 3x3 window for searching minutiae [16,17].

In this method a 3x3 window is used and the minutiae points are extracted by searching and scanning the neighbourhood of very ridge pixel in an image.

After that, the CN is calculated and it refers to the sum of differences between adjoined pixels. For the P point that belongs to 3x3 window, the process is performed counter clockwise for the eight neighboring pixels [16,17].

The CN value for a ridge pixel is calculated as shown below in (2) [18].

$$CN(x, y) = \frac{1}{2} \sum_{n=1}^8 |P_i - P_{i+1}|, P_1 = P_9 \quad (2)$$

The types of points that refer to calculated CN values are as in the below [19].

CN	Property
0	Isolated Point
1	Ridge ending point
2	Continuing ridge point
3	Bifurcation point
4	Crossing point

Figure 3: Crossing number types [19].

The ending point that has '1' value for the CN number and the bifurcation that has "3" value for CN number are as in the Figure 4 stated below [20].

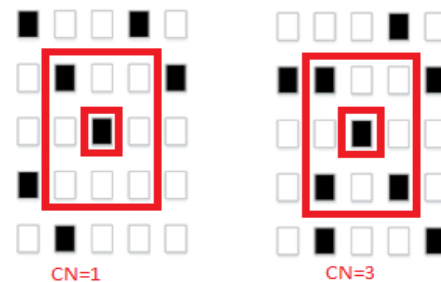


Figure 4: Ending Point and Bifurcation [20].

The identifying process is fulfilled according to the specific properties of the minutiae points. These properties can be stated as the x and y coordinate points, the calculated angle value of the minutiae and also the type of the minutiae. After these processes, it is possible to identify a person through these flocks of minutiae.

2.5. Post-Processing Technique:

This process is the last process and it is performed after the extraction of minutiae points.

After identifying the minutiae points, the wrong points that are not clear and explicit are extracted and then these points are deleted [18, 19]. Using more accurate and clear minutiae points increases the accuracy of the identifying processes.

III. EXPERIMENTAL RESULTS

In this study, the dataset of fingerprint images from Cornell University was used which consists of 6.000 fingerprint images of 600 persons [21]. The dimension of the each fingerprint image is 96x103 size and bmp.

In the proposed method, the fingerprint image is first converted to a black-and-white image. This binary image is then applied to the Thinning process and the Crossing number technique. Finally, minutiae points are deduced according to CN value. After these stages, the value of the minutiae points belonging to the fingerprint picture is recorded in the file in text format. These four fields are the coordinate information (x and y values), angle values and the class name of the minutiae type. The process is completed by keeping the file format as string. The outputs of the steps applied to the original fingerprint image are shown in Figure 5, Figure 6, Figure 7, Figure 8 below.



Figure 5: Fingerprint Image.



Figure 6: Binary Image.



Figure 7: The Image following the thinning process.

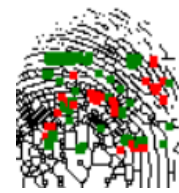


Figure 8: Minutiae Points.

IV. CONCLUSION

In this article, it is aimed to inform about the fingerprint identifying system, which is a very popular technique among the biometric applications in technology. Moreover, the identifying process has been applied to ensure high quality of fingerprint images through the minutiae technique and in the end, an output in the text format that includes specific points to identify a person has been attained.

When compared to the comparisons of images, much faster results have been attained by means of text matching technique.

The necessary steps for the process of extracting minutiae have been defined through the optimum parameters and in each step the optimum techniques have been applied. Especially the design of each step's working in parallel in itself is a very important acquisition as it means the more parallel the processes are, the faster results you gain. In our next studies, it is especially aimed to search the parallel implementation of the process of extracting minutiae by depending on this framework.

REFERENCES

- [1] BANSAL R., SEHGAL P., BEDİ P., (September, 2011), "Minutiae Extraction from Fingerprint Images-a Review", *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 5, No 3, ISSN [Online]: 1694-0814 September 2011.
- [2] Akın N., Takmaz B., Güvenoğlu E., "Image Encryption Using Fingerprint", *XV. Academic Informatics Conference Presentations*, Antalya, January 2013.
- [3] N. U. Ain, F. Shaukat, A.S. Nagra and G. Raja, "AN EFFICIENT ALGORITHM FOR FINGERPRINT RECOGNITION USING MINUTIAE EXTRACTION", *Pakistan Journal of Scienc*, Vol 70, June 2018.
- [4] VARLIK A., ÇORUMLUOĞLU Ö., "Photogrammetric Approach in Person Identification", *TUFUAB IV. Technical Symposium*, June 2007.
- [5] ÖZKAYA N., SAĞIROĞLU Ş., BEŞDOK E., "General Purpose Automatic Fingerprint Recognition System Design and Implementation", *Journal of Polytechnic*, Vol: 8 No: 3 pp. 239-247, 2005.
- [6] C. Saravanan, "Color Image to Grayscale Image Conversion", *2010 Second International Conference on Computer Engineering and Applications*, Vol 2, April 2010.
- [7] Fisher R., Perkins S., Walker A., Wolfart E., "Hypermedia Image Processing Reference", 1996.

- [8] Isnanto R. R., Riyadi M. A., Awaj M. F. , “Herb Leaves Recognition using Gay Level Co-occurrence Matrix and Five Distance-based Similarity Measures”, *International Journal of Electrical and Computer Engineering(IJ ECE)*, Vol.8,pp. 1920-1932, June 2018.
- [9] Bangare S. L.,Dubal A.,Bangare P. S.,Patil S.T.,”Reviewing Otsu’s Method For Image Thresholding”, *International Journal of Applied Engineering Research* ,Vol. 10, pp. 21777-21783,2015.
- [10] Otsu N., “A Threshold Selection Method from Gray-Level Histograms”, Vol. 3, pp. 62-66, January 1979.
- [11]Atasoy H., ”Otsu Threshold Determination Method”, Available:<http://huseyinatasoy.com/Otsu-Esik-Belirleme-Metodu>.,February 2011. [Online](02.03.2019)
- [12] Morphology Thinning, Available: <http://www.cce.hw.ac.uk/hipr/html/thin.html>[Online].
- [13] Image Processing Techniques For Machine Vision, Available: http://www.eng.fiu.edu/me/robotics/elib/am_st_fiu_ppr_2000.pdf [Online].
- [14]The Zyang- Suen Thinning Algorithm: Introduction and applications, Available: <http://jaredgerschler.blog/2018/02/25/the-zhang-suen-thinning-algorithm-introduction-and-applications/> [Online].
- [15] Thakkar D.,”Minutiae Based Extraction in Fingerprint Recognition” ,Available:<https://www.bayometric.com/minutiae-based-extraction-fingerprint-recognition/> [Online].
- [16] Feng Zhao and Xiaou Tang, “Preprocessing and postprocessing for skeleton-based fingerprint minutiae extraction”, *Pattern Recognition Society*,pp. 1270-1281,April 2007.
- [17] Ishpreet Singh Virk and Raman Maini, “Fingerprint Image Enhancement and Minutiae Matching in Fingerprint Verification”, *Journal of Computing Technologies*, vol. 1, June 2012.
- [18] Sudiro S. A. and Yuwono R. T. , “ADAPTABLE FINGERPRINT MINUTIAE EXTRACTION ALGORITHM BASED-ON CROSSING NUMBER METHOD FOR HARDWARE IMPLEMENTATION USING FPGA DEVICE”, *International Journal of Computer Science, Engineering and Information Technology (IJCEIT)*, Vol.2, No.3, June 2012.
- [19] Babatunde I. G., Akinyokun, Charles O., Kayode A. B., “A Modified Approach to Crossing Number and Post-processing Algorithms for Fingerprint Minutiae Extraction and Validation”, *IMS Manthan - Volume VI*, No. 1, June 2011.
- [20] RAVI. J, K. B. RAJA and VENUGOPAL. K. R, “FINGERPRINT RECOGNITION USING MINUTIA SCORE MATCHING”, *International Journal of Engineering Science and Technology* Vol.1(2), pp. 35-42, 2009.
- [21] Shehu Y.I., Ruiz-Garcia A., Palade V., James e. A. ,” Sokoto Coventry Fingerprint Dataset”, *CoRR journal*, vol. abs/1807.10609, 2018.

An Automated GIS Tool For Property Valuation

S.K.M. ABUJAYYAB¹, I.R. KARAS¹, C. AKICI¹, G. OZKAHRAMAN¹

¹ Karabuk University, Karabuk/Turkey, s.jayyab@hotmail.com

¹ Karabuk University, Karabuk/Turkey, irkaras@gmail.com

¹ Karabuk University, Karabuk/Turkey, cerenakicii@gmail.com

¹ Karabuk University, Karabuk/Turkey, gozkahraman@hotmail.com

Abstract - Property value is a reflection of locational, physical, legal and economic factors. Spatial factors are the most important factors among evaluation criteria. Geographic Information System (GIS) provide capable tools that can be used to record spatial information about value properties. The purpose of this study is to developing a property valuation GIS tool, which capable to estimate residential properties values. To achieve this objective, tabular data was developed that geographically represent of property information factors. Then, multi criteria decision analysis MCDA used to evaluate the property value. The tool capable to generate property values as percentage in the tabular data. In this study, Safranbolu-Turkey region has been studied. The property value influence factors are distance to main roads, distance to markets, distance to child parks, distance to schools, age of building, floor of building and distance to city center. The tool capable to help Safranbolu municipality to generate property evaluation for priced fair pricing, renting, buying or taxation and based on the data updating.

Keywords – Property valuation, Automated GIS Tool, MCDA.

1. INTRODUCTION

Property valuation is a serious need for institutions and individuals worldwide for public institutions, individuals, and developing cities, which have developing and renewed properties [1]. Property valuation become a primary task for several applications such as rental, trading and tax payments [2]. Several systems were developed for property valuation based on spatial data. National land information system NLIS [3] is a great example of property valuation systems used in the UK. In NLIS system, 24 million search perform yearly regarding property values [3].

Property value is a reflection of locational, physical, legal and economic factors [4]. Spatial factors are the most important factors among evaluation criteria [5]. Property valuation parameters can be determining by using local experts, investors and institutions opinions [6]. Parameters must be relevant to the local area characteristics and frequently questioning [7].

In order to perform property valuation, spatial data of factors required to be collect for the valuation [8]. ArcGIS software contains both database manager and set of spatial functionalities. ArcGIS allow to end-users to store and analysis spatial data of valuation factors across the desired geographic landscape [9]. GIS can potentially play an important role in real estate research [10].

In order to perform property valuation for new geographic area, several GIS tools are essential. The available tools in ArcGIS are insufficient to perform property valuation. For

this reason, this paper aimed to develop new tool for property valuation in Safranbolu-Turkey using ArcGIS, Python and MCDA.

2. METHODOLOGY

This study intended to develop the tool in Esri ArcGIS application using Python programming language. ArcGIS is the most popular GIS software worldwide among specialists. The final outcome is ArcToolbox. The property evaluation process were divided into four steps, and the necessary scripts are written for each step. Each tool displayed as figure. These scripts are written using the python programming language. The necessary steps for the tool have been completed with four associated scripts. The general steps of the tool displayed in Figure 1.

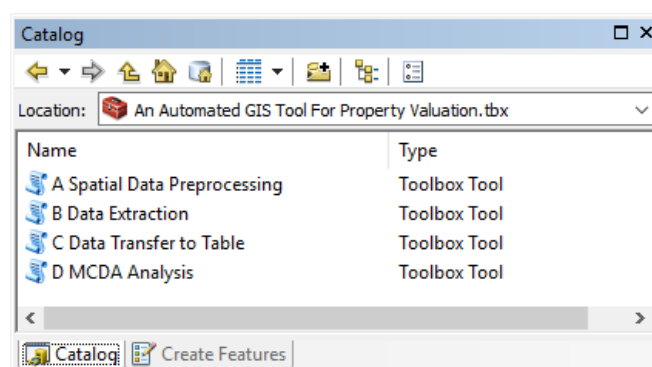


Figure 1 An Automated GIS Tool For Property Valuation

2.1. Spatial Data Preprocessing

In real world, the real value of property is difficult to define. It may vary depending on the prospects of individuals or organizations. At this instant, the furthestmost mutual factors are considered. As stated in [11], 407 parameters can be subtracted according to tastes and preferences for property valuation. Based on the consultation of local experts, seven primary factors were adopted. These factors are based on the distance from property that has been evaluated to schools, streets, parks, markets and city center. In addition, the dataset is taken from the attribute table as the building age, the floor, the area information is taken as a parameter and the result is intended to affect.

In this step, spatial data layers of factors were utilized. From Vector data of factors, several parameters converted to raster data for distance calculations. The objective here to prepare the

data for next stage to obtain and automatic correct distance information from factors objects.

In addition, by converting the property data from polygon into point, next stage will be able to extract the spatial informatics from the raster data based on property points. It is aimed to get results effectively and accurately and away from technical problems. The Spatial Data Preprocessing stage showed in Figure 2. The age, floor and area factors require to be manually inter to the dataset as displayed in Figure 3.

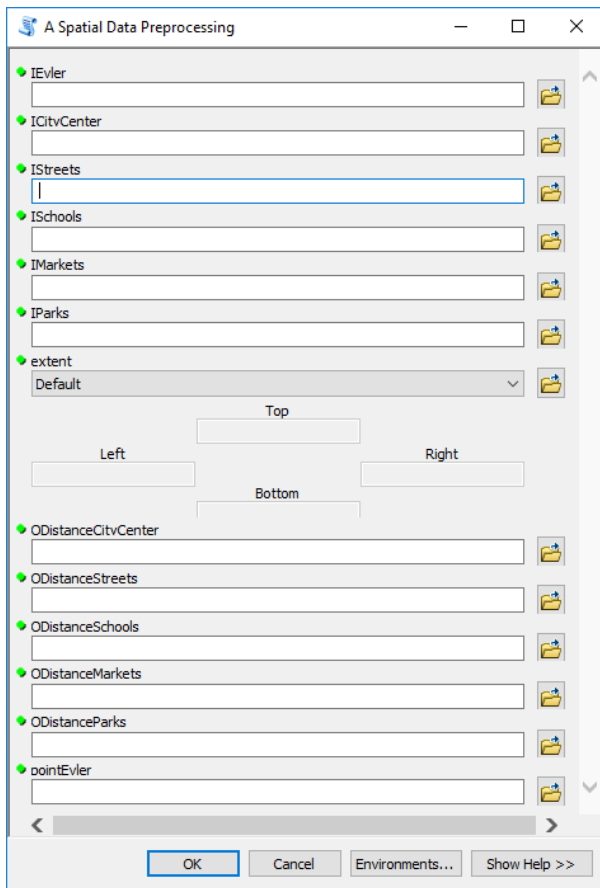


Figure 2 Spatial Data Preprocessing stage

FID	Shape *	Id	area	floor	rooms	age
0	Polygon	0	895,767	10	2	8
1	Polygon	0	803,067	2	1	5
2	Polygon	0	468,972	30	1	2
3	Polygon	0	363,253	4	2	8
4	Polygon	0	239,305	1	1	6
5	Polygon	0	221,078	3	1	15
6	Polygon	0	322,892	8	2	5
7	Polygon	0	339,295	1	1	4
8	Polygon	0	348,151	1	3	4
9	Polygon	0	480,433	2	1	4
10	Polygon	0	372,891	2	4	7
11	Polygon	0	433,692	2	1	5
12	Polygon	0	449,446	3	4	3
13	Polygon	0	342,338	3	3	2
14	Polygon	0	286,089	4	2	1
15	Polygon	0	304,375	4	4	9

Figure 3 property dataset and manual interning factors

2.2. Data Extraction

This step is related to the previous step. The output raster data obtained in the previous step is used as input. In this step, processes were continued by creating the dataset and the attribute table for the property points automatically. Information of the distance to the city center, distance to the street, distance to the school and the distance to the park from the previous step are used as input. Factors placed in the dataset beside the age, floor and area factors for property evaluation. Thus, the dataset extraction performs. The stage tool presented in Figure 4. an example of final dataset were presented in Figure 5.

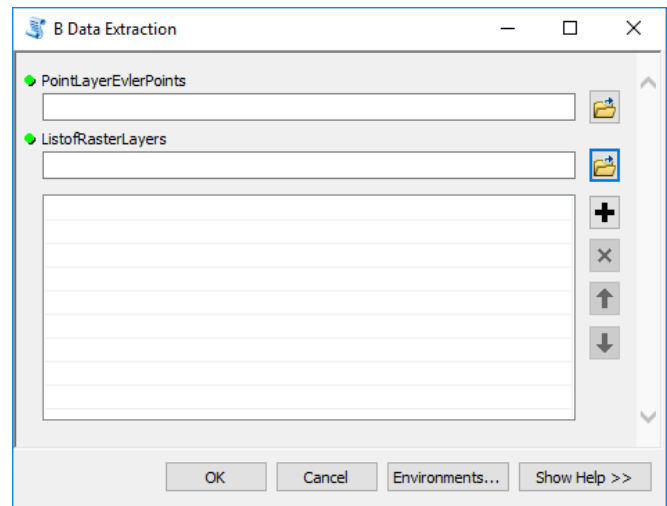


Figure 4 Data Extraction

FID	Shape	Id	area	floor	rooms	age	ORIG_FID	dstnceastre	dstnceascho	dstnceast	dstncepark	dstncecmrk
0	Point	0	895,767	10	2	8	0	30	120	30	80,7775	279,016
1	Point	0	803,067	2	1	5	1	30	171,026	30	135,831	261,964
2	Point	0	468,972	30	1	2	2	15	108,187	15	47,4342	345,326
3	Point	0	363,253	4	2	8	3	30	167,705	30	109,202	328,976
4	Point	0	239,305	1	1	6	4	45	175,57	45	120,834	209,233
5	Point	0	221,078	3	1	15	5	30	208,826	30	152,971	313,209
6	Point	0	339,295	1	1	4	7	21,2132	80,7775	21,2132	159,452	190,919
7	Point	0	348,151	1	3	4	8	61,8466	123,883	61,8466	192,094	162,25
8	Point	0	480,433	2	1	4	9	75	152,971	75	216,333	147,733
9	Point	0	372,891	2	4	7	10	61,8466	135	61,8466	233,345	117,154
10	Point	0	433,692	2	1	5	11	30	33,541	30	120,934	237,171
11	Point	0	449,446	3	4	3	12	15	61,8466	15	96,9469	255,441
12	Point	0	342,338	3	3	2	13	33,541	161,555	33,541	106,066	437,321
13	Point	0	286,089	4	2	1	14	45	192,094	45	129,035	409,695
14	Point	0	304,375	4	4	9	15	45	192,094	45	129,035	409,695
15	Point	0	322,892	8	2	5	6	21,2132	256,32	21,2132	200,125	318,198

Figure 5 an example of final property valuation dataset that extracted from raster data

2.3. Data Transfer to Table

After the calculations of the property were done via point vector data, data were continued to be processed in vector polygon layer in order not to get away from real presentation. With this stage, a new layer creates and switch from point vector data to polygon vector data. The two layers are spatially matched and the tabular data joined. While joining, all raster and static table data of property are move from points layer to original polygon layer. The developed tool were illustrated in Figure 6.

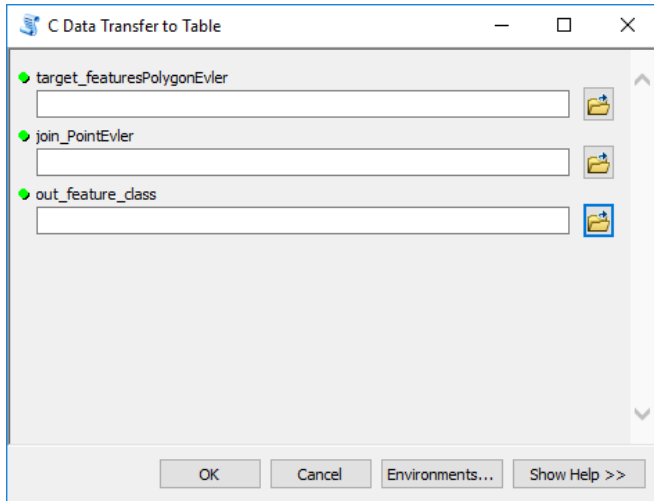


Figure 6 The third stage of the model. Joining the dataset from points layer to polygons layer.

