INTERNATIONAL CONFERENCE ON

ADVANCED TECHNOLOGIES, COMPUTER ENGINEERING AND SCIENCE

KARABUK, TURKEY

3-5 JUNE 2020

2020

PROCEEDING BOOK

ICATCES 2020
Proceeding Book

International Conference on Advanced Technologies,
Computer Engineering and Science

Jun 03-05, 2020 / Karabuk, Turkey

Proceeding Book of the International Conference on Advanced Technologies, Computer Engineering and Science (ICATCES 2020)

Editors
Prof. Dr. Oğuz FINDIK
Asst. Prof. Dr. Caner ÖZCAN
Res. Asst. Yusuf Yargı BAYDİLLİ

Published, 2020.

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

Honorary Committee
Prof. Dr. Refik POLAT, Karabük University, Rector
Prof. Dr. Mehmet AKBABA, Karabük University

Chair
Prof. Dr. Oğuz FINDIK, Karabük University
Asst. Prof. Dr. Caner ÖZCAN, Karabük University

Co-Chair
Asst. Prof. Dr. Hakan KUTUCU, Karabük University

Organization Committee
Prof. Dr. Oğuz FINDIK, Karabük University
Assoc. Prof. Dr. İlker TÜRKER, Karabük University
Asst. Prof. Dr. Caner ÖZCAN, Karabük University
Asst. Prof. Dr. Emrullah SONUÇ, Karabük University
Asst. Prof. Dr. Hakan KUTUCU, Karabük University
Asst. Prof. Dr. Rafet DURGUT, Karabük University
Res. Asst. Yusuf Yargı BAYDİLLİ, Karabük University

Scientific Committee
Prof. Dr. Ali KARCI (İnönü University)
Prof. Dr. Banu DİRİ (Yıldız Technical University)
Prof. Dr. Basel MAHAFZAH (The University of Jordan, Jordan)
Prof. Dr. Bilal ALATAŞ (Fırat University)
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, Pakistan)
Prof. Dr. Haldun GÖKTAŞ (Yıldırım Beyazıt University)
Prof. Dr. Harun UĞUZ (Konya Technical University)
Prof. Dr. İsmail Rakıp KARAŞ (Karabük University)
Prof. Dr. Ivan DEMYDOV (Lviv Polytechnic National University, Ukraine)
Prof. Dr. Kemal POLAT (Abant Izzet Baysal University)
Prof. Dr. Mario KOEPPEN (Kyushu Instutite Of Technology, Japan)
Prof. Dr. Mehmet AKBABA (Karabük University)
Prof. Dr. Mehmet ÖZALP (Karabük University)
Prof. Dr. Murat YÜCEL (Gazi University)
Prof. Dr. Mykola S. NIKITCHENKO (Taras Shevchenko National University of Kyiv, Ukraine)
Prof. Dr. Oleksandr I. Provotar (Taras Shevchenko National University of Kyiv, Ukraine)
Prof. Dr. Oleksandr O. Marchenko (Taras Shevchenko National University of Kyiv, Ukraine)
Prof. Dr. Oğuz FINDIK, Karabük University
Prof. Dr. Oğuz GÜNGÖR (Karadeniz Technical University)
Prof. Dr. Okan ERSOY (Purdue University, USA)
Prof. Dr. Raif BAYIR (Karabük University)
Prof. Dr. Sergiy D. Pogorilyy (Taras Shevchenko National University of Kyiv, Ukraine)
Prof. Dr. Serhii L. Kryvyi (Taras Shevchenko National University of Kyiv, Ukraine)
Prof. Dr. Vasyl M. Tereshchenko (Taras Shevchenko National University of Kyiv, Ukraine)
Prof. Dr. Valentina Emilia BALAS (Aurel Vlaicu University of Arad, Romania)
Prof. Dr. Yaşar BECERİKLİ (Kocaeli University)
Assoc. Prof. Dr. Abdrakhmanov RUSTAM (Ahmet Yesevi University)
Assoc. Prof. Dr. Adib HABBAL (Karabük University)
Assoc. Prof. Dr. Ahmad Taher AZAR (Benha University, Egypt)
Assoc. Prof. Dr. Amirtayev KANAT (Ahmet Yesevi University)
Assoc. Prof. Dr. Andrew D. BAGDANOV (University of Florence, Italy)

Secretary

Res. Asst. Ayşe Nur ALTINTAŞ, Karabük University
Res. Asst. Berna YENİGÜN, Karabük University
Res. Asst. Elif KABULLAR, Karabük University
Res. Asst. Furkan SABAZ, Karabük University
Res. Asst. Hasan KIVRAK, Karabük University
Res. Asst. İdris KAHRAMAN, Karabük University
Res. Asst. Mehmet Zahid YILDIRIM, Karabük University
Res. Asst. Said DEMİR, Karabük University
Res. Asst. Yasemin SANDAL, Karabük University

Welcome Address

It is a pleasure for us to offer you Abstracts Book for the 3rd International Conference on Advanced Technologies, Computer Engineering and Science ICATCES 2020. 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 2020 for accepting our invitation. We hope that you enjoy the symposium and look forward to meeting you again in one of the forthcoming ICATCES 2021 event.