2.4. MCDA Analysis

In this step, the vector data has been moved to the layer, and the data is presented to the user's choices. In addition, numerical result will be obtaining. As a result of this step, people or end-users can make a real property valuation.

By applying MCDA for the input factors using property valuation will be performing. The tool access to the created spatial dataset and migrate the data to a calculator. The calculator can be used to perform MCDA and produce final scores for property. The results are extracted as a percentage. The tool exhibited in Figure 7, and an example of output table demonstrated in Figure 8. Then, the user need to sort the output values and take the highest percentage.

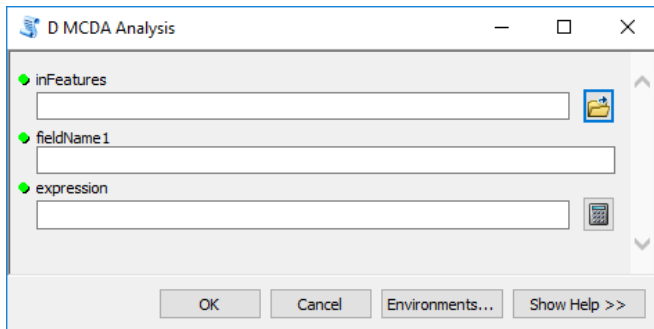


Figure 7 MCDA Analysis stage

[illegible]

Figure 8 an example of output table based on MCDA

3. CASE STUDY

This study was carried out gradually and the result of each stage was used as the input of the next stage. Parameters and specified input values are left to the user. The study was conducted on Safranbolu-Karabuk as illustrated in Figure 9. However, this tool was provided to be use on the desired region by leaving the inputs to the user.

In the study, location factors are used, and access to accurate distance information is targeted as illustrated in Figure 10. The distance between the determined property and selected parameters (city center, school, park, street, market) is calculated and converted into raster data. Then the results are transferred to the database to obtain a result value. Attention was paid to produce final value to be able to evaluate the result value based on MCDA.

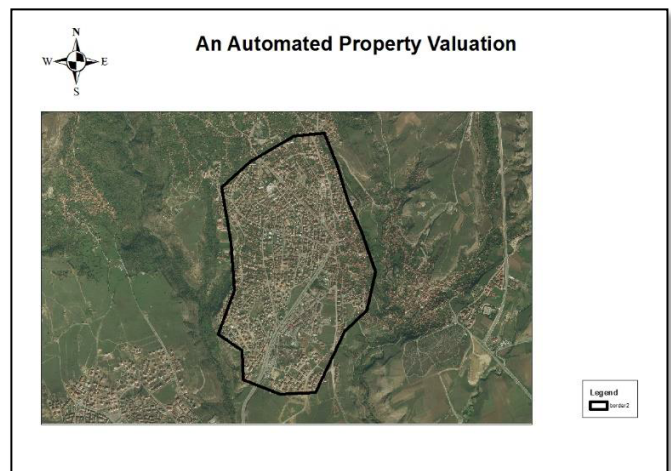


Figure 9 Study Area, Safranbolu

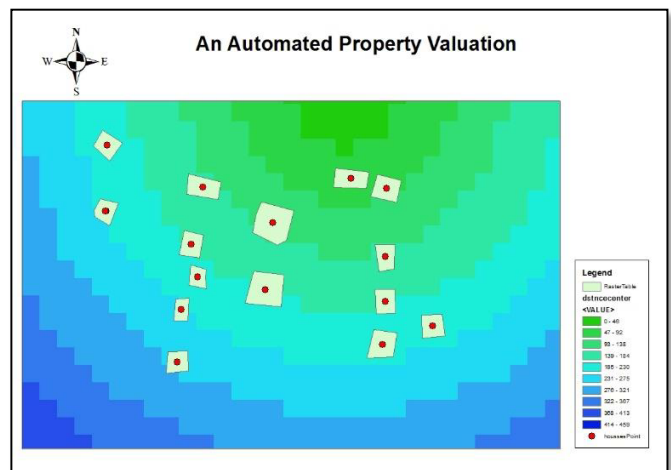


Figure 10 Property Positioning on Raster Model

4. DISCUSSION AND CONCLUSION

Location is an important factor when conducting real estate analysis. It has been determined as the public needs to obtain accurate results and improve our decision in the life such as rental, trading and tax payments. For this reason, accurate results were obtained with the same conditions in Safranbolu regions. Results were obtained with GIS technology and spatial database.

In the study, a general evaluation was made for a real estate. At the time of the study, the distinction between the distance to each point of the city center of Safranbolu Atamerkez was determined as the most important factor. Atamerkez city center was observed during the day, the distance of the restaurant, market, shopping center and cinema locations were observed at this point. In addition to this, it is reinforced that the main parameter for the people who want to live in this environment is the city center.

The Scripts were written and tool developed because the ArcGIS toolbox were inadequate at this point. Thus, the problem of spatial decision is solved. In order to help the decision making, a numerical result was produced and the decision mechanism was left to the user by knowing that the expectations of the individuals would be different.

Consequently, it has been proving that the tool was able to generate property valuation for priced fair pricing, renting, buying or taxation.

ACKNOWLEDGMENT

This study has been supported by 2221 – Fellowship Program of TUBITAK (The Scientific and Technological Research Council of Turkey). We are indebted for their supports.

REFERENCES

- [1] S.W.Gatheru, D.Nyika, Application Of Geographic Information System In Property Valuation International Journal of Scientific & Technology Research Volume 4 Issue 08, August 2015
- [2] M.Lemmens, J.Kurm, Integrating Gis with a Land Valuation Information System: Some non-Technical Consideration for the Estonian Case, Cadastre and Land Administration II, 2000
- [3] <https://www.nlis.org.uk/index.aspx>
- [4] M.Rodriguez, C. F. Sirmans, Allen P. Marks, Using Geographic Information Systems to Improve Real Estate Analysis, American Real Estate Society, 1994
- [5] S.Metzner, A.Kindt, Determination of the parameters of automated valuation models for the hedonic property valuation of residential properties: A literature-based approach, Emerald Publishing Limited, 2017
- [6] A.C. Aydinoglu, R. Bovkir, An Approach for Calculating Land Valuation by Using Inspire Data Models, a Dept. of Geomatics Engineering, Gebze Technical University, 41400, Gebze, Kocaeli, Turkey, 2017
- [7] P.J. WYATT, The development of a GIS-based property information system for real estate valuation, International Journal of Geographical Information Science, 1997
- [8] F.Marzuki, Residential Property Valuation Using Gis, Conference: Signal Processing & Its Applications (CSPA), 2015 IEEE 11th International Colloquium, 2015
- [9] S. W. Gatheru, D. Nyika, Application Of Geographic Information System In Property Valuation, International Journal of Scientific & Technology Research Volume 4, Issue 08, August 2015
- [10] A.C. Aydinoglu, R. Bovkir, An Approach for Calculating Land Valuation by Using Inspire Data Models, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,

Volume XLII-4/W6, 2017 4th International GeoAdvances Workshop, 14–15 October 2017, Safranbolu, Karabuk, Turkey

- [11] S. Metzner, A. Kindt, Determination of the parameters of automated valuation models for the hedonic property valuation of residential properties: A literature-based approach, International Journal of Housing Markets and Analysis, 2017

Intelligent Examination Glove System Design for Use in Medical Education and User Interface Application

Y.ÖZKAN¹ and M. TOZ²

¹ Düzce University, Düzce/Turkey, s.yasin09@gmail.com

² Düzce University, Düzce/Turkey, metintoz@duzce.edu.tr

Abstract - In this study, a smart examination glove system design which is developed with wearable technology approach is recommended for the manual examination application which requires more experience in medical education. The recommended smart glove allows the user to record the pressure information applied by the hand fingers. When the proposed glove is worn by a trainer for an examination and then by a student for the same examination, the pressure information applied by the hand fingers during the examinations are separately recorded. The two obtained application data are compared both statistically and visually with each other by a user interface that is developed. According to the results of the comparison, the student can receive some feedback about how well the examination is performed. Therefore, the student is able to complete manual examination training more accurately and faster. In the scope of this paper, smart examination glove system design and developed user interface are presented with sample data.

Keywords – Medical Education, Wearable Technology, Palpation, Smart Glove System

I. INTRODUCTION

Thanks to being in the technological age, we have several technological opportunities to develop technological products in all the areas such as life, business, health, education, social and so on. Also, today, with the rapid development of technology, the size of system components such as microprocessor / microcontroller, sensor, transmitter and wireless network units in electronic and computer systems are getting to be smaller. In addition to being portable with minimization in size, this situation also helps the wearability feature of such products to become widespread. In this way, a new wearable technology product design is made almost every day. Some examples from the field of medicine can be given as follows. Çetin et al. [1] designed a system that instantly measures the body temperature and pulse information and allows monitoring by the doctor via the internet. Altın et al. [2] monitored the changes in body temperature using infrared sensor technology. Borges et al. (2009) [3] developed a wearable monitoring belt mounted on a flexible sensor to help of the pregnant women. The proposed belt is a mechanism to monitor the movement of the baby. The aim of this study is to establish a system that will provide remote control of low-risk pregnancies. The use of technology in education has a facilitating and sustaining effect on learning [4]. As there are multiple stimuli in the training material used, it helps individuals to perceive their perceptions more clearly and in a more assertive manner. In the literature, it is seen that the wearable technology has been used frequently in applied medical education. Some examples of studies in this area can be given as follows. Goodwin and his friends [5], used Google Glass as a wearable technology product and performed remotely-conducted exercises for medical students' education in obstetric emergencies. A different research in medical education was conducted on robot-assisted gait training [6]. In order to restore motor functions of walking, in the study, it was aimed to re-teach movement with function and job-specific robotic systems [6]. One of the important applications in medical education is patient examination training. In the physical examination of the patients, the general conditions of

the patients are evaluated. All organs are examined by palpation, percussion and auscultation techniques [7]. Palpation; is one of the subheadings of the examinations performed by hands. The patient's temperature, tenderness, swelling and injury should be palpated [8]. In this paper, an intelligent glove system design is recommended which computes the pressure force applied by the specialist physician to diagnose the problem on the patient during palpation. The user interface program developed for this system is also introduced. By using this system in medical education, a student is expected to learn not only by watching the specialist but also by visually seeing the results of his/her trials. The intelligent glove provides instant recording of the pressure information applied through force sensors replaced in the structure. Thanks to the user interface program that communicates with this glove, the pressure profiles of the examinations made by the specialist and the student can be compared with each other. The paper is organized in four sections. In the second section, the proposed smart glove system design is explained, the developed user interface program is introduced in the third section, finally the paper is concluded by the last section.

II. RECOMMENDED INTELLIGENT EXAMINATION GLOVE SYSTEM DESIGN

This section introduces the proposed smart examination glove system design. A general drawing of the proposed design is presented in Figure 1.

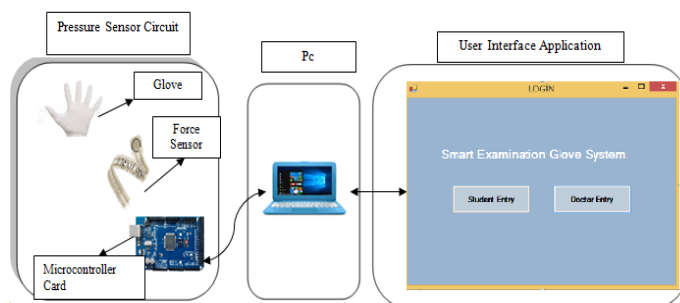


Figure 1: Recommended smart examination glove system design

As shown in the figure, the proposed design consists of an examination glove with force sensors installed in it, a sensor reader card, a computer and a user interface program. Thanks to the sensor-containing glove, the force information from each force sensor is recorded and converted into pressure information depending on the time during the palpation examination and can be displayed instantly through the user interface. In the study presented, the user interface program which is included in the design of this system has been realized. In our study that is supported by Duzce University Scientific Research Projects, the production of smart gloves continues. In the next section, the developed user interface application is introduced using sample data.

III. INTELLIGENT INSPECTION GLOVE SYSTEM USER INTERFACE APPLICATION

This section introduces the user interface program developed for the proposed smart examination glove system design. However, some basic information about palpation examination is provided before introducing the interface. Palpation is the collection of data on the physical state of the patient using the sense of touch. For this purpose, hands are used especially the fingertips. The use of fingertips is mainly due to the presence of nerve endings in these areas [9]. In this examination, firstly some properties of the tissue to be examined should be known such as shape and structure. In this way, the pathologies can be identified by distinguishing them. Palpation is performed by a doctor by touching or pressing the tissue. It gives more information than what can be seen and revealing information about those who cannot be seen. With palpation, the structure, size, density, temperature and functional status of tissue are evaluated [9]. Palpation is a complex examination type that affected from the medical knowledge, the motor skills and the perceptual skills etc. at the same time. Moreover, the attitude of the doctor and the design of the environment that the examination is performed in also play an active role in the palpation process [10]. Palpation is inherently complex because it is an open-ended task seeking an undefined solution [11]. In this study, two samples of palpation were selected to introduce the system. These are abdominal and liver palpations. The basic information about these two examinations can be briefly stated as follows.

A. Abdominal Palpation

For palpation, the patient first takes the supine position. Other fingers, except thumbs of the hand, are held parallel to the abdomen such as being to lightly touch with fingertips. The palpation is started from the right upper quadrant. The hand moves in a clockwise direction and all quadrants are palpated. The hands used for palpation should be warm enough to not disturb the patient [12].

B. Liver Palpation

In the palpation of the liver, the right hand is placed in the abdomen of the patient. Palpation is started by placing on the right lower quadrant [9]. If the liver of the patient is enlarged, the edge of the liver will be touched by the doctor's fingertips. In normal individuals, the edge of the liver is palpated just below the rib edge [9].

These both type of the examinations can be performed by the developed user interface program. The main screen of this program is presented in Figure 2. There are input buttons on this screen and switching to doctor and student screens is possible with the help of these buttons.

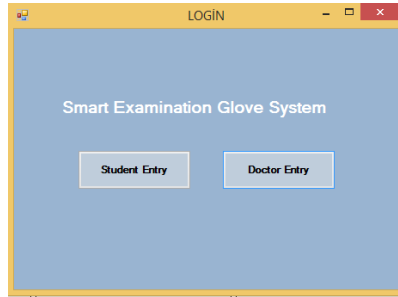


Figure 2: Intelligent examination glove system user interface main screen

When the doctor screen is selected (Figure 3), it is possible to select one of two sample palpation applications from that figure.

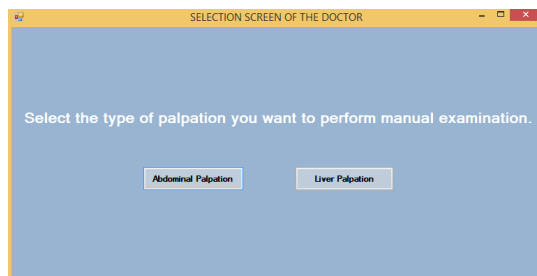


Figure 3: Doctor selection screen

If the abdominal palpation is selected then the doctor abdominal palpation screen is shown (Figure 4).

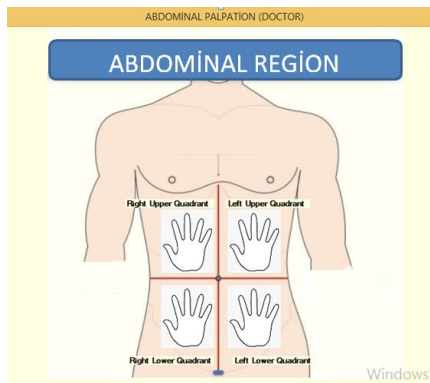


Figure 4: Doctor abdominal palpation screen

The relevant examination screen will be displayed according to the selection from this screen. For example, when the palpation of the right upper quadrant palpation is selected, the screen in Figure 5 will appear.

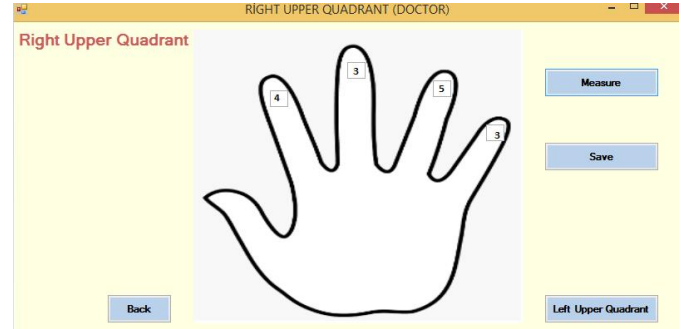


Figure 5: Doctor right upper quadrant palpation screen

Measurements can be made at this stage. The physician wearing the glove presses the "Measure" button before examining the right upper quadrant. Sensors located on each finger are used instantly during the examination. The pressure values calculated with the obtained force values are recorded. After completing the examination, the doctor may also switch to the other quadrants and complete the examination in those areas too. The pressure data obtained after the completion of the examination are recorded under the name of "Expert Data" for evaluation. For example, suppose that the data given in Table 1 is obtained after the palpation of the abdomen performed by the doctor, and the data given in the Table 2 is obtained for the liver palpation examination performed by the doctor.

Table 1: Data of abdominal region right upper quadrant palpation performed by the doctor

Abdominal Region Right Upper Quadrant Palpation Data **	
Time	Palpation Data*
0,5. sn	3,3,4,3
1. sn	4,3,5,3
1,5. sn	3,4,4,3
2. sn	3,4,4,4

. * In the data order in the dials; the first data points to the index finger, the second to the middle finger, the third to the ring finger and the last to pinky finger.

** Pressure data are given in kPa(kiloPascal)

Table 2: Data of liver region right lower quadrant palpation performed by the doctor

Liver Region Right Lower Quadrant Palpation Data **	
Time	Palpation Data
0,5. sn	4,3,4,3
1. sn	4,3,3,4
1,5. sn	4,2,3,3
2. sn	4,4,3,4

** Pressure data are given in kPa (kiloPascal).

After completing the examination performed by the doctor, it is the student's turn. Student will wear the same glove and examine the patient. Similarly, student data is also recorded in a separate field. For example, some sample student data is given in Table 3 and Table 4:

Table 3: Data of abdominal region right upper quadrant palpation performed by the students

Student's, Abdominal Region Right Upper Quadrant Palpation Data **						
Time	1.student	2. student	3. student	4.student	Doctor	Educational Success Rate
0,5. sn	3,3,2,3	3,3,4,3	3,3,3,3	4,4,3,4	3,3,4,3	%25
1. sn	4,3,5,4	4,3,5,3	4,3,3,4	4,3,5,3	4,3,5,3	%50
1,5. sn	3,4,4,3	3,4,4,4	3,4,3,4	3,4,4,3	3,4,4,3	%50
2. sn	3,4,4,4	3,3,3,4	3,4,4,4	3,4,4,4	3,4,4,4	%75
Student Success R.	%50	%50	%25	%75		

** Pressure data are given in kPa (kilopascal).

Table 4: Data of liver region right lower quadrant palpation performed by the students

Student's, Liver Region Right Lower Quadrant Palpation Data **						
Time	1.student	2. student	3. student	4.student	Doctor	Education Success Rate
0,5. sn	4,3,4,3	4,3,4,4	5,4,4,5	5,4,4,4	4,3,4,3	%25
1. sn	5,4,4,4	4,3,3,3	4,3,3,4	4,3,3,4	4,3,3,4	%50
1,5. sn	4,2,3,3	4,2,3,3	4,2,3,3	4,5,4,4	4,2,3,3	%75
2. sn	4,4,3,4	4,4,3,4	4,4,3,4	3,4,4,4	4,4,3,4	%75
Student Success R.	%75	%50	%75	%25		

** Pressure data are given in kPa (kilopascal).

After the examination process is completed, evaluation process is started. The developed user interface compares the data obtained from each student examination with the expert's data. It sets an accuracy rate for each. It is also presented visually for the student which finger is faulty or which performed the correct examination. For example, the evaluation screen for the 1st student is given below.

When the right upper quadrant examination data of the abdomen in Table 3 is examined, the students achieved success rates of 50, 50, 25 and 75%, respectively. Table 4 shows the right lower quadrant palpation data of the liver. In this table, students achieved success rates of 75%, 50%, 75% and 25% respectively.

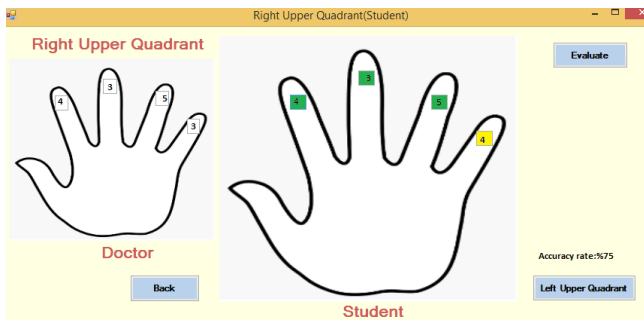


Figure 6: Upper right quadrant examination evaluation screen for the first student

IV. RESULTS AND ADVANCED STUDIES

In this study, the design of a smart examination glove system is recommended. The proposed smart glove makes a comparison between the results of manual examinations performed by the expert and the student. The design of the smart examination glove system and the user interface developed for use with this system are presented with sample data. In the next study, it is planned to perform the physical examination glove of the proposed system as a whole and to test it.

ACKNOWLEDGMENT

This study was supported within the scope of Düzce University Scientific Research Projects Support Program. Thanks to Düzce University for their support. Project Number: 2018.07.01.847

REFERENCES

- [1] Çetin, G. D., Bayılmış, C., Kaçar, S., & Kırbaz, İ. Çevrimiçi Sağlık İzleme Sistemi Uygulaması Application Of An On-Line Medical Monitoring System.
- [2] Altın, E., & Demirel, S. [2016]. Applications Of Infrared Sensors İn Field Of Health Kizilötesi Sensörlerin Sağlık Alanındaki Uygulamaları.
- [3] L. M. Borges, N. Barroca, F. J. Velez, A. S. Lebres, "Smartclothing wireless flex sensor belt network for foetal health monitoring", 3rd International Conference on Pervasive Computing Technologies for Healthcare [PervasiveHealth], London, 1-4, 2009.
- [4] Kabakçı, I., & Odabaşı, H. F. [2004]. Teknolojiyi kullanmak ve teknogerçekçi olabilmek.
- [5] J. Goodwin, R. A. Elkattah, and M. Olsen, "Wearable technology in obstetrical emergency simulation: A pilot study," International Journal of Health Sciences Education, vol. 2, no. 2, p. 3, 2014.
- [6] Demir, S. Ö. (2015). Omurilik Yaralanmalı Hastalarda Robot Yardımlı Yürüme Eğitimi. Turkish Journal of Physical Medicine & Rehabilitation/Türkiye Fiziksel Tıp ve Rehabilitasyon Dergisi, 61.
- [7] Birol L [2002]. Hemşirelik Süreci. 5. Baskı, Etki Matbaacılık Yayıncılık Ltd. Şti., İzmir.
- [8] Erkinç, M. Vaka Değerlendirme ve Fizik Muayene.
- [9] Kaymaz, M. [Online] Available: <http://med.gazi.edu.tr/posts/download?id=98373>
- [10] Brydges R, Butler D. A reflective analysis of medical education research on self-regulation in learning and practice. Med Educ 2012; 46:71e9.
- [11] Fossum C. Palpation: from past to present. In: Teaching palpation. Potsdam: OsEAN; 2011.
- [12] Gül, Y., & İssi, M. (2015). Karın Bölgesi Ve Karın Boşluğunun Muayenesi. Türkiye Klinikleri Journal Of Veterinary Sciences-Internal Medicine-Special Topics, 1(2), 1-8.

Brain Tumor Detection via Active Contours and Scale Invariant Feature Transform

P.GÖRGEL¹

¹ Istanbul University-Cerrahpasa, Computer Eng., Istanbul/Turkey, paras@istanbul.edu.tr

Abstract - In cancer diagnosis, segmentation of brain tumors reliably and accurately is essential for evaluation and treatment planning. In this study a novel brain tumor detection method is developed by integrating morphological pre-processing, Scale Invariant Feature Transform (SIFT) based feature extraction and classification. Firstly morphological operations are applied to the raw brain MR images for pre-processing. Then initial contours are obtained to implement active contours using median filtering and threshold base rules. After that, segmentation of the suspicious parts of the images via active contours method is done using the initial contours. A further feature extraction step is applied to the segments to achieve classification. For feature extraction Scale Invariant Feature Transform (SIFT) is implemented. Finally the segments are classified by Support Vector Machines (SVM) using these features to eliminate false positive tumors. According to the experimental results the performance evaluators like system accuracy, F1-score, precision, true positive rate and true negative rate are 0.85, 0.88, 0.85, 0.92 and 0.83 respectively.

Keywords - Brain tumor, scale invariant feature transform, active contours, support vector machines.

I. INTRODUCTION

BRAIN tumors are a particular group of neoplasm growing in the central nervous system and affect up to one third of patients with cancer. Tumor cells are comprised of brain cells around the membranes of the brain, nerves or glands and morphologically different. Tumor cells have lower contrast compared to around tissue cells in brain. By the help of Magnetic Resonance Imaging (MRI) technology the abnormalities and contrast variation are identified [1].

Computer aided tumor detection in medical images is essential for stages such as biopsy, radiotherapy planning and more recently, for advanced biomedical image analysis techniques. Tumor segmentation is also a difficult process because inadequate contrast is not able to emphasize the variation between tumor regions and the normal tissue around. Today for brain images MRI technology is preferred much more than Computer Tomography (CT) due to its high sensitivity and high contrast with respect to various intensity tissues.

Recently, many CAD (Computer Aided Diagnosis) techniques have been used to deal with brain tumors. Reza et al. [2] proposed a brain tumor type classification using Multi-Fractal Detrended Fluctuation Analysis (MFDFA) and texture features known as multi-fractional Brownian motion in MR images based on Random Forest method.

Nabizadeh et al. [3] utilized from Histogram-based Gravitational Optimization Algorithm (HGOA) to detect brain tumors. Ting et al. [4] proposed a brain lesion segmentation method to segment the region pertaining to brain injury using Adaptive Tuning Spatial (ATS) pre-processing and case-control atlas and parametric based segmentation. A multiclass brain tumor classification was done by Sachdeva et al. [5], using PCA-ANN (Principal Component Analysis – Artificial Neural Network) approaches. The initial contour was determined by Content Based Active Contour to find texture and intensity values outside and inside ROI (Region of Interest).

In the proposed computer aided tumor detection system morphological operations are applied to the raw brain MR images for pre-processing. Then initial contours are obtained to be able to implement active contours using median filtering and threshold base rules. After that, the segmentation of the suspicious parts of the images via active contours is done. A further feature extraction step is applied to the segments for the classification step. For feature extraction Scale Invariant Feature Transform (SIFT) is implemented. Finally the segments are classified by Support Vector Machines (SVM) according to these features to eliminate false positive tumors.

II. MATERIALS AND METHODS

A. Morphological Opening for Enhancement

Opening is one of the basic morphological operators in image processing. The properties of geometrical continuous-space such as geodesic distance, convexity, size, shape, connectivity are the interest area of morphological operations [6].

Opening is made with dilation of the erosion of a set A by a structuring element B. It can be considered as morphological noise removal by removing small regions from the foreground of an image. They are placed in the background in opening. On the other hand closing eliminates small holes in the foreground. These techniques also work to find the defined shapes in an image. Specific structuring elements can be used in opening to detect edges and corners. The opening operation is represented in (1) where \oplus and \ominus denote dilation and erosion respectively.

$$A \circ B = (A \ominus B) \oplus B \quad (1)$$

B. Median Filtering for Noise Removal

Median filtering is used for reducing noise in an image processing application. While doing this, it cares to keep the fine details in the image. A median filter takes into account the neighborhoods to calculate each pixel's value. The neighbor pixels are sorted and the value in the middle of an array is taken as the new value. If there are even numbers of pixels in the region of interest, the average of the two pixels in the middle is used as the median [6]. In (2) w is the required neighborhood and $[m, n]$ is the central location in the image.

$$y[m, n] = \text{median}\{x[i, j], (i, j) \in w\} \quad (2)$$

C. Tumor Segmentation using Active Contours (Snakes)

Active contours [7] are a set of points to be used for feature extraction. Active contours enclose a target feature interactively to describe it. Firstly an initial contour is acquired and later this contour is evolved to enclose it. The goal is to find the optimum contour which shrinks to get closer to the target. To match the feature the contour points are adjusted after sequential iterations. Anyone can consider the Active contours process as a minimisation of the energy. The target feature is a minimum of an appropriate energy functional which includes properties that control the way the contour can stretch and curve. Hence the active contour represents a compromise between its own properties such as bend and stretch and image properties such as the edge magnitude. The energy functional includes the contour's internal energy (E_{int}), its constraint energy (E_{con}) and the image energy (E_{image}) functions. A snake consists of these functions and $v(s)$ represents the set of x and y co-ordinates of the points which make up that snake. The energy functional E_{snake} (3) is the integral of the active contour functions and $s \in [0, 1]$ is the normalised size [7]:

$$E_{\text{snake}} = \int_{s=0}^1 E_{\text{int}}(v(s)) + E_{\text{image}}(v(s)) + E_{\text{con}}(V(s)) ds \quad (3)$$

For internal image energy a weighted summation of first- and second-order derivatives is computed as demonstrated in (4). The first-order differential measures the elastic energy due to stretching. When this value is high it suggests a high rate of change in that part of the contour. The second-order differential measures the bending energy due to the curvature features. The parameter $\alpha(s)$ arranges the contribution of the elastic energy according to point spacing and $\beta(s)$ arranges the contribution of the curvature energy according to point variation. If α is low, the points can change greatly in spacing. Otherwise snake aims to attain evenly spaced contour points. When β is low, curvature may not be minimised and contour can evolve to corners and when it is high the corners may soften.

$$E_{\text{int}} = \alpha(s) \left| \frac{dv(s)}{ds} \right|^2 + \beta(s) \left| \frac{d^2v(s)}{ds^2} \right|^2 \quad (4)$$

D. Scale Invariant Feature Transform Feature Extraction

Scale Invariant Feature Transform (SIFT) is mostly used for feature extraction in pattern recognition applications. The main benefit of SIFT algorithm is it is scale and orientation invariant. It is used in digital images to detect and extract local features [8]. SIFT produces a large set of features from the images, which reduces local variation errors in the average error of all errors in feature matching. SIFT is able to identify objects in case of partial occlusion among clutter. SIFT is processed in four steps: Scale-space extrema detection, keypoint localization, orientation assignment, keypoint descriptor.

- Scale-space extrema detection: The keypoints are the points of interest in an image. In this stage Gaussian filter is used for convolution in different scales before the difference of successive Gaussian-blurred images. Difference of Gaussians (DoG) image $D(x, y, \sigma)$ at multiple scales is given by

$$D(x, y, \sigma) = L(x, y, k_i \sigma) - L(x, y, k_j \sigma) \quad (5)$$

In (6), $I(x, y)$, $L(x, y, k\sigma)$, $G(x, y, k\sigma)$ and $k\sigma$ represent original image, convolution of the original image and Gaussian blur and scale respectively [8].

$$L(x, y, k\sigma) = G(x, y, k\sigma) * I(x, y) \quad (6)$$

To detect scale space extrema the convolution of the images with Gaussian-blurs in different scales is implemented.

- Keypoint localization: As a result of the detection of keypoints, many endpoints that are not actually keypoints may be detected. Such ambiguous points should not be included in the key points set. Hence with the help of some filters and thresholds, low contrast and weak keypoints are eliminated [8]. $D(x)$ represents the Taylor expansion in (7) and X is the computed offset in (8).

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \quad (7)$$

$$X = (x, y, \sigma)^T \quad (8)$$

E. Support Vector Machines Classification for Decision of Tumor / Normal Tissue

Support Vector Machine (SVM), introduced by Vapnik in 1995 [9], is a frequently used data classification function. The goal of an SVM is to implement a hyperplane as the decision surface by maximizing the separation margin between the negative and positive examples. To determine an optimized SVM classifier boundary in a given label pairs (x_i, y_i) , the relation should be maximized in (9). Here P is a user defined positive parameter to control the tradeoff between the complexity and non-separable points.

$$L(c) = \sum_{i=1}^l c_i - \frac{1}{2} \sum_{i,j=1}^l y_i y_j c_i c_j K(x_i, x_j), \quad 0 \leq c_i \leq P \quad (9)$$

$$w = \sum_{i=1}^N c_i y_i x_i, \quad c_i [y_i (w^T x_i + b) - 1 + \xi_i] = 0 \quad (10)$$

l , $K(x_i, x_j)$ and c_i are the sample number, SVM kernel and Lagrange coefficient respectively. The slack variable ξ_i in (10) is used to relax the canonical hyperplane equation's constraints.

F. Proposed Approach

In the proposed approach morphological opening with a structural element of disk shape is applied to the raw brain MR images for enhancement. Then initial contours are obtained to implement active contours (snakes) using median filtering and threshold based rules. After that, the segmentation of the suspicious parts of the images via active contours (snakes) is provided. A further feature extraction step is applied to the segments for the classification step. SIFT based feature extraction is preferred and the segments are classified by SVM according to those features to determine whether the segmented regions are tumors or normal tissue (Figure 1).

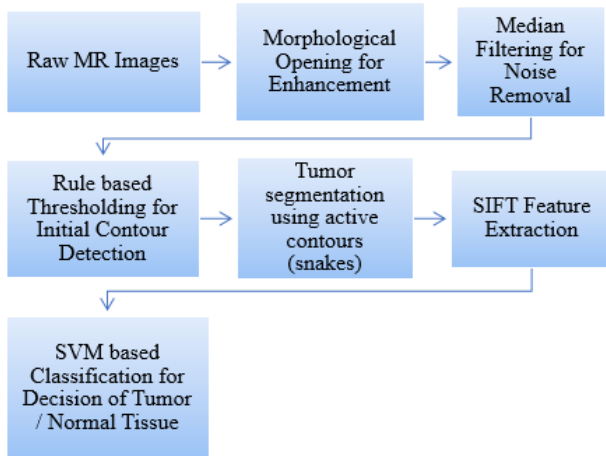


Figure 1: The flow chart of the proposed system

III. EXPERIMENTAL RESULTS

In this study a brain dataset including 20 brain images from BRATS 2015 [10-12] was used for training and testing sets. The proposed system segmented 24 brain tumors and 23 normal tissue after the detection process. These 47 regions were given to SVM classification. The system produced a confusion matrix as seen in Table 1. The performance parameters are also demonstrated in Table 2. Before classification, the morphological opening for enhancement, median filtering for noise removal, rule based thresholding for initial contour detection and tumor segmentation using active contours (snakes) process results are given in Figure 2 and Figure 3.

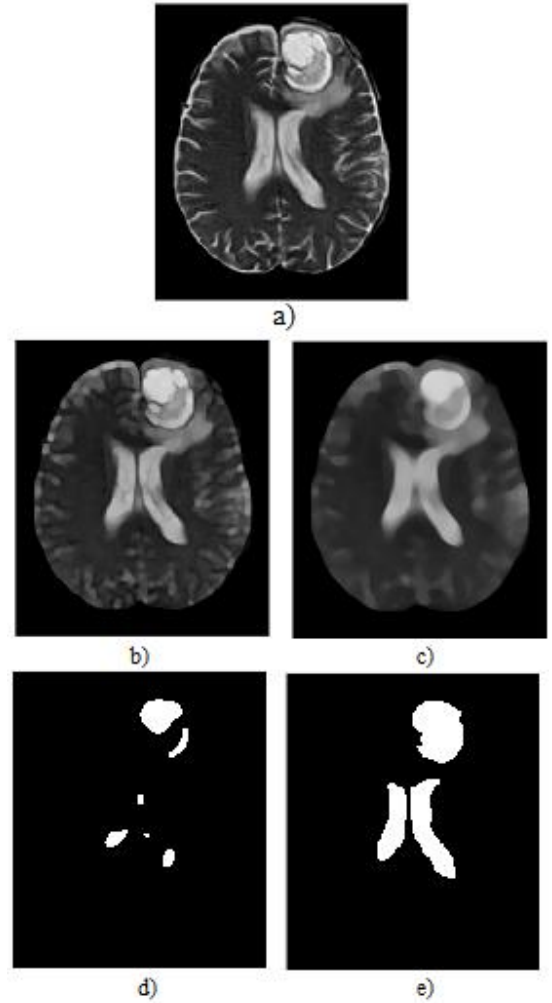


Figure 2: a) Raw image b) Morphological opening c) Median filtering d) Rule based threshold for the initial contours e) Segmented image with Active contours- The region at the top is a tumor, the two regions at the bottom are normal tissue

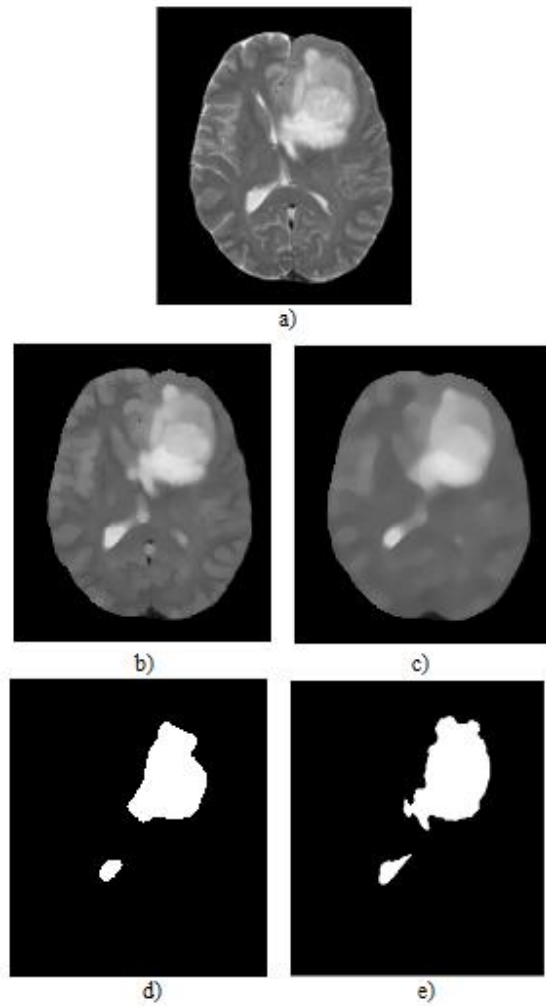


Figure 3: a) Raw image b) Morphological opening c) Median filtering d) Rule based threshold for the initial contours e) Segmented image with Active contours- The region at the top is a tumor and the region at the bottom is normal tissue

Table 1: The confusion matrix

Tumor	TP: 22	FN: 2
Normal Tissue	TN: 19	FP: 4

Table 2: Performance parameters

Accuracy	True Positive Rate	Precision	F1 Score	True Negative Rate
0.85	0.92	0.85	0.88	0.83

IV. CONCLUSION

In this computer aided brain tumor detection system morphological operations and median filtering are used to enhance the image emphasize disk shaped objects and remove the noise for optimum segmentation. After pre-processing the

initial contours of the suspicious regions are obtained. These contours are given to active contours algorithm and the segmented regions are provided. To eliminate the false positive detections a classification step is needed. Therefore a number of features are extracted via SIFT and the segments are classified by Support Vector Machines (SVM) using these features. According to the experimental results, the performance evaluators like system accuracy, F_1 -score, precision, true positive rate and true negative rate are 0.85, 0.88, 0.85, 0.92 and 0.83 respectively.