Contents

Proceedings

Ship Detection on Highly Complex Satellite Imageries using CNN

Ibrahim GULATAS [1][0000-0002-0804-3588] and Metin TURAN [2][0000-0002-1941-6693]

[1] İstanbul Commerce University, İstanbul, Turkey
[2] İstanbul Commerce University, İstanbul, Turkey
lncs@springer.com

Abstract. Satellite imagery has a wide variety of usage for various markets such as agriculture, defense and intelligence, energy and finance. The recent improvements on the machine learning techniques help those markets to gather more information from the satellite imageries. One of the most challenging areas is detecting ships on large satellite imagery with complex environment. In this paper, we experiment a convolutional neural network model for detecting ships on the common shared satellite imageries. With the recent developments on the processing powers of the computers make it possible to use deep learning algorithms. The feature extraction part and deep learning layers in the model were optimized. Moreover thanks to the simple heuristic approach used for detecting ships on the complex backgrounds, the proposed method resulted with both remarkable increment of accuracy and decrement of training speed than the researches have been done until now. Probability of detection metric and quality factor metric were obtained approximately 69% and 94% respectively.

Keywords: Ship Detection, Convolutional Neural Network, Deep Learning, Machine Learning.

1 Introduction

The usage of machine learning techniques, and especially image processing, to solve daily problems is increasing day by day. One of these topics is also the shipping traffic on the ports. It has been getting increase even faster than expected. The vast increase on the marine traffic also makes it possible of infractions like environmentally destructions due to ship accidents, and moreover piracy, illegal fishing, drug trafficking, and illegal cargo movement. Current problems have compelled many organizations, from environmental protection agencies to insurance companies and national government authorities, to have a closer watch over the open seas. As a result, automatic ship detection is one of the most challenging applications that gains attention to deal with nowadays.

Aerial and satellite imageries are very important resource technology in order to detect movements on time. They are also used in many fields for object detection such

as, detection of roads, buildings, trees, plants, vehicles currently [1-2]. Some different type of applications has been developed in order to detect ships on satellite and radar imageries. The first example of these research has been done by Shengxiang Qi and his colleagues [3]. They proposed a novel descriptor that determines ship histogram of oriented gradient to discriminate real ships, which characterizes the gradient symmetry of ship sides. The second example of the rest is the research done by Gu-Hong Nie and his friends [4]. They introduce a transfer learned Single Shot Multi-Box Detector (SSD) for ship detection. An object detection model pre-trained from a large number of natural images was transfer learned for ship detection with limited labeled satellite imageries. On the other hand, Harm Greidanus and others work on Search for Unidentified Maritime Objects (SUMO) algorithm [5]. SUMO is a pixel based CFAR (Constant False Alarm Rate) detector for multi-look radar imageries used for 20 years approximately. They investigated and discussed the algorithmic approach in all of the steps of the ship detection: land masking, clutter estimation, detection threshold, target clustering, ship attribute estimation and false alarm suppression. The other technique was proposed by Zikun Liu and others [6]. They first analyzed the possibility of accurately covering all ships by labeling rotated bounding boxes. Later, they constructed nearly closed-form ship rotated bounding box space in order to reduce search space. Then, they selected a small number of highly potential candidates by scoring each candidate in the space using a two-cascaded linear model.

Lately, Convolutional Neural Networks (CNN) has been used widely for object detection. Aerial and satellite imageries are appeal the researchers for object detection problems. These imageries are used for many fields. Chen et al. used satellite imagery to detect vehicles with CNN [7]. Wu et.al detects fast aircraft in satellite imagery based on CNN [8]. Janalipour proposed a method for detecting damaged buildings by the help of the aerial imageries [9]. Moreover, Li et al. proposed a method for detecting palm tree located on the images which are taken by an UAV [10]. One of the most impressive researches of CNN applications has been done by Zhang et al. [11] lately. They suggested two ship models, the “V” ship head model and the “||” ship body one, to localize the ship proposals from the line segments extracted from a test image. Due to relatively small sizes, lack of reliable line segments, they propose an improved saliency detection method to find these proposals. The other valuable research has been achieved by Liu and his colleagues [12]. They observed that satellite imageries expose strip-like rotated assembled object detection which is common in remote sensing images. Hence, they introduced the rotated region based CNN (RR-CNN), which can learn and accurately extract features of rotated regions and locate rotated objects precisely. Finally, Zou et al. were used linear SVM classifier and CNN in order to detect ships, called that method SVDnet [13].