REFERENCES

- [1] S. Pereira, A. Pinto, V. Alves, C.A. Silva, "Brain tumor segmentation using con- volutional neural networks in MRI images," *IEEE Trans. Med. Imaging*, vol. 35, pp. 1240–1251, 2016.
- [2] S.M. Reza, R. Mays, K.M. Iftekharuddin, "Multi-fractal detrended texture feature for brain tumor classification," in *March 2015, SPIE-The International Society for Optical Engineering Conf.*, DOI: 10.1117/12.2083596.
- [3] N. Nabizadeh , N. John , C. Wright , "Histogram-based gravitational optimization algorithm on single MR modality for automatic brain lesion detection and segmentation," *Expert Syst.Appl.*, vol.41 (17), pp. 7820–7836, 2014.
- [4] F. F. Ting, K. S. Sim, C. P. Lim, "Case-control comparison brain lesion segmentation for early infarct detection," *Computerized Medical Imaging and Graphics*, vol. 69, pp. 82–95, 2018.
- [5] J. Sachdeva, V. Kumar, I. Gupta, N. Khandelwal, C.K. Ahuja, "Segmentation, feature extraction, and multiclass brain tumor classification," *J. Digit. Imaging*, vol. 26, pp. 1141–1150, 2013.
- [6] R. C. Gonzalez, R. E. Woods, *Digital Image Processing (Second Edition)*, USA: Prentice Hall, 2002.
- [7] M. S. Nixon, A. S. Aguado, *Feature Extraction and Image Processing*, England: Newnes, 2002.
- [8] D. G. Lowe, "Object recognition from local scale-invariant features," in *1999 International Conference on Computer Vision*, vol. 2, pp. 1150–1157.
- [9] C. Cortes, V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.
- [10] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imaging*, vol. 34 (10), pp.1993–2024, 2015.
- [11] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, JS Kirby, JB Freymann, K. Farahani, C. Davatzikos, "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
- [12] V. Rao, M. Sarabi, A. Jaiswal, "Brain tumor segmentation with deep learning", in: *2015 MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS)*, pp.56–59.

Spatial Preference System For Roads Maintenance

S.K.M. ABUJAYYAB¹, I.R. KARAS¹, Ozlem DEMIRCAN¹ and Hanife INCE¹

¹ Karabuk University, Karabuk/Turkey, s.jayyab@hotmail.com

¹ Karabuk University, Karabuk/Turkey, irkaras@gmail.com

¹ Karabuk University, Karabuk/Turkey, demircanozlem4@gmail.com

¹ Karabuk University, Karabuk/Turkey, incehanife@gmail.com

Abstract – Information technologies have made great progress in the last 50 years and have started to be at the center of our lives. Information can be access quickly, share it easily and efficiently.

The objective of this paper is to develop spatial preference system for roads maintenance valuation operations. The tool will determine the road damage percentages based on relevant parameters such as, distance from main roads, distance from city center, distance from city halls, distance from schools, distance from markets, distance from bus terminals and distance from hospitals. In this study, the author has adopted GIS application software – ArcMap. The tool developed in ArcGIS-Arctoolbox environment using Python programing language. multi criteria decision analysis MCDA used to evaluate preference. Data were mange, processed, and assist through the tool. The tool effectively and systematically evaluates roads for maintenance. Safranbolu-Karabuk data is used as a case study to further clarify the application of tool in road maintenance valuation. The tool prove that it is capable to effetely, accurately and speedily evaluate the roads for maintenance operations.

Keywords – Road maintenance, ArcMap, Geographic Information System.

I. INTRODUCTION

The nations development and economic growth are closely associated to its accessible transportation system. Efficient road transportation infrastructure abilities promote industrial and socio-economic development [1]. To provide safety, good condition and comfort roads to drivers at all times, a comprehensive and schedule roads maintenance must be formulating and adopt. Preventive maintenance works, such as road rehabilitation help to reduce the major road repairs, expenditure, accidents and reduce the traffics [2].

Suitable Geographical information systems GIS database is necessary step to effectively conducting the road preference for maintenance. As a result of the integration of spatial data with tabular data, GIS is strongly recommended to perform road preference valuation. GIS has been use to service the human to make their life easier, accurate and faster [3].

GIS is an information system that performs several functions such as collecting, storing, processing and presenting the graphic and non-graphic information obtained from several sources. GIS enable us to record, edit, structure, model and analysis our complex, location-based information [4]. Along with the developing in GIS Technologies, roads maintenance

system still need several improvements. There is a need for a fast, economical and accurate system for redundant roads maintenance valuation operations [2].

The purpose of this paper is to develop spatial preference system for roads maintenance valuation operations. The tool will determine the road damage percentages based on relevant parameters. The development of this tool will help the roads maintenance departments in municipalities to manage the limited budget and guide the maintenance operations for the most urgent roads. The tool

II. METHODOLOGY

This section will mainly focus on the methodology that involves handling and managing data. Within the process of data collection, two sets of data are classified; raster data and attribute data. Raster data are achieved. Processes involved in this research are illustrated in Figure 1.

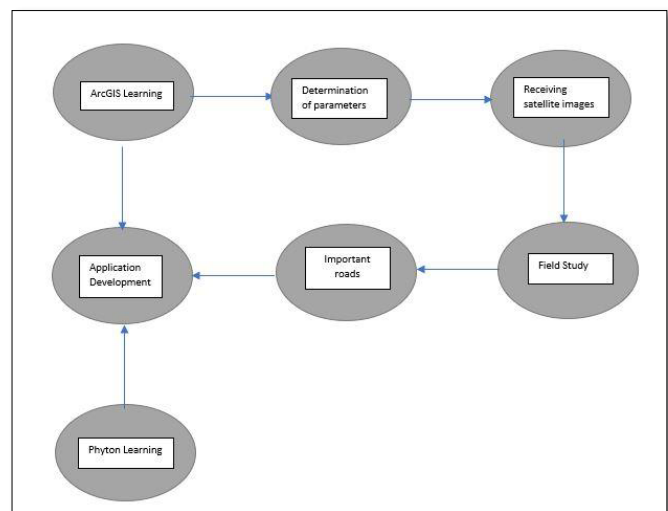


Figure 1 Research Methodology

A. Generating Vector Data

In this study, ArcMap software used to construct the vector data [5]. Roads represented as line segment for evaluations. Essentially, all most relevant factors for road valuation constructed. Each factor represented as vector layer in ArcMap. Roads represented as line segments for evaluations. Other parameters were determine based on experts such as main streets, hospital, park, market, city hall, and school etc.

B. Generating Raster Data

In raster dataset, each cell (which is also known as a pixel) has a value. The cell values represent the phenomenon density. In this tool, raster data calculated from vector data convert using the Python programming language. For each road segment, the distance from evaluation parameters (distance from main roads, distance from city center, distance from city halls, distance from schools, distance from markets, distance from bus terminals and distance from hospitals) were calculated automatically.

C. Attribute Data Extraction

Essentially, attribute tables are constructed for each road segment from all input raster parameters (distance from main roads, distance from city center, distance from city halls, distance from schools, distance from markets, distance from bus terminals and distance from hospitals). Data extracted based on the develop tool using Python language.

D. MCDA analysis

In this step, vector road segment data with its attributes data has been feed to this stage. The end-user need to decide and obtaining the final preferences by applying MCDA.

E. Tool development

The tool was developed in Arctoolbox environment using python programming language. The tool exhibited in Figure 2,

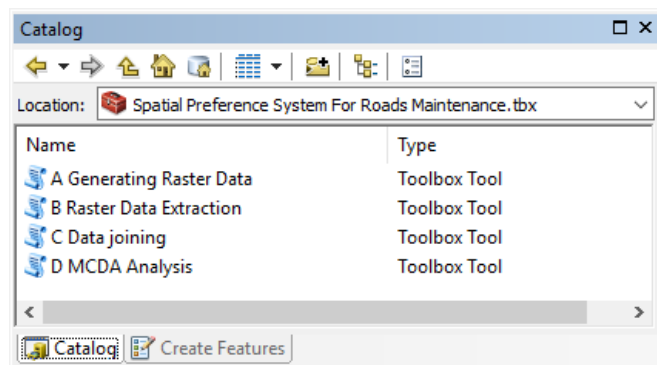


Figure 2 Spatial Preference System For Roads Maintenance

III. APPLICATION

In this study, Safranbolu-Karabuk data is used as a case study to further clarify the application of tool in road maintenance valuation. Necessary parameters determined for roads maintenance in Safranbolu. Then, parameters data were generated for Safranbolu in ArcMap. Satellite images used to generate the parameters maps and roads segment map as illustrated in Figure 3.

Our parameters: distance from main roads, distance from city center, distance from city halls, distance from schools, distance from markets, distance from bus terminals and distance from hospitals. The reason of choosing these parameters is to find out which way is more important to make maintenance. In addition, maintenance will be according budget set by the municipality.

After creating spatial data of parameters, raster data generated and tabular data extracted. The last stage is to perform MCDA in order to calculating most important road segment for maintenance.

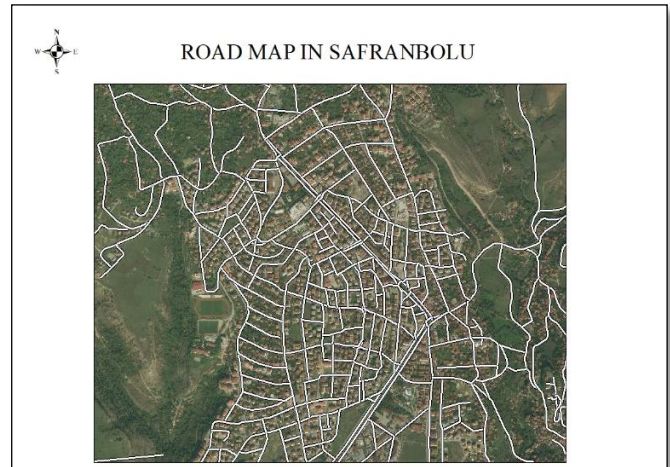


Figure 3 Safranbolu-Karabuk study area

IV. CONCLUSION

The adoption of GIS will lead to a more organized management of digital data especially those related to road data. Particularly, this system application will also increase work productivity in managing road maintenance.

ACKNOWLEDGMENT

This study has been supported by 2221 – Fellowship Program of TUBITAK (The Scientific and Technological Research Council of Turkey). We are indebted for their supports.

REFERENCES

- [1] Coğrafi Tabanlı Karayolları Bakım Yönetim Sistemi Oluşturulması – Ankara, 2013.
- [2] Managing Road Maintenance Using Geographic Information System Application, Mohd Zulkifli B. Mohd Yunus, Hamidah Bt. Hassan, 2010.
- [3] Journal of Geographic Information System, 2010, 2, 215-219.
- [4] <http://sbpturkiye.com/cografı-bilgi-sistemleri.html>
- [5] <http://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/what-is-raster-data.htm>.

Spatial Preferences Decision Tool For Historical Building Restoration

S.K.M. ABUJAYYAB¹, I.R. KARAS¹, K.SEN¹, B.BAKAR¹, F.UNAL¹

¹ Karabuk University, Karabuk/Turkey, s.jayyab@hotmail.com

¹ Karabuk University, Karabuk/Turkey, irkaras@gmail.com

¹ Karabuk University, Karabuk/Turkey, kbrasenn@hotmail.com

¹ Karabuk University, Karabuk/Turkey, busra_bakar@hotmail.com

¹ Karabuk University, Karabuk/Turkey, feyzauanal95@gmail.com

Abstract - Historical buildings in Safranbolu is one of the most popular historical places in Turkey. These historical buildings determined by UNESCO that it should be preserved and transferred to future generations as a world heritage site. The purpose of this paper is to develop spatial preferences decision tool for historical building restoration. This tool will contribute to administrations of historical buildings to perform evaluation and find most deserve building for the restoration based on several factors. Evaluation factors data were collected through GIS software. Python programming language were used to develop the code of the tools. Multi criteria decision analysis MCDA used to evaluate preference. Final output will be the percent of house damage. The final tool will be among ArcGIS-ArcToolbox. The tool systematically assesses historical buildings for restoration. The tool demonstrate that it is able to accurately and rapidly estimate the preferences of historical buildings.

Keywords - Multi criteria decision analysis, decision support system, Historical buildings restoration, damage rate.

I. INTRODUCTION

Throughout the history of turkey, several civilizations inherited from Ottoman Empire stayed untill curen data. Turkish architecture is a representatio of richness of Turkish civilization history. Historical buildings in Safranbolu are one of most unique remained archaeologic sites. The Turkish Houses are unique due to it hava multi-storey, different roof style and construction system of wooden roof.

These historical buildings determined by United Nations Educational, Scientific and Cultural Organization UNESCO. UNESCO recommend that it should be conserved and transmitted to upcoming generations as a world heritage site. However, in addition to owning these riches, it is necessary to systematicly and comprehensively record of surviving historical textures for future generations. Additionally, it is prerequisite to adhering the original form. For this sreason, relevant administrations have limited budget for historical buildings restoration. The main challenge is that restoration operations have limited budget. Developing preferences decision tool is urgent need.

The objective of this paper is to develop spatial preferences decision tool for historical building restoration. This tool will make it easier to find the most deserve buildings for the limited budget of administrations by autumatically calculating the percentage of deserve buildings for restorations based on their construction conditions.

II. METHODOLOGY

This methodology section explining handling and managing spatial data. Processes involved in this research are presented in Figure 1.

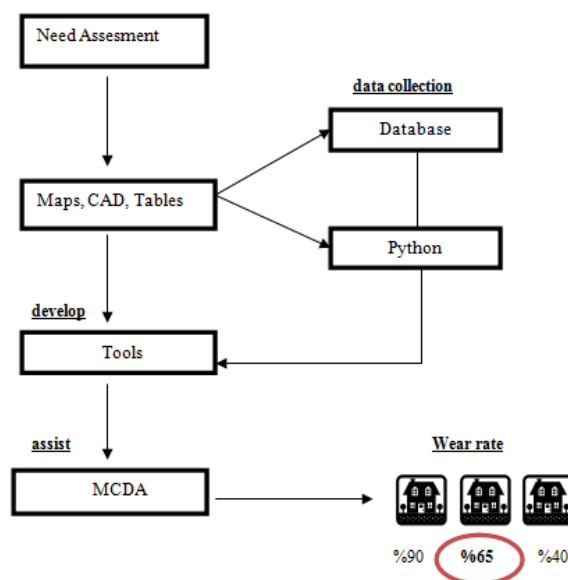


Figure 1 Research Methodology

ArcGIS toolbox using Pyhton language from the database of Safranbolu buildings used to develop the tool. The tool exhibited in Figure 2.

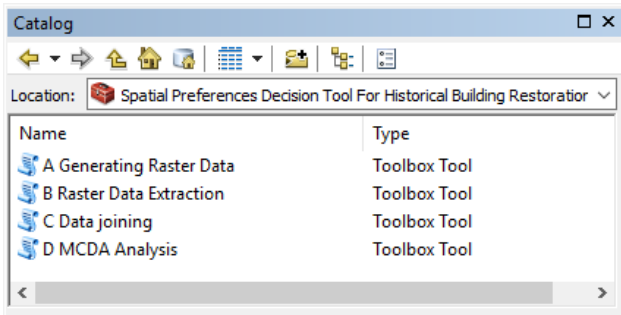


Figure 2 Spatial preferences decision tool for historical building restoration

Based on the consultation of local experts, primary factors were adopted. These factors are based on the construction stutes of buildings (Figure 3) such as walls of the buildings, number of floors, number of balcony and distance from transportation . houses close to tourist areas has been taken into consideration. Several historical buildings in Safranbolu in Eskişehir examined and location collected by using GPS and added in geodatabase. Based on evaluation parameters, the infrastructure of the selected houses were examined and the damage rates of the buildings were calculated and presented in Figure 4. damage rates based on MCDA.



Figure 3 historical house status

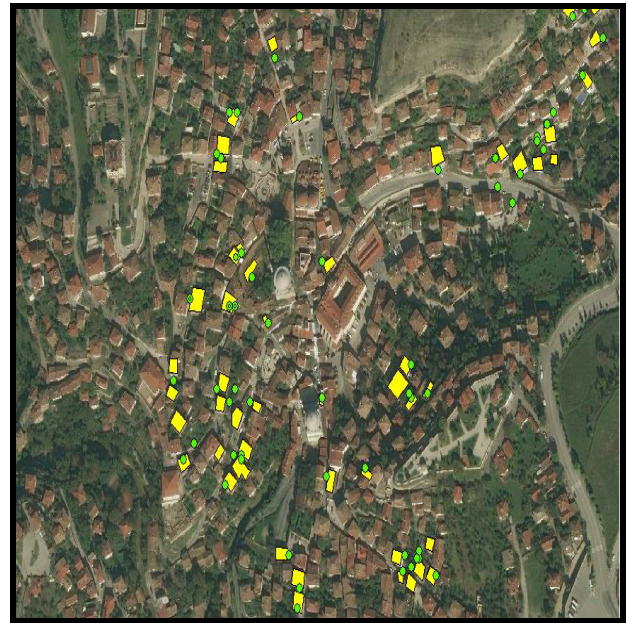


Figure 4 evaluated historical house

CONCLUSION

Safranbolu historical buildings determined by UNESCO that it should be preserved and transferred to future generations as a world heritage site. In the system, historical buildings attribute information and spatial information about objects in different layers have been transferred to the system. The final tool will be among ArcGIS-ArcToolbox. The tool systematically assesses historical buildings for restoration. The tool demonstrate that it is able to accurately and rapidly estimate the preferences of historical buildings.

REFERENCE

- [1] Ö. Ergin, H. Akçın, S. Karakış, H. Şahin Zonguldak Karaelmas Üniversitesi, Jeodezi ve Fotogrametri Mühendisliği Bölümü, 67100 Zonguldak, hakanakcin@karaelmas.edu.tr
- [2] MDPI and ACS Style Della Spina, L.; Lorè, I.; Scrivo, R.; Viglianisi, A. An Integrated Assessment Approach as a Decision Support System for Urban Planning and Urban Regeneration Policies. *Buildings* 2017, 7, 85.
- [3] ISPRS Ann. Photogramm. Remote Sens. *Spatial Inf. Sci.*, IV-4/W7, 19-25, 2018 <https://doi.org/10.5194/isprs-annals-IV-4-W7-19-2018> © Author(s) 2018. This work is distributed under the Creative Commons Attribution 4.0 License.
- [4] Putting Urban Regeneration Projects in Order: An Example of a Decision Support System
- [5] Lin LC., Watada J. (2010) Building a Decision Support System for Urban Design Based on the Creative City Concept. In: Jain L.C., Lim C.P. (eds) Handbook on Decision Making. Intelligent Systems Reference Library, vol 4. Springer, Berlin, Heidelberg

Satellite Images Classification in Geographic Information Systems

S.K.M. ABUJAYYAB¹, I.R. KARAS¹, B.N. BOGAZ¹, F. ZENCIR¹ and M. GONUL¹

¹Karabuk University, Karabuk/Turkey, s.jayyab@hotmail.com

¹ Karabuk University, Karabuk/Turkey, irkaras@gmail.com

¹ Karabuk University, Karabuk/Turkey, busranurbogaz@hotmail.com

¹ Karabuk University, Karabuk/Turkey, fundazencir.fz@gmail.com

¹ Karabuk University, Karabuk/Turkey, mihriban556gnl@gmail.com

Abstract – Satellite image classification is a significant portion of the pattern recognition, image analysis and remote sensing RS. The aim of this study is to classify the Sentinel-2A satellite image to produce build up area maps by using geographic information systems GIS. Build up maps classification is carried out based on supervised classification. ArcGIS software were used to perform the analysis. Düzce area used as cases study to produce the outcomes. The performance accuracy was 85% for classification. The result showed that classification in GIS environment can be applied easily, accurately and without complexity for end-users.

Keywords - satellite images classification, Sentinel-2A images, build up area, GIS, RS.

I. INTRODUCTION

Remote sensing became significant and effective technology for land cover data gathering and analysis. Remote sensing can quickly provide reliable data about the earth at different resolutions and different time intervals. The outcomes of data collection stage in remote sensing is satellite images.

In order to obtained useful data (for instance maps of land cover and urban areas) from raw satellite image, several steps need to be apply. Numerous statistical analyzes and statistical interpretation techniques utilized to gain information about urban areas from the raw satellite images. The most important method used to urban areas is to classify the images of remote sensed data.

The general purpose of the classification process is to group objects with the same spectral characteristics on the earth. The goal is to separate each pixel in the satellite images into different groups according to their spectral characteristics. Then, assign the pixel to the corresponding cluster on the earth according to their reflectance values.

In the classification process, two approaches, commonly referred to as controlled and uncontrolled classification are used. The controlled classification approach, which performs classification by using the training data set, is the most preferred method due to its high performance accuracy [1].

ArcGIS software contains both database manager and set of spatial functionalities. ArcGIS allow to end-users to store and analysis spatial data of valuation factors across the desired geographic landscape [2]. So, ArcGIS is suitable environment to perform satellite image classification.

The objective of this paper is to perform satellite image classification in order to produce urban areas maps. Classification environment is geographic information systems GIS.

II. METHODOLOGY

This section handling the methodology of applying satellite image classification in geographic information systems environment Figure 1. Düzce city in north west of turkey used as cased study. Sentinel-2A satellite image used to perform classification and produce urban areas map. Sentinel-2A image is Multispectral image and consisting 12 different spectral bands.

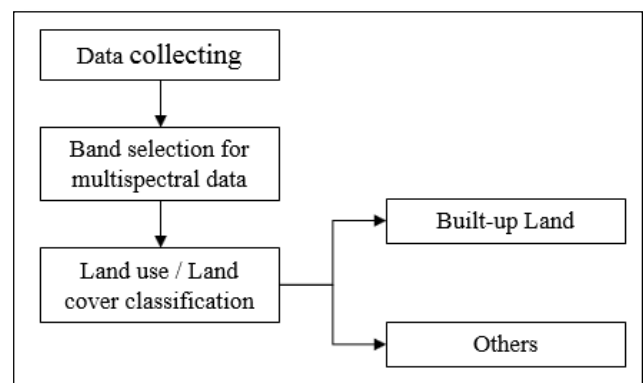


Figure 1 Satellite image classification methodology

2.1. Satellite image collection

Sentinal-2A image obtained from United States Geological Survey (USGS) earth explorer website. Based on the

availability of the data, an up to date (March 5, 2019.) and not cloudy image downloaded Figure 2. Then, image migrated to ArcGIS software.



Figure 2 Satellite image collection (Sentinal-2A image collection)

2.3 Urban areas classification

The most important stage in this paper is to perform satellite image classification. The first step is build the training samples. The samples were built in ArcGIS training sample manger Figure 3. Samples were classified to two classes (urban or building areas and others).

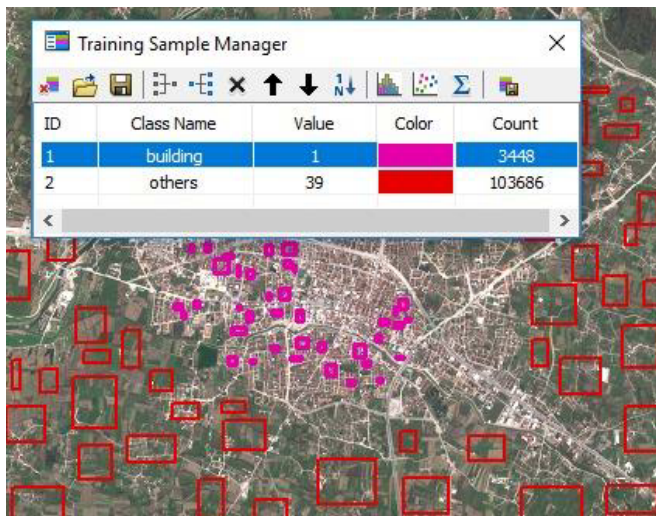


Figure 3 ArcGIS training sample manger

The second step is to performing the classification using any classification method. In this paper Maximum Likelihood Classification tool utilized to execute the classification for Düzce study area Figure 4. After train the model, all pixels in Düzce- Sentinal-2A image assigned to the corresponding classes. Since there is target data in the training sample manger, thus this is controlled classification.

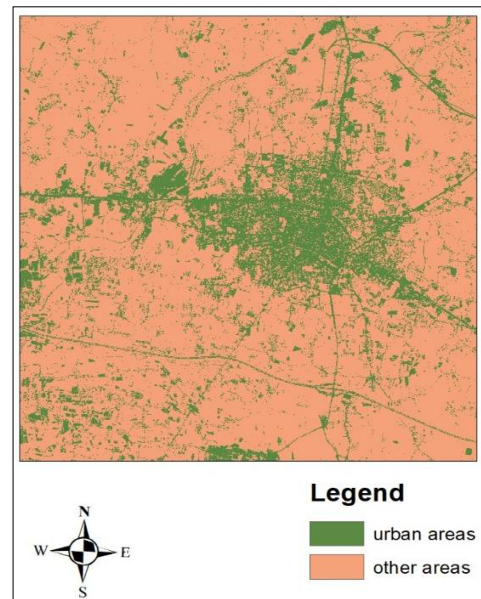


Figure 4 urban areas of Düzce study area after Sentinel-2A satellite image classification

III. DISCUSSION AND CONCLUSION

Spatial information about urban areas is an important factor for decision makers to perform planning. The aim of this article is to implement satellite image classification for urban areas in GIS environment. Sentinel-2A satellite image used to perform classification. Sentinel-2A satellite image obtained from USGS earth explorer website. Cloudy and unclear satellite images is not suitable for image classification and producing urban areas maps. ArcGIS training sample manger used to build the training samples. Maximum Likelihood Classification tool utilized to execute the classification for Düzce study area. The consequence exhibited that performing satellite image classification in GIS environment can be implement straightforwardly, precisely and without difficulty for end-users.

ACKNOWLEDGMENT

This study has been supported by 2221 – Fellowship Program of TUBITAK (The Scientific and Technological Research Council of Turkey). We are indebted for their supports.

REFERENCES

- [1] http://www.irfanakar.com/turkish/sunumlar/UA_CBS/Uzaktan%20Algýlama%20ERDAS%20Bilgi/uydu_goruntulerinin_siniflan_dirilmesi_dogruluk_analizi.PDF
- [2] S. W. Gatheru, D. Nyika, Application Of Geographic Information System In Property Valuation, International Journal of Scientific & Technology Research Volume 4, Issue 08, August 2015
- [3] Hagen, Nathan; Kudenov, Michael W. "Review of snapshot spectral imaging technologies". Spie. Digital Library. Optical Engineering. Archived from the original on 20 September 2015. Retrieved 2 February 2017.

Machine Learning And Satellite Images For Agricultural Areas Determination

S.K.M. ABUJAYYAB¹, I.R. KARAS², G. ERBEK³ and G. KILINÇKAYA⁴

¹Karabuk University, Karabuk/Turkey, s.jayyab@hotmail.com

²Karabuk University, Karabuk/Turkey, irkaras@gmail.com

³Karabuk University, Karabuk/Turkey, miraerbek@gmail.com

⁴Karabuk University, Karabuk/Turkey, kilincKayagizem@gmail.com

Keywords - Machine learning, Sentinel-2A, Satellite images Classification, Agricultural areas.

Abstract -The determination of agricultural areas, agricultural products and their yields distribution are extremely important in directing domestic and foreign market conditions. This paper introducing the ability of using machine learning and satellite images to determine agricultural areas. Classification of satellite images for agricultural areas maps is performed based on supervised classification. ArcGIS environment utilized to applied the analysis. Karabuk Satellite images utilized as cases study for agricultural areas determination. Classification accuracy was 87%. The outcomes displayed that agricultural areas determination in GIS environment is accurately and easily can be implement.

I. INTRODUCTION

Considering agricultural practices, it is very difficult to monitor and determine the agricultural activities in large areas by local methods. At this point, it is possible to determine the spatial distribution of agricultural lands by using satellite images from remote sensing. Remote sensing is the most widely used method to produce field boundaries of agricultural areas. In order to determine the spatial distribution of the product range and then to monitor the growth periods of the products, accurate cell size of the satellite images is highly recommended.

To determine agricultural areas from satellite images, several processing steps should be performed and the properties of the objects to be considered should be extracted. Recently, different methods were recommended in studies to detect or identify agricultural lands. Machine learning techniques among other methods area extremely recommended. Machine learning methods can be utilizing to perform satellite image classification to determine agricultural areas. The overall objective of the classification method is to assembly entities with the similar spectral features in the surface. The aim is to distinct every pixel in the satellite pictures into dissimilar sets based on their spectral features. Later, allocate the pixel to the matching sets on the earth based on their reflectance content. In addition, ArcGIS software comprises both training data manager and relevant tools for classification. ArcGIS permit to

end-users to collection and examination spatial data of assessment features across the desired geographic landscape. So, ArcGIS is appropriate atmosphere to execute satellite image classification machine learning techniques.

The objective of this paper is presenting the ability of using machine learning and satellite images to determine agricultural areas. Classification atmosphere is geographic information systems GIS.

II. METHOD

2.1. Satellite image collection Landsat 8

In 1972, the LANDSAT satellite program was launched by NASA in order to investigate the natural resources on Earth. The first Landsat satellite was sent to space on July 23, 1972 and served for about six years. The last satellite in the Landsat satellite program was launched on February 11, 2013. Landsat 8 has 2 new sensors: OLI (Operational Land Imager) and TIS (thermal infrared sensor).

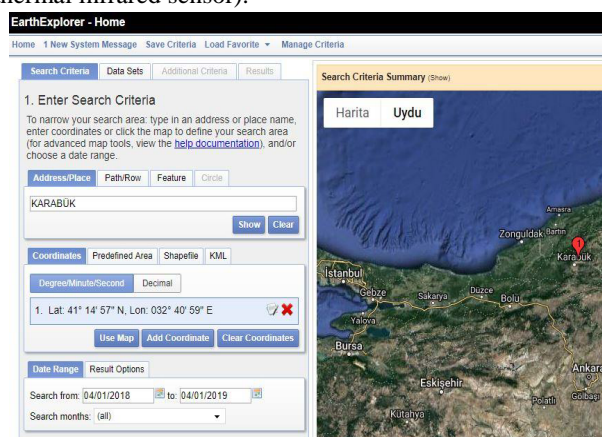


Figure 1 USGS satellite images platform

For Landsat 8 satellite images of Karabük province, 'earthexplorer.usgs.gov' address was used to obtain the image (United States Geological Survey (USGS) earth explorer website) Figure 1. After the location data was entered, a specific

date range was selected and satellite images of Landsat 8 were downloaded Figure 2. Then, image transferred to ArcGIS software.



Figure 2 Satellite image collection

2.2 Agricultural areas classification

A controlled classification algorithm requires a collection of data points for each class. Therefore, classification is based on how the point is close other training samples to be classified. Training samples represent the known interest classes of the analyst. The three basic steps in a typical controlled classification procedure are as follows: (A) Training stage: The analyst defines representative training areas and develops numerical explanations of spectral signatures of each type of agricultural areas at the scene. (B) Classification stage: each pixel in the image data set is classify, to its most similar. If the Pixel does not look like any training data, it is usually labeled 'Unknown'.

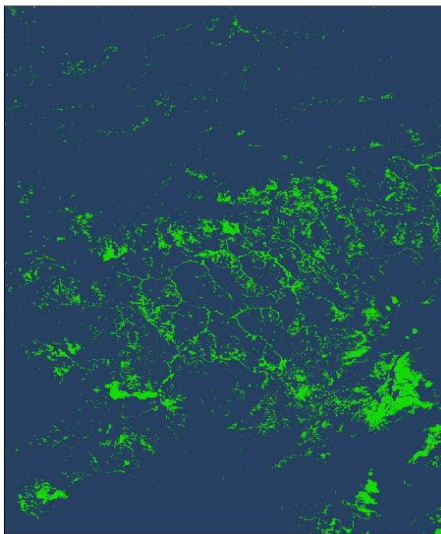


Figure 3 Classified satellite image (Agricultural areas)

ArcGIS, the GIS software, was used in our project. Landsat footage uploaded to ArcGIS. Later, training areas were created to classify the image. Image classification toolbar used. Then, polygons are drawn by selecting the agricultural areas. The polygons saved as signature file for training samples reflectance. Then Maximum Likelihood Classification tool from the classification tab were utilized. The obtained agricultural areas map as a result of these applications is given in Figure 3.

III. CONCLUSION

In this study, satellite images of agricultural areas in Karabük province were obtained from Landsat 8. The appropriate classification tool used to determine the agricultural lands from the available satellite images. Classification of satellite images for agricultural areas maps is performed based on supervised classification. ArcGIS environment utilized to applied the analysis. The identification and classification of agricultural lands is the most important element to achieve these objectives. The outcomes displayed that GIS environment accurately and easily determin agricultural areas.

IV. ACKNOWLEDGMENT

This study has been supported by 2221 – Fellowship Program of TUBITAK (The Scientific and Technological Research Council of Turkey). We are indebted for their supports.

V. REFERENCES

- Uğur ALGANCI-İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü Doktora Tezi Ağustos 2014 Uydu Görüntüleri, Meteorolojik Veriler ve Kamera Fotoğrafları ile Pamuk ve Mısır Bitkileri İçin Rekolte Tahmin Modeli Tasarımı: Şanlıurfa Örneği
- TUFUAB IX. Teknik Sempozyumu 2017 / AFYONKARAHİSAR
- International Journal of Applied Earth Observation and Geo information
- “www.gislounge.com/using-machine-learning-satellite-imagery-estimate-corn-crop-production”
- http://www.ktu.edu.tr/dosyalar/ormanamenajmani_09c24.pdf

TDMA scheduling for real-time flows in cluster-based wireless sensor networks

Gohar Ali¹ and Mohammed Ilyas²

¹ Department of Information Systems and Technology, Sur University College, Sur, Oman
goharali@suc.com

² Department of Information Systems and Technology, Sur University College, Sur, Oman
md_ilyas@suc.edu.om

Abstract - Existing work using Time-division multiple access (TDMA) scheduling in Wireless Sensor Networks (WSNs) try to minimize the schedule length. The problem in previous cluster-based scheduling algorithm is that the node with in the clustered utilized only the intra-cluster time slots and cannot utilized empty time slots of the inter-cluster therefore increase the schedule length. In this paper, we proposed new interference-aware TDMA scheduling algorithm to utilized the empty time slots of inter-cluster to reduce the schedule length. The objective is to improve the delay and throughput by scheduling intra-cluster and inter-cluster non-interference transmissions in the same slot with TDMA scheduling algorithm. Simulation results shows the performance of our proposed scheme.

Keywords – Scheduling, real-time, interference, clustering

I. INTRODUCTION

Wireless Sensor Networks (WSNs) are recently used in many applications. For example, in fire monitoring system the packet generated by nodes must be transmitted in short period to the fire monitoring station to take the action on time. In patient monitoring systems the patient breathing activity, blood glucose, body temperature etc. should reach to the monitoring station timely, to keep doctor aware about patient current situation. Similarly in disaster systems the events detected must report within deadline to base station. Other examples includes, industrial process monitoring, traffic monitoring, structural health monitoring and so on [1], [2].

In WSN, reducing the schedule length is significantly more difficult than wired networks due to the interference among nodes. Because communication between pair of nodes is affected by other nodes that are within the interference range of these nodes and transmit at the same time. Thus, the performance of communication can significantly enhance by considering interference-aware scheme.

In WSNs, MAC designed can be divided into contention based or contention free based schemes. Contention based MAC such as CSMA(Carrier Sense Multiple Access) try to resolve the collision when it occur so it does not provide guarantee on delay. While In time division multiple access-based Media Access Control, a bounded delay can be determine through time slot scheduling. For this purpose Time Division Multiple Access protocol in WSN reduce data retransmission as this protocol avoid the collision by allowing different nodes to access the shared medium without interfering with each other [3-4].

Several TDMA scheduling algorithms are proposed to transmit data from source to destination using multi-hop [5,6,7,8,9,10]. In [5], proposed two centralized, node-based and level-based TDMA scheduling algorithm. In both algorithms base station assign slots to nodes by considering interference relation. These algorithms attempts to find TDMA schedule that minimize the number of time slots. In [6,7,8], conflict-free TDMA scheduling algorithm for multi-hop intra and inter-cluster is provided. The algorithm schedule every node across 3-hop neighbor for the purpose to avoid the interference hence to improve the delay, throughput and energy efficiency. Similarly in some TDMA scheduling the knowledge of 2-hop neighborhoods is assume to avoid the interference. In adaptive DRAND [9] the scheduler assign slot to node across 2-hop neighborhood to avoid the interference while in [10] assign colors to node different from 2-hop neighbor node.

The mentioned scheme assign slots to nodes across 2-hop or 3-hop away to avoid the interference. That consider that if the time slots is not used in 2-hop or 3-hop neighborhood

then that time slots can be reused. However these schemes do not consider the concurrent transmission of nodes in the network. That is due to cumulative nature of interference, where transmissions of other nodes that are n-hop away can interfere with a transmission as they transmit simultaneously. Therefore instead of using 2-hop or 3-hop, we use SNIR (Signal-to-noise plus interference) interference model [11, 12]. In this model a set of transmission is interference-free if the SNIR of all receiver exceeds a threshold.

In this paper, we proposed new interference-aware TDMA scheduling algorithm for reducing TDMA scheduling length in WSNs. So if all the flows are not arrived at the cluster-heads then intraSend can schedule the flows in the empty interComm time slots. Similarly if all the InterComm time slots flows are scheduled then the intraRecv can schedule the flows both in IntraSend and InterComm time slots. The aim is to schedule more flows per time frame and reduce the schedule length by new TDMA scheduling algorithm.

The rest of the paper is organized as follows. Section II, discussed about the proposed work. System performance evaluation are provided in Section III. The paper concludes the section IV.

II. PROPOSED SCHEME

The proposed scheme consists of three scheduling algorithms. The first scheduling algorithm is intraSend scheduling algorithm. This algorithm schedules the flow from source node to source cluster head in the intraSend time slot. The second scheduling algorithm is interComm scheduling algorithm. This algorithm schedules the flow from source cluster head to destination cluster head. The third scheduling algorithm is intraRecv scheduling algorithm. This scheduling algorithm schedules the flow in the IntraRecv time slot from destination cluster head to destination node.

During the initialization phase each node within the cluster sends their interference, arrival to the source cluster head and arrival to the destination cluster head to the global cluster-

head. Each cluster-head shares their interference information with the neighbor cluster heads. As shown in the above Figure 1 the schedule starts from the IntraSend Scheduling. If the intraSend flow is ready for schedule then in the IntraSend time slots intraSend flow is scheduled. Next in InterComm time slots if the intercomm flows are not ready then intraSend flow

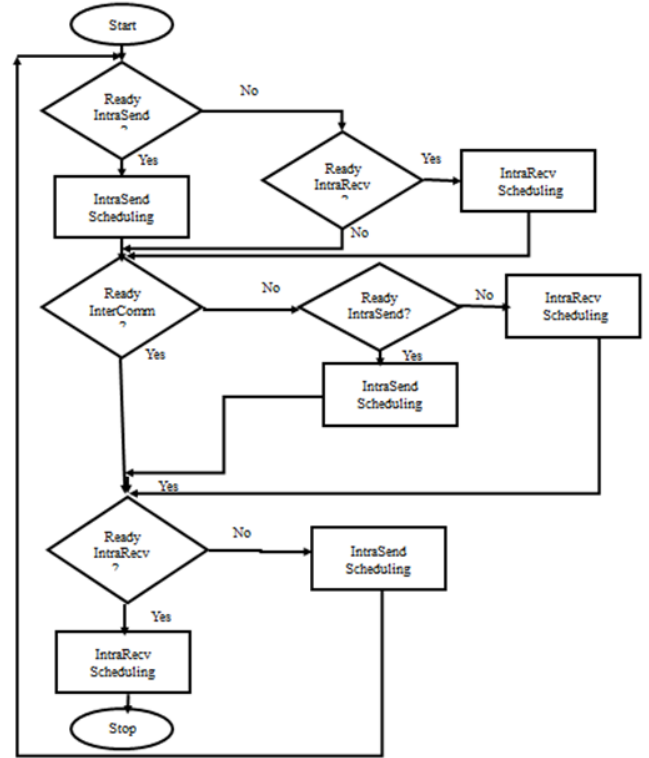


Figure. 1 Flow chart for proposed scheme

will be scheduled otherwise the intraRecv flow will be scheduled. We explain the Figure 1 with help of an example as shown in Figure 2 five flows from F1 to F5. F1 has high priority while F5 has the lowest priority.

As shown in the Figure 3 the TDMA frame is divided into three parts. The first time slots is for intraSend scheduling, the second and third time slots for InterComm Scheduling. Finally the last time slot for IntraRecv scheduling. As shown in Figure 4(a) the scheduling is starting from the IntraSend scheduling. The F1 has high priority therefore transmission *za* will be scheduled in the IntraSend time slot 1. Transmission *sg* in F2 and transmission *za* in F1 has no interference so it will be scheduled in the same IntraSend time slot 1. Transmission *te* in F4 and *re* in F3 has interference therefore transmission *te* in

F4 will be delayed due to low priority. Transmission xk in F5 will be schedule in time slot 1 due to no interference with high priority flows.