There are a few researches on the literature about ship detection on aerial and satellite imageries. Most of these researches are applied to synthetic aperture radar (SAR) imageries to detect ships. SAR has wide

angle vision, it does not affected from the clouds on the sky and it can collect data at night as well as at daylight. For all these reasons researchers used SAR imageries to detect ships. Lin et al. [14], Kang et al. [15] and Lend et al. [16] are some of the recent researches on this field. Faster R-CNN based algorithms are the mostly preferred algorithms for SAR imageries.

Recent developments on the optical sensors and space technology bring us the vast increase on the satellite numbers on the earth orbit. This increase on the satellite numbers yields us to use these satellite imageries to detect ships. However this new technology brings new challenges. The satellite imageries are taken with wide angle cameras which cover a huge field. On the satellite imageries ships are demonstrated on only a few pixels.

This research mainly focuses on ship detection on the locations which have heavy marine traffic such as harbours, anchorage areas and narrow straits, to help to authorities to regulate the traffic. These heavy traffic area imageries also contain lots of objects that are located on the shore and they create very complex backgrounds for detecting ships. With the recent developments on the processing powers of the computers make it possible to use deep learning algorithms. One of the most popular application fields of deep learning technique is the object detection on the images. A novel approach for ship detection on satellite imageries is experimented in this research to get the advantage of the convolutional neural networks (CNN). The results showed that thanks to the simple heuristic approach for detecting ships on the complex backgrounds results with remarkable increment of accuracy and decrement of training speed than the researches done until now.

This paper continues with the details of proposed method in the section II. The performance evaluation and results are presented in the section III. Finally, conclusion and discussions are given in Section IV.

2 Methodology

The data set used for this project is obtained by constellations of small satellites of Planet Company [17]. The dataset contains 4000 number of cropped area images of satellite imagery. The size of the images is 80x80 pixels (Fig. 1). The 25 percent of the data set are ship images and the rest of them are non-ship images, namely there are 1000 ship and 3000 non-ship images. On the satellite imageries ships are demonstrated on only a few pixels.

Fig. 1. Shows the some sample of ship and non-ship

The entire data set is saved as .png image file. Each individual file name contains specific information about the image. The each filename follows a specific format: [label] [scene id] [longitude][latitude].png. The images of the data set are converted and written into a .json file to speed up training process and avoid loss of time when trying different training models. During the creation of the .json file, the object is saved as four lists. The first list consists of 19.200 integer values which show the pixel information. These 19.200 integers are divided by 3. Each 6400 integer value belongs to red-green-blue channels. The second list is the labels of the images. The "0" value refers to non-ship images and "1" refers to ship images. The third list gives the GPS location of the images. The fourth and the last list contain the image id's. Since this research is not focused on the geographical locations of the ships the last two lists are ignored in this study. Fig.2 shows the first four records of the .json object.

```
                                                 data  labels  \
0  [82, 89, 91, 87, 89, 87, 86, 86, 86, 86, 84, 8...       1
1  [76, 75, 67, 62, 68, 72, 73, 73, 68, 69, 69, 6...       1
2  [125, 127, 129, 130, 126, 125, 129, 133, 132, ...       1
3  [102, 99, 113, 106, 96, 102, 105, 105, 103, 10...       1
4  [78, 76, 74, 78, 79, 79, 79, 82, 86, 85, 83, 8...       1

                                locations             scene_ids
0    [-118.2254694333423, 33.73803725920789]  20180708_180909_0f47
1    [-122.33222866289329, 37.7491755586813]  20170705_180816_103e
2  [-118.14283073363218, 33.736016066914175]  20180712_211331_0f06
3   [-122.34784341495181, 37.76648707436548]  20170609_180756_103a
4   [-122.34852408322172, 37.75878462398653]  20170515_180653_1007
```

Fig. 2. The first four records of the .json object

The "sequential" model of the "Keras" library is used for the creation of the training model. First step of our research is to import the ship images to the convolutional neural network for feature extraction. Then polling layers are added to the model in order to downsample the image. As it is mentioned herein before, the input images are 80x80 pixels. It is shown on the Fig. 1, as ship figure on the images involves most of the pixels, so bigger kernel size could be used for feature extraction of ship images. However, on the images which do not contain ship figure may need smaller kernel size for feature extraction. For these reasons the convolutional network designed with four convolutional layers each contains 32 neurons with 3x3 kernel size and 4 polling layers with 2x2 pool size. This is generally well accepted and used structure in the literature [18-19]. Fig. 3 shows the proposed CNN model and Fig. 4 shows the training model code.