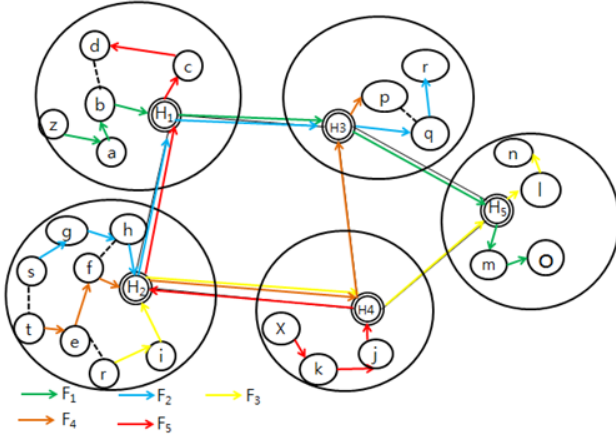


Figure. 2 Example of cluster-based real-time flow

Flow#	IntraSend		InterComm		IntraRecv
	ri	1	2	3	
F1	0	$z \rightarrow a$	$a \rightarrow b$		$b \rightarrow H1$
F2	0	$s \rightarrow g$	$g \rightarrow h$		$h \rightarrow H2$
F3	0	$r \rightarrow i$	$i \rightarrow H2$	$H2 \rightarrow H4$	
F4	0		$t \rightarrow e$		
F5	0	$x \rightarrow k$			$j \rightarrow H4$

(a)

Flow#	IntraSend		InterComm		IntraRecv
	5	6	7	8	
F1		$H1 \rightarrow H3$	$H3 \rightarrow H5$	$H5 \rightarrow m$	
F2	$s \rightarrow g$		$H2 \rightarrow H1$		
F3					
F4	$e \rightarrow f$			$f \rightarrow H2$	
F5		$H4 \rightarrow H2$			

(b)

Flow#	IntraSend		InterComm		IntraRecv
	9	10	11	12	
F1	$m \rightarrow o$				
F2		$H1 \rightarrow H3$		$H3 \rightarrow q$	
F3			$H4 \rightarrow H5$	$H5 \rightarrow l$	
F4		$H2 \rightarrow H4$			
F5			$H2 \rightarrow H1$	$H1 \rightarrow c$	

(c)

Flow#	IntraSend		InterComm		IntraRecv
	13	14	15	16	
F1					
F2	$q \rightarrow r$				
F3	$i \rightarrow n$				
F4		$H4 \rightarrow H3$	$H3 \rightarrow p$		
F5	$c \rightarrow d$				

(d)

Figure.4 Example of proposed scheduling scheme

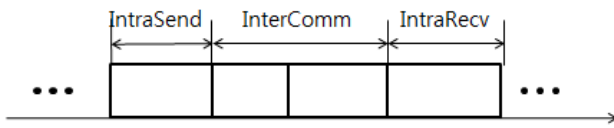


Figure. 3 TDMA frame

Interestingly in the InterComm Scheduling there is no ready flow to be schedule in time slot 2. Therefore the IntraSend will be schedule flows in the time slot 2 as shown in Figure 4(a). Similarly F2,F3 and F4 will also schedule their next IntraSend flow in the InterComm slot 2 and will reduce the schedule length as compared to previous scheme. InterComm transmission $H2H4$ of F3 will be schedule in the slot 3 therefore IntraSend and IntraRecv cannot be schedule in this time slot. Finally the transmission $bH1$ of F1 will be schedule in the timeslot 4 as there is no IntraRecv flow is read to be schedule. Similarly IntraSend transmission $hH2$ of F2 will be also schedule in the IntraRecv time slot since there is no ready IntraRecv flow. Similarly in figure 4(b), 4(c) and 4(d) the flows are schedule from source node to their destination.

III. PERFORMANCE EVALUATION

We use GENSEN[13] tool to perform the simulation of algorithm. The nodes are placed randomly in the area of 100m x 100m area.

Our scheme compared the acceptance rate flows with previous scheme [14]. The previous scheme schedule the intra-cluster flows in intra-cluster time slots while the inter-com flows schedule the flows in interComm time slots. For the same number of cluster and number flows 5 Figure 5 shows comparison of acceptance rate of the previous and proposed scheme. The proposed scheme schedule more flow per time frame than the previous scheme. This is because in previous scheduling the intraSend schedule the flow only in intraSend time slots while intraRecv schedule the flow in the intraRecv time slots even interComm slots are empty. While in proposed scheduling scheme, the intraSend and intraRecv can utilized the interComm time slots if the interComm flows are not schedule therefore increase the acceptance rate. In Figure 6 the number of flows are increased by 8. In proposed scheme due to the inference of the increasing number of flows most of the flows will be not reached to their cluster-heads in the IntraSend Scheduling so they will utilized the InterComm time slots. So the acceptance rate of the proposed scheme is higher than the previous scheme.

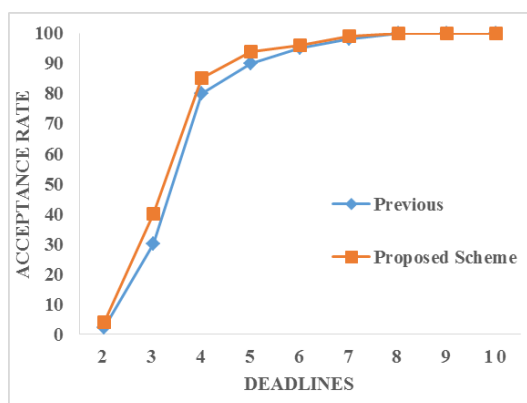


Figure 5 Number of flows=5

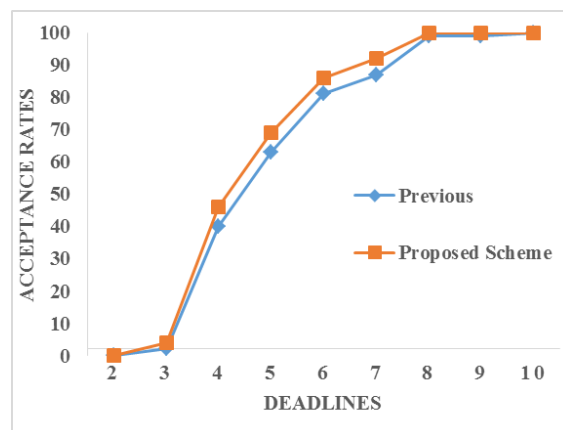


Figure 6 Number of flows=8

IV. CONCLUSION

In this paper a new interference-aware real-time flow scheduling algorithm is proposed. The proposed scheduling algorithm schedule the flows from source to destination in cluster-based approached. The IntraSend and InterRecv scheduling can utilized the empty time slots of InterComm Scheduling.

ACKNOWLEDGMENT

The authors would like to thank Sur University College (Sur, Sultanate of Oman) for their support and sponsorship of this research.

REFERENCES

- [1] Cao, X., Chen, J., Xiao, Y., Sun, Y.: Building-environment control with wireless sensor and actuator networks: Centralized versus distributed. *IEEE Transactions on Industrial Electronics*, 57(11), 3596-3605 (2010).
- [2] O. Diallo, J. C. Rodrigues, M. Sene. Real-time data management on wireless sensor networks: A survey. *Journal of Network and Computer Applications*, 35(3):1013-1021, May 2012.
- [3] Y. Li, C. S. Chen, Y. Q. Song, and Z. Wang. Real-Time QoS Support in Wireless Sensor Networks: A Survey. In *Proceedings of the FeT*, pages 101-108, 2007.
- [4] P. Suriyachai, J. Brown, and U. Roedig. Time-critical data delivery in wireless sensor networks. In *Proc. of the 6th IEEE International Conference on Distributed Computing in Sensor Systems*, Santa Barbara, USA, pp. 216-229, 2010.
- [5] H. Zhang, P. Soldati, and M. Johansson. Optimal link scheduling and channel assignment for convergecast in linear WirelessHART networks. In *Proc. of WiOPT*, 2009.
- [6] T. Herman and S. Tixeuil, A distributed TDMA slot assignment algorithm for wireless sensor (2004).
- [7] Pawar, P.M.; Nielsen, R.H.; Prasad, N.R.; Ohmori, S.; Prasad, R., "GCF: GreenConict Free TDMA scheduling for wireless sensor network," *Communications (ICC)*, 2012 IEEE International Conference on, vol., no., pp.5726, 5730, 10-15 June 2012
- [8] Pawar, P.M.; Nielsen, R.H.; Prasad, N.R.; Ohmori, S.; Prasad, R., "M-GCF: Multicolor-Green Conict Free scheduling algorithm for WSN," *Wireless Personal Multimedia Communications (WPMC)*, 2012 15th International Symposium on ,vol., no., pp.143,147, 24-27 Sept. 2012.

- [9] DRAND: Distributed Randomized TDMA Scheduling for Wireless Ad Hoc Net-works
- [10] Sinem Coleri Ergen and Pravin Varaiya. 2010. TDMA scheduling algorithms for wireless sensor networks. *Wirel. Netw.* 16, 4 (May 2010), 985-97.
- [11] L. Jin, L. Fang and H. Zailu, "Initialization Algorithm Based on SNIR Model for Wireless Sensor Networks," 2006 International Conference on Wireless Communications, Networking and Mobile Computing, Wuhan, 2006, pp. 1-4.
- [12] F. Durand and T. Abrão, "Distributed SNIR Optimization Based on the Verhulst Model in Optical Code Path Routed Networks With Physical Constraints," *J. Opt. Commun. Netw.* 3, 683-691 (2011).
- [13] T. Camilo, J. Sa Silva, A. Rodrigues, and F. Boavida. GENSEN: A topology generator for real wireless sensor networks deployment. In *Proc. of 5th IFIP Workshop on Software Technologies for Future Embedded & Ubiquitous Systems (SEUS 2007)*, Santorini, Greece, 2007.
- [14] G. Ali, S. Kang, K. H. Kim and K. Kim, "Towards Cluster-Based Real-Time Flow Scheduling in Interference-Aware Wireless Sensor Networks," 2013 IEEE 16th International Conference on Computational Science and Engineering, Sydney, NSW, 2013, pp. 523-530.

Deep Learning Approaches for Traffic Flow Predictions on Signal-controlled intersections

Fatih Gunes¹, Dr. Selim Bayraklı² and Prof. Dr. Abdul Halim Zaim³

¹ Istanbul Commerce University, Istanbul/Turkey, fatih.gunes@istanbulticaret.edu.tr

² Maltepe University, Istanbul/Turkey, selimbayrakli@maltepe.edu.tr

³ Istanbul Commerce University, Istanbul/Turkey, azaim@ticaret.edu.tr

Abstract – In today's world, the number of vehicles is increasing rapidly in developing countries and in city centers. Developing technology, economy and increasing population cause to use more vehicles in today's cities. More vehicle use causes long queues in the junction arms, unnecessary time losses, fuel consumption and environmental pollution. As known, the main reason of the congestion in city traffic is the intersections. Signalized intersections, which are the points where many roads meet, inevitably cause the traffic to slow down or even stop. The control of the traffic is provided by signaling control systems. With the control system of an intersection, it is aimed to minimize delays and downtime and to maximize the benefits provided to all road users. The basic control mechanism used for this purpose in the signalized interconnections depends on the optimization of the signal timing. Until now several researches have been done about signal timing optimization. In this context, the data, collected in and around the junction arms, represent the most important parameter. In this study, we studied deep learning approaches in traffic signal timing and estimation.

Keywords - Deep Learning, Machine Learning, Traffic Flow Prediction, Signalized intersections, Smart Cities

I. INTRODUCTION

Nowadays, one of the main factors that affecting the human life negatively in city centers, is traffic problem. Irregular and distorted urbanization and the increase in vehicle ownership have become the most important factors behind the transportation problems. The fixed and sequential working times of the lights at the intersections cannot cope with variable traffic density and cause unnecessary long queues at the intersections. Average vehicle delays in signalized intersections directly affect the entire traffic system. The increase in delay times plays a major role in the choice of drivers and accordingly the distribution of the traffic on the road network.

The best optimization of signal times at signalized intersections is based on a realistic estimation of vehicle delays and queue lengths at the junction. This enables full utilization of the intersection capacity and minimizes the total delay. The estimation of vehicle delay is very important especially in signal planning and traffic management. In line with this

estimation, intersection control has a more flexible control mechanism and can be managed more efficiently [1].

The main parameter used in the calculation of signaling times is traffic volume values. In order to determine the traffic volume values, counts are performed by sensors and cameras placed at the intersections. In today's technology, it is possible to design intelligent signaling systems with the use of low-cost cameras, processors with high computations and vehicle counting modules. Especially in recent years, rapidly developing image processing techniques and successful applications also strengthen this idea. The most important parameter considered in the improvements in signalized intersections is the delay times of the vehicles. In order to minimize the average vehicle latency and average stop counts in a road network, the plan times of the signalized junctions are optimized according to the parameters such as traffic volume and queuing. If these techniques are used in signaling systems, it can be ensured that the traffic lights operate in variable time depending on the vehicle density at the intersections.

Systems that operate in this way can be monitored instantly from traffic control centers. It is aimed to produce the best predictive results by evaluating the data obtained from intersections. In the next step, the signaling sequence and times can be changed momentarily by comparing the available data with current and past situations in other intersections.

Nowadays, it is easier to calculate traffic flow or density due to developing technology and more easily collected GPS data from field. Real-time traffic data collected from a variety of sensor sources such as loop detectors, radars, cameras, GPS, social media has opened the way to new developments in traffic and transport technologies. We are in a period where big data analytics play an important role in determining the density maps, routes and flow in traffic.[2]

For this reason, this study investigated of the prediction of signal timing, vehicle delay and density in the aspect of machine learning and deep learning techniques. In order to develop the most appropriate predictive model for such problems, information extraction methods are used more frequently. In this context, deep learning is very closely related to such problems. Deep learning algorithms have been applied to many different areas especially in recent years.

Deep learning algorithms are applied to a wide range of problems, most of them achieving the best achievements in the field. The main aim of the study is to gain a vision for researchers who want to learn about deep learning or apply these methods to problems in traffic and transportation. In this context, another aim of the study is to give an idea about which problems can be applied to deep learning approaches.

In the first part of this study, developing cities and approaches to increasing traffic problems are mentioned. In the second chapter, the researches about deep learning in literature and applications in traffic analysis are mentioned. The following section summarizes the current information about deep learning algorithms and especially the hardware, software and libraries used in the field of deep learning. As a result, a comprehensive analysis has been done by researching advanced solution methods in the field of traffic analysis, which is a current issue that is open to research and development.

II. LITERATUR REVIEW

Although deep learning was very popular under the science of artificial intelligence and machine learning since the 1980s, it has not found much practice in these years due to hardware deficiencies. These new technologies, called Deep Learning, were inspired by the neural networks in the human brain. We can imitate intelligent organisms in nature by means of these technologies, such as human, which sometimes perceive, see, hear, speak and learn better than humans. Intelligent systems can detect and make sense of the environment as a living thing. Beyond that, we can develop machines that can see, hear, perceive and recognize that people can't. In summary, deep learning, which is the subject of this study is one of the artificial intelligence approaches. And it is a type of machine learning, especially where computers develop themselves using data and experiences. It has been suggested that this kind of machine learning can be an artificial intelligence that can solve real world problems.

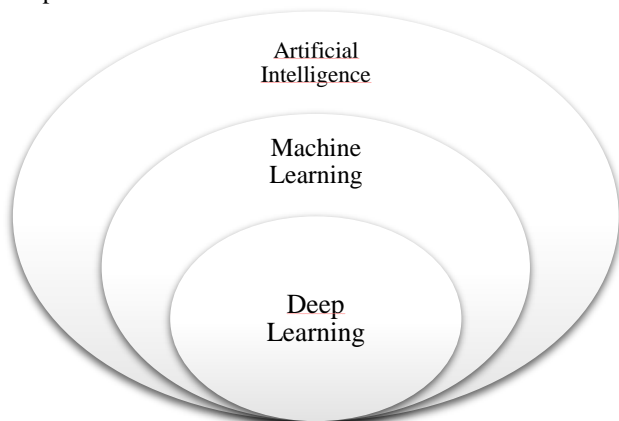


Figure 1 shows that deep learning is a type of machine learning used in many approaches to artificial intelligence.

It has applications in many fields such as computer vision, natural language processing, autonomous vehicles, music, art, defense industry, security and finance. However, high-speed graphics processing units (GPUs), high-capacity memory and /

or cloud computing environments are needed when working in the deep learning area.

In recent years, many models have been developed to improve the efficiency of transport, such as route management, vehicle routing, signal optimization, traffic management, traffic flow and density prediction. We can classify the predictive approaches in three categories such as parametric, nonparametric methods and simulation tools. Parametric models include methods such as linear regression, time series models, Kalman filtering models. Nonparametric models include methods such as the nearest neighbor (k-NN) methods, artificial neural networks (ANN), decision trees [2]. In such models, the structure is determined only by the data. Simulation approaches use traffic simulation tools to estimate traffic flow such as Vissim, Sumo, Sidra etc. Artificial neural networks, fuzzy logic methods and decision support systems were frequently used in the improvement of traffic and transportation problems [1].

Tzes and colleagues have developed a simulation model based on fuzzy logic and have compared it with fixed-time signal model.[3] Hoyer and Jumar developed a model that compares the volume of traffic in the intersection and the duration of the phase times depending on the red-light signal duration [4]. Huang et al. designed control of signal systems with genetic algorithms.[5] Anokye and his colleagues measured the performance on the links of the intersection over real data using the queuing model [6]. Canaud and his friends studied probability hypothesis density (PHD) methodology for real-time traffic state estimation. They suppose that this model brings novel tool to the state estimation problems. And in this article, they compared PHD filter performance with a particle filter (PF), both considering the measurement origin uncertainty and show that they can provide high accuracy in a traffic setting and real-time computational costs [7]. The supervised learning algorithm, which has a multilayered structure for use in artificial neural networks that are inspired by the human nervous system, was first published by Ivakhnenko and Lapa in 1965 [8]. In the late 1970s, was used Box-Jenkins techniques for analysis of freeway traffic time-series data [9]. Time series models, which aim to make predictions for the future with the help of observation values related to previous periods, are widely used in many fields such as medicine, engineering, business, economy and finance. There are different models which are created by using different methods to make estimation with time series. Among these models, ARIMA models are the most widely known and widely used. ARIMA models, which assume a linear relationship between serial data and can model this relationship, can be applied successfully to time series with various statistical methods [10]. In the case of traffic flow estimation, models based on time series methods are quite common. Levin and Tsao found that Box - Jenkins time series analysis was used to estimate highway traffic flow and ARIMA (0, 1, 1) was the most significant statistical method for all estimates [11]. Due to the stochastic and non-linear nature of the traffic flow, researchers are more likely to apply non-parametric methods to predict traffic flow. Zhang and his friends presented KNN model for short-term traffic flow prediction [12]. Sun et al. studied a local linear regression model for short-term traffic forecasting [13]. Pang et al. used

non parametric Regression algorithm for forecasting of traffic flow [14]. Kimber and Hollis presented techniques for forecasting queue length and delay at unsignalized intersection [15]. Madanat et al. introduced a probabilistic delay model at stop-controlled intersection [16]. Brilon has aimed to calculate average delay at unsignalized intersections for periods with variable traffic demand [17]. Wu presented a new model for estimating critical gap and its distribution at unsignalized intersections based on the equilibrium of probabilities [18]. Sabrina et al used K-NN algorithm in road traffic prediction [19]. Intelligent traffic planning is an important factor in establishing smart cities. It is an important challenge to overcome non-linear and random properties in the transport network. In a study[25], a deep learning model called DeepSense was used to estimate the traffic flow by preparing the taxi with GPS data. 5% improvement has been achieved by the present methods with the proposed model.

III. DEEP LEARNING APPROACHES

Numerous techniques have been reviewed and applied in the literature. But here we want to talk about a few methods that are most emphasized, especially in predictive approaches.

A. K-NN (K nearest neighbor) Algorithm

K-NN is the simplest, non-parametric and most widely used supervised learning algorithm among all machine learning algorithms. In the training phase of the algorithm, class-specific data are used. Unlike other supervised Learning algorithms, it does not have a training phase. Training and testing are almost the same thing. It is a kind of lazy learning. Therefore, K-NN is not an ideal candidate for the algorithm required to process a large data set.

With K-NN you are basically looking for points that are closest to the new spot. K represents the amount of the nearest neighbors of the unknown point. To estimate the results, we choose the k amount (usually an odd number) of the algorithm. K neighbors are a simple algorithm that stores all existing cases and classifies new cases based on a similarity measure (eg distance functions). KNN was used in statistical estimation and pattern recognition as a nonparametric technique in the early 1970s.

The KNN algorithm, which is one of the most frequently used methods in estimated approaches, was also used in traffic problems. This algorithm identifies the nearest neighbors within the data set and groups them for clustering. Data set contains some values such as density of the flow, count of the car, queue length at the signalized intersections [19]. Training of these data is so critical to estimate next step of the traffic flow. It is also likely to be used in several problems such as vehicle classification in city centers, estimating queue length at signalized intersections, prediction of traffic flow and intensive time zones [29].

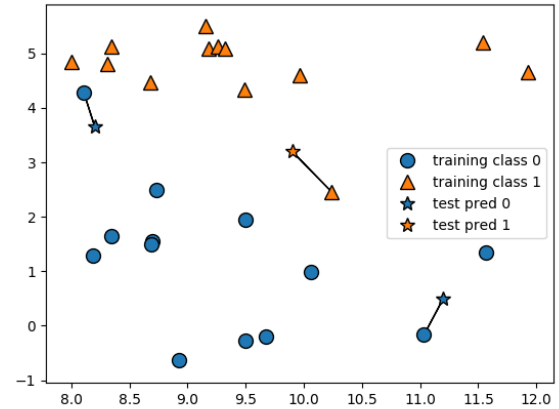


Figure 2 KNN aim to minimize distance between test and training observation for high classification accuracy [20]

B. Autoregressive Integrated Moving Average Model (ARIMA)

As mentioned above, time series models, which aim to make predictions for the future with the help of observation values of past periods, are widely used in many areas. There are different models which are created by using different methods to make estimation with time series. Among these models, ARIMA models are the most widely known and used.

It is very important to be able to predict the future behavior of series in traffic problems or applications in different areas. ARIMA “autoregressive integrated moving average” is a common way to model stationary time series. This approach, developed by George Box and Gwilym Jenkins, is also called the Box-Jenkins (BJ) method. The main emphasis of the Box-Jenkins method is to explain the time series only with their historical values and the probabilistic error term [9], [2].

The general representation of the models is ARIMA (p, d, q). Here, p and q are respectively the degree of autoregressive (AR) model and moving average (MA) model, and d is the degree of difference.

The general model ARIMA (p, d, q) is formulated as follows.

- p is the number of autoregressive terms,
- d is the number of nonseasonal differences needed for stationarity, and
- q is the number of lagged forecast errors in the prediction equation.

$$Y_t = \theta + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + U_t + \beta_1 U_{t-1} + \dots + \beta_q U_{t-q}$$

$$Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$$

denotes the observation values that are different from d.

$$\alpha_1, \alpha_2, \dots, \alpha_p$$

are the coefficients for the observation values,

$$U_t, U_{t-1}, \dots, U_{t-q}$$

describe error terms

$$\beta_1, \beta_2, \dots, \beta_q$$

are coefficients of the error terms.

Once the values of p, d and q are determined, the second stage in the Box-Jenkins method is estimating of the model. After a certain ARIMA model has been estimated, the next step is to examine how data is close to the model.

C. Artificial Neural Networks (ANN)

Mathematical modeling of the nervous system and decision mechanisms of living things is the main subject of artificial neural networks. It is aimed to model the learning structure of live brain by using artificial neural networks which can be trained as adaptable and self-organized.

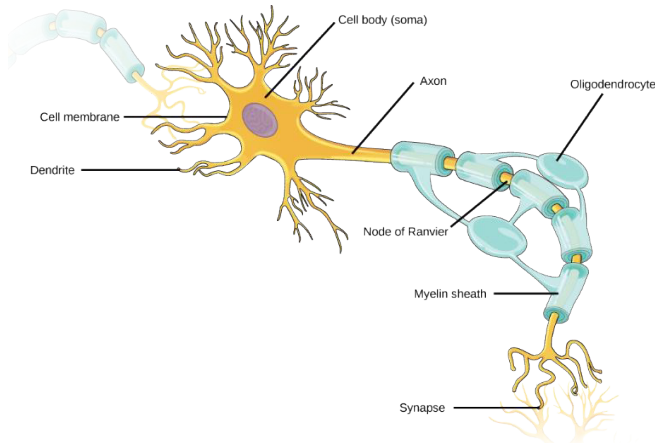


Figure 3 Human neural cell structure [23]

Artificial neural networks (ANN), which can model both linear and nonlinear relationships, have been one of the most used deep learning methods in the analysis of the prediction models in recent years. The most important advantage of artificial neural networks is that when the functional structure of the data set cannot be determined precisely, it can successfully extract many different analytical models based on the available data [22]. With artificial neural networks, it is possible to model very successful forecasts by training the existing data without any assumptions on any model. Artificial neural networks are used in many areas such as control and system identification, image and voice recognition, prediction problems, fault analysis, predictive maintenance, medicine, communication, traffic and production planning.

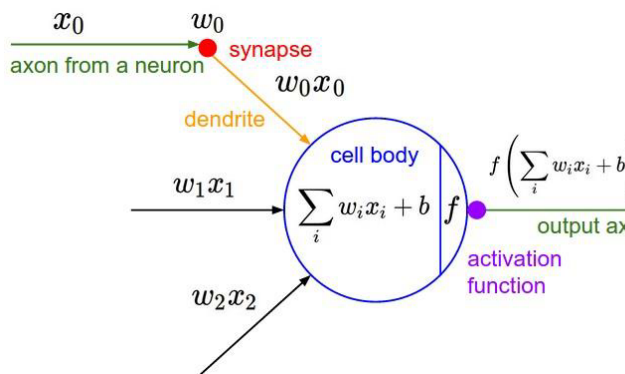


Figure 4 Mathematical model of a neuron (nerve cell) [23]

To summarize the above model, artificial neural network is determined by a specific mathematical model to determine

the weights and produce the outputs. According to this model we can assume that we have the data set as an input layer values (x_0, x_1, x_2) and after receiving the input data, the artificial neural network will produce an output according to the initial weight values (w_0, w_1, w_2) . Processes will be repeated until appropriate weight values are obtained according to the accuracy of the outputs [21].

The mathematical model of the neural network is the process by which each input is multiplied by its own weight before the incoming information goes to the activation function, where different aggregate functions can be used instead of sum operation. Activation function determines the appropriate output by processing the values from the total function. Activation functions can vary depending on the problem. For example, if we take the threshold function, the output value can produce 1, 0 output depending on whether it is smaller than or equal to the appropriate threshold value.

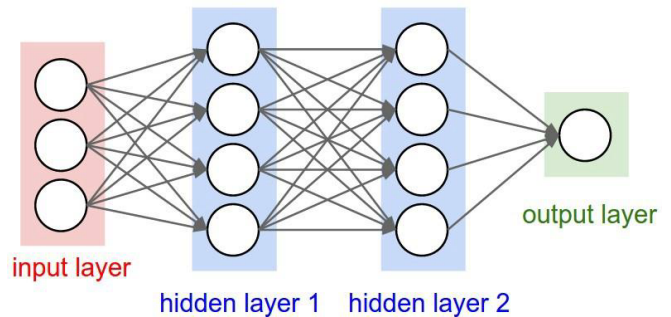


Figure 5 Single and multi-hidden layer artificial neural networks [23]

D. Multilayer Perceptron Neural Networks

In artificial neural networks, there are a wide variety of models of neuron cells called neurons connected to each other in advanced and feedback form. Multilayer Perceptron Neural Networks are the most commonly used among these models. There are many different frameworks and libraries developed for deep learning. The following table shows the most popular technologies and languages. Neurons in MLP networks are organized as layers. The first layer in MLP is the input layer. The input layer makes the data available to be entered the neural network. The other layer is the output layer in which the information processed within the network is transmitted to the outside. The layer between the input and output layers is called the hidden layer. MLP networks can also have multiple hidden layers. Figure 6 shows the structure of a typical MLP network.

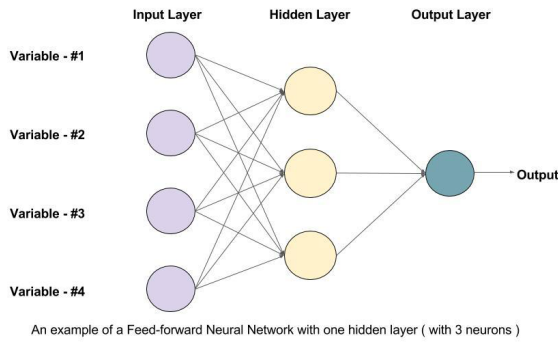


Figure 6 MLP Network with one hidden layer

The main task of a model used in artificial neural networks is to learn the structure in the sample data set and to generalize to fulfill the desired task. For the network to perform this task, a model can be generalized by training the related event with the sample data sets.

Artificial neural network with the output value y_t and the observation inputs x_i can be expressed mathematically as follows;

$$y_t = w_0 + \sum_{j=1}^p w_j f(\sum_i w_{ij} x_i) + b \quad (1)$$

In equation(1); p represents the number of hidden neurons and f shows the nonlinear activation function used in the hidden layer. The most commonly used activation functions are sigmoid(2) and hyperbolic tangent(3).

$$f(x) = \frac{1}{1+e^{-x}} \quad (2)$$

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3)$$

E. Convolutional Neural Networks

Convolutional neural networks (CNN), a special type of multi-layer perceptron-MLP neuron networks, which is modeled by the modeling of the human visual system, is now widely used for achieving the most successful results in the field of computer vision.

One of the most frequently used problems of CNN is image processing. The CNN can be technically explained as follows; the cells in the center of vision are subdivided to cover the entire image. Simple cells concentrate on edge-like properties, while complex cells concentrate on the entire image with larger receptors. CNN, an advanced neural network, was inspired by the vision center of animals. The mathematical convolution process here can be thought of as a response from a neuron to its stimuli. Convolutional neural networks are one of the most commonly used methods for object recognition, classification, tracking, natural language processing, meaning separation, sentence modeling and forecasting problems [26].

Considering the input layer in artificial neural networks, the input data used in CNN is an image. In the field of image processing an image can be described as a matrix of RGB values. Convolution occurs on the image matrix by moving a matrix of size 3x3, 5x5, 7x7, 9x9, 11x11. Specified small size

filter matrices move over the entire image matrix to highlight the attributes in the image. In this way a new feature is obtained from the image matrix and with different filter mask can be obtained different features.

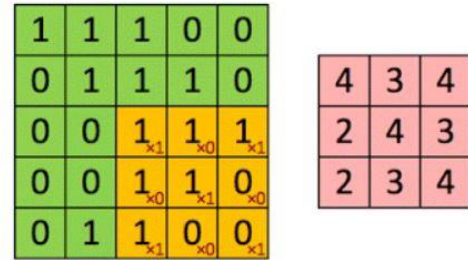


Figure 7 Image and convolved feature [27]

Although machine learning is a method that has been used for many years, two new trends have led to the widespread use of machine learning: a powerful and efficient parallel computation with GPU and large data to be trained from many sources. In recent years the tech-companies as well as the open source communities have been working in the field of deep learning and continue to develop several tools for data science world.

Library	Language	Developer
Theano	Phyton	MILA Lab
Caffe	Phyton	Berkeley Vision and Learning Center
Torch	Lua	Ronan Collobert et al.
Digits	C++	NVIDIA
Tensorflow	Phyton	Google
DeepLearning	Java	Adam Gibson
Keras	Phyton	Francois Chollet-Google
MxNet	R, Python ve Julia	Pedro Domingos, Amazon AWS
CNTK	C++	Microsoft

Table 1 Deep Learning libraries [24], [28]

IV. CONCLUSION

In this study, we studied about the approaches to the traffic problem, which has become an important problem in recent years. As described in the previous sections, deep learning algorithms are used in many areas and problems. With the increasing amount of vehicle, in city centers there are problems such as traffic density estimation, vehicle identification, prediction of queue length and optimization of traffic signal timing. Deep learning methods, which prove themselves as a method that can cope with traffic problems, are frequently used and give the most successful results. Besides, GPU-based deep learning algorithms are used for real-time data processing capability in autonomous vehicle

technologies. In our study, we tried to summarize the most commonly used methods in the literature. We are also trying to achieve successful results with the estimation methods that we used in our research within the scope of the doctoral thesis. With the analytical model we developed at the signalized intersections, we can optimize the signal times according to the vehicle density. By doing this study, we aimed to examine the methods we will use to estimate queue lengths and densities at signalized intersections. We want to focus on this issue in the next study.

REFERENCES

- [1] Y. Zayi Murat, M. Başkan, "İzole Sinyalizasyon Kavşaklarındaki Ortalama Taşıt Gecikmelerinin Yapay Sinir Ağları ile Modellenmesi", January 2006
- [2] Yisheng Lv, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang, "Traffic Flow Prediction With Big Data: A Deep Learning App", IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, APRIL 2015
- [3] Tzes A., McShane and Kim, S., "Expert Fuzzy Logic Traffic Signal Control for Transportation Networks", Institute of Transportation Engineers 65th Annual Meeting, Denver USA, 154-158, (1995).
- [4] Hoyer, R., Jumar, U., "Fuzzy Control Traffic Control of Traffic Lights", Proc. IEEE International Conference on Fuzzy Systems, 1526-1531, (1994).
- [5] Zhen-Jin Huang, Chun-Gui Li, Zeng-Fang Zhang, "Traffic signal control based on genetic neural network algorithm", 2009 March
- [6] Martin Anokye, A.R. Abdul-Aziz, Kwame Annin, "Application of Queuing Theory to Vehicular Traffic at Signalized Intersection in Kumasi-Ashanti Region, 2013 Ghana
- [7] Canaud, M., Mihaylova, L., Sau, J. et al. (1 more author) (2013) Probabilistic hypothesis density filtering for real-time traffic state estimation and prediction. Networks and Heterogeneous Media (NHM), 8 (3). 825 - 842. ISSN 1556-1801.
- [8] A. G. Ivakhnenko, V. G. Lapa, Cybernetic Predicting Devices, Purdue University School of Electrical Engineering, 1965.
- [9] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using Box-Jenkins techniques," *Transp. Res. Rec.*, no. 722, pp. 1-9, 1979.
- [10] M. vanderVoort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transp. Res. C, Emerging Technol.*, vol. 4, no. 5, pp. 307-318, Oct. 1996.
- [11] M. Levin and Y.-D. Tsao, "On forecasting freeway occupancies and volumes," *Transp. Res. Rec.*, no. 773, pp. 47-49, 1980.
- [12] Zhang L., Liu Q., Yang W., Wei N., Dong D., "An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction", 13th COTA International Conference of Transportation Professionals (CICTP 2013)
- [13] H. Y. Sun, H. X. Liu, H. Xiao, R. R. He, and B. Ran, "Use of local linear regression model for short-term traffic forecasting," *Transp. Res. Rec.*, no. 1836, pp. 143-150, 2003, Initiatives in Information Technology and Geospatial Science for Transportation: Planning and Administration.
- [14] Pang X., Wang C., Huang G., "A Short-Term Traffic Flow Forecasting Method Based on a Three-Layer K-Nearest Neighbor Non-Parametric Regression Algorithm," *Journal of Transportation Technologies* 06(04):200-206 DOI: 10.4236, 2016
- [15] R. Kimber and E.M. Hollis, "Traffic queues and delays at road junctions" 0266-7045, 1979
- [16] S. M. Madanat, et al., "Probabilistic delay model at stop controlled intersection," *Journal of transportation engineering*, vol. 120, pp. 21-36, 1994.
- [17] W. Brilon, "Average Delay at Unsignalized Intersections for Periods with Variable Traffic Demand," *Transportation Research Record: Journal of the Transportation Research Board*, pp. 57-65, 2015.
- [18] N. Wu, "A new model for estimating critical gap and its distribution at unsignalized intersections based on the equilibrium of probabilities," in *Proceeding of the 5th International Symposium on Highway Capacity and Quality of Service*, 2006.
- [19] Sabrina, Chhabra A., "Road Traffic Prediction Using KNN and Optimized Multilayer Perceptron", in *International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 11 (2018) pp. 9843-9847*
- [20] Internet: Grogan M., MGCodesandStats | K-Nearest Neighbors (KNN): Solving Classification Problems, <http://www.michaeljgrogan.com/k-nearest-neighbors-classification-problems/>, 2019.
- [21] Fausett, L., "Fundamentals of Neural Networks: Architectures, Algorithms and Applications", Prentice Hall, 1994
- [22] Donaldson, R.G. ve Kamstra, M. (1996), Forecasting combining with neural networks", *Journal of Forecasting*, 15, ss.49-61.
- [23] Internet: A. Karpathy, Stanford University, Stanford CS class CS231n: Convolutional Neural Networks for Visual Recognition, Course Notes, 20.03.2018.
- [24] "Comparison of deep learning software." [Online]. Available: https://en.wikipedia.org/wiki/Comparison_of_deep_learning_software. [Accessed: 01-May-2017].
- [25] X. Niu, Y. Zhu, and X. Zhang, "DeepSense: A novel learning mechanism for traffic prediction with taxi GPS traces," in 2014 IEEE Global Communications Conference, 2014, pp. 2745-2750.
- [26] Y. LeCun, K. Kavukcuoglu, and C. Farabet. Convolutional networks and applications in vision. In *Circuits and Systems (ISCAS)*, Proceedings of 2010 IEEE International Symposium on, pages 253-256. IEEE, 2010.
- [27] Internet: Computer Science Department, Stanford University | Feature extraction using convolution, http://ufldl.stanford.edu/wiki/index.php/UFLDL_Tutorial, 2019
- [28] Ayyüce Kızrak M., Bolat B., "Derin Öğrenme ile Kalabalık Analizi Üzerine Detaylı Bir Araştırma", BİLİŞİM TEKNOLOJİLERİ DERGİSİ, CİLT: 11, SAYI: 3, TEMMUZ 2018
- [29] Filmon G., Cetin M., "Short-Term Traffic Flow Rate forecasting based on identifying similar traffic patterns", *Transportation Research Part C: Emerging Technologies*, Volume 66, May 2016, Pages 61-78

Securing TLS from MITM Incursion using Diffie-Hellman

Y. JAVED^{1, 2}, S.K. Adnan¹, H.AHTESHAM², S.ALMUQHIM³, Z.I.KHAN² and J.ABDULLAH¹

¹Network Security Research Group, Faculty of Computer Science and Information Technology,
Universiti Malaysia Sarawak.Malaysia, yjaved@psu.edu.sa

¹Network Security Research Group, Faculty of Computer Science and Information Technology,
Universiti Malaysia Sarawak.Malaysia, skadnan@unimas.my

²Prince Sultan University, Riyadh, KSA, hahtesham@psu.edu.sa

³College of community, Shaqra University , Shaqra, KSA, salmuqhim@su.edu.sa

²Prince Sultan University, Riyadh, KSA, zkhan@psu.edu.sa

¹Network Security Research Group, Faculty of Computer Science and Information Technology,
Universiti Malaysia Sarawak.Malaysia, ajohari@unimas.my

Abstract - Computer security is to protect our computer from the unauthorized usage. Prevention measures help to stop unapproved users (hackers) from getting to any piece of the computer framework. Identification decides if somebody endeavoured to break into the framework. When the mugger intercepts the traffic with MiTM incursion in such a way that the client downgrades the connection and hence TLS gets exposed. We suggested using Diffie-Hellman, which aimed to enforce security level by implementing in TLS network environment. The mathematical algorithm that is used Mathematical Algorithm of Diffie-Hellman. The solution that we used to secure the network is providing a digital signature. Therefore, it can help users to make sure they got the message from a trusted source. This solution is very useful since it can prevent from the attackers. The communication between the users is successful without any interruption. For future works, Elliptic curve cryptography (ECC) is suitable as it can overtake due to increasing in smaller device uses and security needs that become increasing.

Keywords - TLS, MITM Incursion, Third Handshake attack, Diffie-Hellman and RSA.

I. INTRODUCTION

Security in the network is very significant. The major requirements of the network security are secrets, rectitude and verification. Nowadays, there are numerous attacks to web servers and to prevent this attacking security measures should be implemented in the networks as well as at the hardware and software level. The attackers usually aim the software which builds the web server so that they can make an illegal authorized entry to the server.

There are number of security algorithms that have been proposed to take care of security issues. The issues may vary from small confidentiality or eaves dropping attack to completely overcoming the admin privileges and having full access rights. The algorithms for security must handle all kind of issues or at least the major issues. One of the biggest threat or attack is Man in the Middle attack or simply MiTM. This

kind of attack is harmful in case of success or failure such as in form of success it will result in complete control and in case of failure will result in Denial of service attack.

Number of researcher have been conducted in this area to avoid these kind of attacks and most successfully it was solved but due to arrival of new devices the security algorithms needs to be revamped to be executed on low computational devices. This research looks into one of the promising algorithm Diffie Hellman and Elliptic curve Cryptography.

Section II looks into common literature that has been studied and conducted in the area. Section III and IV four presents the system model and Problem statement while Section V discuss the proposed solution. Next section highlights the mathematical analysis of algorithm while conclusion section provides an overview of the research along with results

II. PROCEDURE FOR PAPER SUBMISSION

S. Mehrotra and A.K.Agrawal replaced RSA with elliptic curves to build a trapdoor function. An integer factorization method called as Quadratic Sieve (QS) was used that enhanced time to calculate prime factorization. As QS helps in quicker factorization, thus a large key size can be used for same algorithm. On the other hand the security of the algorithm will be enhanced due to large key size [1]. The key size needs to grow quickly as the resources available to decrypt number increases. In the ideal situation of Trapdoor Function, the rate of easiness and hardness depends on the number size.

Peng Zhou and Xiojing Gu used HTTPAS as a new framework that is not only secure, but it can also increase the comprehensive verification against MITP incursion. CAs is utilizing the diversity of the Internet routes. Nevertheless, the MITP incursion can still affect the internet routes to the tunnel points from a network server. A solution to this issue from the client-side has been designed, that will allow a network user to earnestly gather the site certificates from a number of various websites. If the n certificates are issued from less than μ CAs

than they can be subject to MITP attack. The quantification of CAs specified indirectly by n can lead to overthrow MITP variant. There is an increased possibility that if the size of n is larger various CAs issues the certificates of the site. If we take the sample of (n) web sites with $|T|$ different CAs issued certificates, than all the CAs are beneficial in preventing MITP attack. [2] [3]

The real time example of application is Imperva SecureSphere Web Application Firewall (WAF) [4], It provides security to the data from cyber-attacks by analyzing the critical business web applications accessed by the users. SecureSphere WAF provides the high standard real time protection by vigorously analyzing the application by comparing their normal behavior with updated threat intelligence crowd-source around the whole world. It can block the attacks of SQL injection, cross-site scripting and remote file inclusion that makes full use of any exposed weaknesses in the web application. This application also shields the system from botnets and DDoS incursion and hence protects it from the attempt to take over the account in real-time before the false transaction can be performed [5] [6]. Faiza Babakano Jada used Elliptic Curves Cryptography (ECC) so solve the attacks from attackers. The attackers tend to take users' information from the system without bringing into notice of the user. The Elliptic Curves Cryptography helps to attain the security with few mathematical calculations of elliptic curves that uses the location of points to encrypt and decrypt the information. However, the problem still exists due to some reasons that can be solved by Diffie-Hellman, which aimed to enforce security level by implementing in TLS network environment. This solution can prevent an extensive protection to the users. [7][8]

Next, F5 Silverline DDoS shielding is the platform that analyzes and identifies the threat from the incoming traffic and sends back the clean traffic to the destination website by removing malicious threats and hence reduces or totally eliminates its effects to the end user or the network. This application gives hybrid solution for DDoS shielding by combining the high-capacity cloud service and DDoS protection capabilities and hence reduces or completely eliminates the complete spectrum of DDoS attacks [9].

The solution is by using a Mathematical Algorithm of Deffie-Hillman. In this algorithm, a prime number is selected. For example, to send Hamid a message, Yasir chooses a random secret. To send Yasir a message, Hamid chooses a random secret. Hamid receives and calculate the shared secret key same as the Yasir receives and calculate the shared secret key[10] [11].

For future works, Elliptic curve cryptography (ECC) is suitable as it can overtake due to increased use of it in smaller devices [10][12]. ECC is used with SSL/TLS due to increase in security needs which are the premier web security protocols. It is suitable for mobile cloud algorithms due to its low cost in power consumption. The world is swiftly transforming towards cloud computing and Internet of things (IOT) due to the gradual decrease in the size of the devices, e.g. sensors and home appliances may not have enough computational resources to use RSA as their cryptographic system. This new

generation of the algorithm is much better than the prevailing encryption methods in terms of providing security [14] [15]. The NIST recommended the utilization of standard curves give strong, reliable security in web communications on the Cloud or any other platform

III. PROPOSED ALGORITHM AND PROBLEM

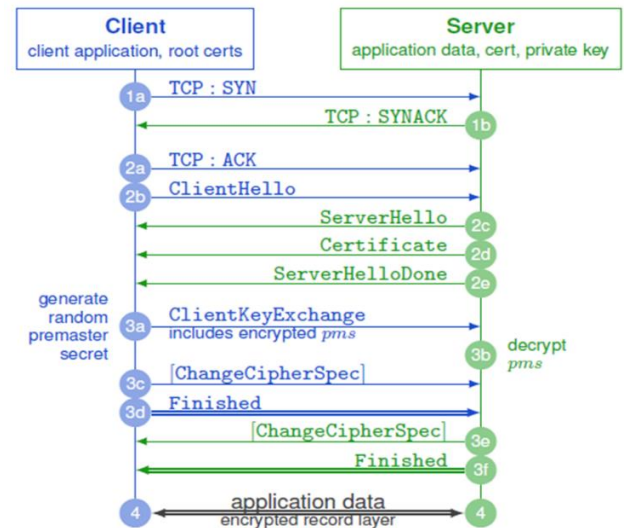


Figure 1: Network Model of TLS

TLS has some imperfections due to which the attacker can intercept the client's traffic performing MITM incursion by impersonating between the Server and the Client that makes the Client agree to downgrade the connection to the vulnerable TLS[16] [17]. If the attack on a security protocol that is designed to protect users is successful, it negates the purpose and hence endangering the integrity, confidentiality and authenticity of information transmitted [13][18].

IV. PROPOSED SOLUTION

A. Man-In-The-Middle (MITM)

Existing solution against issues of MITM incursion in TLS is discussed in a variety of literatures. This study defines the explanation and the novice solution which is proposed for the task. The goal of this suggested solution is to provide an extensive protection to MITM attack. MITM is an attacker that secretly impersonates and potentially altered the communication between two parties such as Client and Server. Both parties believing they are directly communicating with each other. One example of MITM attack in this study is Triple Handshake attack. The proposed solution is using Diffie-Hellman, which aimed to enforce security level by implementing in TLS network environment.

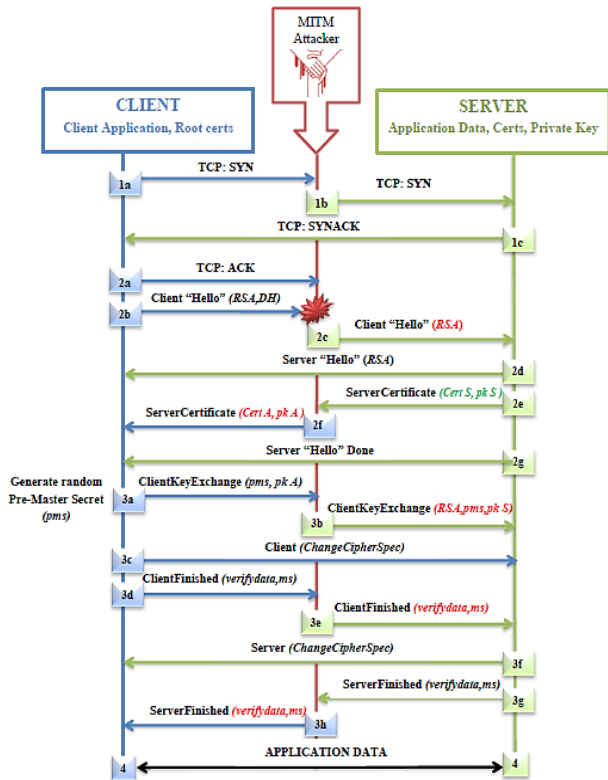


Figure 2: The triple handshake incursion

Triple handshake incursion is found in the negotiation phase. In the Triple Handshake attack against TLS, the server uses filched cookies on the client as the attacker pretends to be a client. The figure 2 above shows the processing of Triple Handshake attack on authenticated TLS renegotiation where (MITM) attacker is a malicious server.

The attacker proceeds the operations in four levels:

1. The client sends synchronize (SYN) TCP to the server. Server to synchronize acknowledges (SYNACK) TCP and send back to the sender.
2. The acknowledge TCP is sent but the sender thought it were directly communicating through the server. As the client sends "Hello" as a message to the server, the attacker maliciously connects to the server. When the Server sends the certificate to the client the attacker impersonates as a Client receives the information.
3. In Pre-Master Secret (PMS), the server does not participate in the key exchange. Then, the attacker receives the PMS generated and encrypted by the Client. Client decrypts using a private key and then sends it to the server that re-encrypts using server public key. The end of the handshake, attacker accomplishes the session after the client has done on verifying the data.
4. In the last level, the attacker stole the client information in the application data.

B. Diffie-Hellman Key Exchange (DHKE)

We suggest using Diffie-Hellman (DH). In Public Key Cryptography (PKC), Diffie-Hellman is an algorithm of the asymmetric key. IPsec is also used in the systems like SSL, SSH, PGP and PKI. DHKE can create a secret communication for exchanging data over a public network between the two parties.

It works between two peers generating a public and a private key. The Client will communicate by sending its public key to the Server and the Server would send its public key to the Client. After that the Client will use the public key that was received from the Server with its own private key to generate a symmetric key by using the Diffie-Hellman techniques. The same procedure is repeated at the end of the Server when Client communicates with the Server, enabling them to communicate securely over the insecure internet by using symmetric keys.

Still, there are some issues that were discovered by us later with the Diffie-Hellman algorithm such as MITM attacks, due to lack of authentication in place before the keys are exchanged that can lead to spoofing the Client's identity of the hacker. This issue can be overcome by using an authentication method, e.g. pre-shared keys and digital certificates to authenticate VPN gateways. Thus, it is suggested to use Diffie-Hellman Key Exchange alongside authentication algorithms as a protected and agreed solution.

V. MATHEMATICAL ANALYSIS

In this section, we provide the mathematical analysis as a proof that how an attacker can attack the traffic and its solution that can prevent from being attacked. Firstly, we will show as a mathematical how key Exchange (Diffie-Hellman) and attacker (Man-In-The-Middle) works.

To easily understand, we illustrate the scenario. For example Hamid sent a message to Yasir.

1. Hamid and Yasir are agreeing which the prime number $p = 41$ and $a = 4$.
2. Then, Hamid chooses $x = 16$ and sent Yasir $37 (\equiv 4^{16} \text{ mod } 41)$.
3. Next, Yasir chooses $y = 21$ and sent Hamid $37 (\equiv 4^{21} \text{ mod } 41)$.
4. Yasir receives 4 and computes $37 (\equiv 4^{21} \text{ mod } 41)$.
5. Hamid receives 4 and computes $37 (\equiv 4^{16} \text{ mod } 41)$.

That is why Hamid and Yasir have 4 as the mutually agreed secret key. For the example, the larger value of p used, will make the key agreement more secure and therefore such key space cannot be exhausted by the existing computing power. The important task of the attacker is to attack the network by searching a prime number because the attacker cannot continue to attack if he/she cannot find a prime number. The figure below, demonstrates which parties know what information is provided. In this case MITP is referred as Attacker.

Table 1: General Mathematical Algorithm of Diffie-Hellman and MiTM

Key Exchanges (Diffie-Hellman)	Attacker (MiTM)
<ol style="list-style-type: none"> 1. Select a prime number p and generator α of $Z_p (2 \leq \alpha \leq p-2)$. 2. To send Yasir a message $\alpha^x \bmod p$, Hamid choose a secret x haphazardly $1 \leq x \leq p-2$. 3. To send Hamid a message $\alpha^y \bmod p$, Yasir choose a secret y haphazardly $1 \leq y \leq p-2$. 4. Yasir receives α^x and calculate the shared key as $K = (\alpha^x)^y \bmod p$. 5. Hamid receives α^y and calculate the shared key as $K = (\alpha^x)^y \bmod p$. 	<ol style="list-style-type: none"> 1. Calculate $p-1/R$ to make sure it is an integer or not. 2. If it is an integer, then determine whether $p-1/R$ a prime number is or not. 3. If it is a prime number, then after message by raising to $p-1/R$. The attacker will determine the secret key is one number of R value.

The solution that we used to secure the network is providing a digital signature. Therefore, here it can help Yasir to make sure he got the message from a trusted source.

Table 2: The protocol of RSA signature scheme

Signature generation	Signature verification
<ol style="list-style-type: none"> 1. Compute $\tilde{m} = R(m)$, where an integer in the range $[0, n-1]$. 2. Compute $s = \tilde{m}^d \bmod n$. 3. Therefore, Hamid's signature for m is s. 	<ol style="list-style-type: none"> 1. Obtain Hamid's authentic public key (n, e). 2. Compute $\tilde{m} = s^e \bmod n$. 3. Recover $m = R^{-1}(\tilde{m})$.

Then, from Table 2, we implement it into terms of digital signature scheme. Here, we let E as symmetric encryption algorithm and $S_A(m)$ as Hamid's signature on m .

Step 1:

1. Select a prime number p and generator α of $Z_p (2 \leq \alpha \leq p-2)$.

Step 2:

1. To send Yasir a message $\alpha^x \bmod p$, Hamid chooses a random secret $x, 1 \leq x \leq p-2$
 $A \rightarrow B : \alpha^x \bmod p$ (message 1)

2. To send Hamid a message $\alpha^y \bmod p$, Yasir chooses a random secret $y, 1 \leq y \leq p-2$ and computes the shared key $k = (\alpha^x)^y \bmod p$

$$B \rightarrow A : \alpha^y \bmod p, E_k(S_B(\alpha^y, \alpha^x)) \text{ (message 2)}$$

3. Hamid computes the shared key as $K = (\alpha^x)^y \bmod p$. He decrypts the encrypted data, and uses Yasir's public key to verify the received value as the signature on the hash of the clear text exponential received and the exponential sent in message 1. If the verification is successful, Hamid accepts that k is actually shared with Yasir, and sends Yasir an analogous message.

$$A \rightarrow B : E_k(S_A(\alpha^x, \alpha^y)) \text{ (message 3)}$$

4. Yasir decrypts the received message and verifies Hamid's signature there. If the verification is successful, Yasir will accept that k is actually shared with Hamid.

So, it is proved that the solution given is a best solution to secure the network. In addition, if the process of receiving and sending a message is taking a long time, then there is a higher possibility of the network to be hacked or interrupted.

VI. CONCLUSION

Last but not least, the web servers can easily become the target for attackers and that is why security is very important for both internet facing and intranet servers. By providing security encryption and message integrity in a client-server environment can give security to the system. As the security challenges increases in the communication, the security measurements should also be increased. The security level will increase if we use a large key size prime number, because it is not easy to break. Therefore, it must be powerful enough so that the hackers cannot break the key. Security and privacy always a major concern in the data communication. Data security is an imperative part of nature, of administration accordingly. Security must be imposed on information by utilizing encryption techniques to accomplish secured information data storage and access. Due to the obscurity nature of the cloud, it is as yet having security issues. The cloud foundation much more dependable and effective than individualized computing, however extensive variety of interior, outside dangers of information put away on the cloud. Since the information are not put away in customer zone, carrying out safety efforts can't be connected specifically..

ACKNOWLEDGMENT

We would like to express my gratitude to Network Security Research Group, Faculty of Computer Science and Information Technology at UNIMAS Sarawak.

REFERENCES

- [1] Agrawal, A. K., & Mehrotra, S. (2016). Application of elliptic curve cryptography in pretty good privacy (PGP). 2016 International Conference on Computing, Communication and Automation (ICCCA).

- [2] Apostolopoulos, G., Peris, V., & Saha, D. (1999). Transport layer security: how much does it really cost? IEEE INFOCOM 99. Conference on Computer Communications. Proceedings. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. The Future is Now (Cat. No.99CH36320). doi:10.1109/infcom.1999.751458
- [3] E. S. Alashwali, "Cryptographic vulnerabilities in real-life web servers." pp. 6-11.
- [4] SecureSphere Web Application Firewall (WAF) – Real-time protection against web attacks. (n.d.).Retrieved May 11, 2017, from <https://www.imperva.com/Products/WebApplicationFirewall-WAF>
- [5] Haque, M. S. (2016). Web Server Vulnerability Analysis in the context of Transport Layer Security (TLS). International Journal of Computer Science Issues, 13(5), 11-19. doi:10.20943/01201605.1119
- [6] Joppe W. Bos et al., "Elliptic Curve Cryptography in practice," Financial Cryptography and data security, 18th international conference FC 2014, Christ Church, Barbados, March 3-7, 2014, Springer, LNCS 8437.
- [7] Li, X., Xu, J., Zhang, Z., & Feng, D. (2016). On the security of TLS resumption and renegotiation. China Communications, 13(12), 176-188. doi:10.1109/cc.2016.7897542
- [8] Onomzawaziri, V., Danladi, H., Alhassan, J. K., & Jada, F. B. (2015). Web-Cloud-based Security Services Based-on Elliptic Curves Cryptosystem. International Journal of Computer Applications, 115(23), 12-18.
- [9] Parmar, H., & Gosai, A. (2015). Analysis and Study of Network Security at Transport Layer. International Journal of Computer Applications, 121(13), 35-40. doi:10.5120/21604-4716
- [10] Javed, Y., Khan, A. S., Qahar, A., & Abdullah, J. (2017). EEoP: A Lightweight Security Scheme over PKI in D2D Cellular Networks. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 9(3-11), 99-105.
- [11] W. Diffie and M. E. Hellman. New Direction in Cryptography. IEEE Transaction on Information Theory, IT-22(6):74-84, June 1997
- [12] Stebila, D., & Sullivan, N. (2015, August). An analysis of TLS handshake proxying. In Trustcom/BigDataSE/ISPA, 2015 IEEE (Vol. 1, pp. 279-286). IEEE.
- [13] Waziri, V. O., Danladi, H., Alhassan, J. K., & Jada, F. B. (2015). Web-Cloud-based Security Services based-on Elliptic Curves Cryptosystem. International Journal of Computer Applications, 115(23).
- [14] Web Server Vulnerability Analysis in the context of Transport Layer Security (TLS). (2016). International Journal of Computer Science Issues, 13(5), 11-19. doi:10.20943/01201605.1119
- [15] Javed, Y., Khan, A. S., Qahar, A., & Abdullah, J. (2017). Preventing DoS Attacks in IoT Using AES. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 9(3-11), 55-60.
- [16] Y. Sheffel, R. Holz, P. Saint-Andre May 2015 Recommendation for secure use of TLS and DTLS, <http://www.ietf.org/html/rfc7525>, RFC 7525.
- [17] Khan, A. S., Javed, Y., Abdullah, J., Nazim, J. M., & Khan, N. (2017). Security issues in 5G device to device communication. IJCSNS, 17(5), 366.
- [18] Zhou, P., & Gu, X. (2016). HTTPAS: active authentication against HTTPS man-in-the-middle attacks. IET Communications, 10(17)

ICATCES 2019

INTERNATIONAL CONFERENCE ON
ADVANCED TECHNOLOGIES,
COMPUTER ENGINEERING AND SCIENCE



WWW.ICATCES.ORG

ISBN: 978-605-9554-18-3