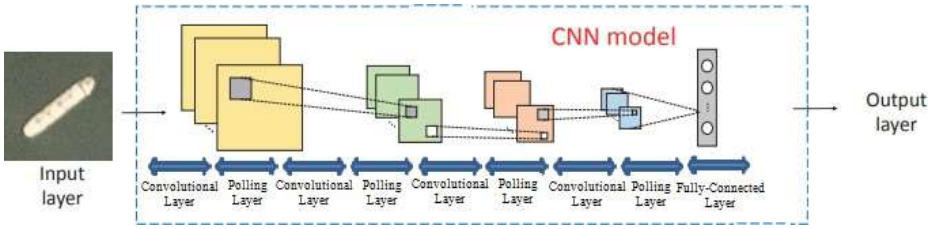


Fig. 3. Proposed CNN model

```
model.add(Conv2D(32, (3, 3), padding='same', input_shape=(80, 80, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2))) #40x40
model.add(Dropout(0.25))

model.add(Conv2D(32, (3, 3), padding='same', activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2))) #20x20
model.add(Dropout(0.25))

model.add(Conv2D(32, (3, 3), padding='same', activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2))) #10x10
model.add(Dropout(0.25))

model.add(Conv2D(32, (10, 10), padding='same', activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2))) #5x5
model.add(Dropout(0.25))

model.add(Flatten())
model.add(Dense(512, activation='relu'))
model.add(Dropout(0.5))

model.add(Dense(2, activation='softmax'))
```

Fig. 4. Training code

Gradient Descent based learning methods are used for the much of the successful machine learning applications [20]. Gradient Descent methods are divided into two sections as stochastic and batch gradient descent methods. Both of them have their own characteristics. In general stochastic method is faster than batch method especial-

ly on large datasets. Also stochastic method has better accuracy performance on non-linear problems. Besides, stochastic method is more useful for the datasets which changes gradually. For these reasons, "Stochastic Gradient Descent" (SGD) is chosen as optimizer for this research. Different epoch numbers are tried to find the best training model. Besides, "categorical cross entropy" is used for loss function and "accuracy" is used as metric.

3 Results

The training model is saved into a .json file to prevent loss of time. The training model tested on various test images. These test images are belongs to some harbour entrances which has lots of ship images as well as lots of images on the land. Since these test images belongs to inshore positions they have very complex backgrounds. This complex background makes the problem even harder to detect one ship on an offshore image. Fig. 5 shows the one of the sample test image

Fig. 5. A sample test image

Our training model is applied to the test image with sliding windows technique. Our method uses 10 pixels steps to search for ships on the image. From the top left corner of the testing image, sliding windows are iterated per 10 pixels. Each window has 80x80 pixels size. However if the window only covers some part of the ship but does not cover the whole ship, the window enlarged to locate the ship on the center of the window. For the each window our training method is applied if the score of the window is below the 0.9 value this window is marked as non-ship.

If the score of the window is above the 0.9 this window is highlighted and marked as ship. Also the coordinates of the highlighted area are shown on the screen.

We have tested our model on various test imageries. Our dataset is trained with 80x80 pixels ship and non-ship samples. However the size of the our test imageries are about 2800x1800 pixels and the biggest ship image only takes 94x39 pixels on the image. According to the epoch numbers, the accuracy of the model is shown on Fig. 6 and loss is shown on Fig. 7.

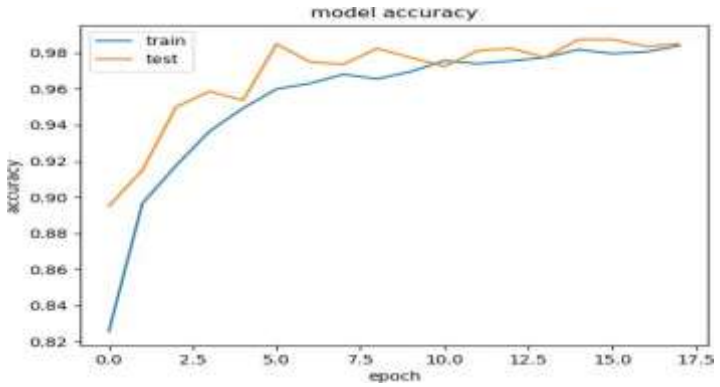


Fig. 6. Accuracy of the model

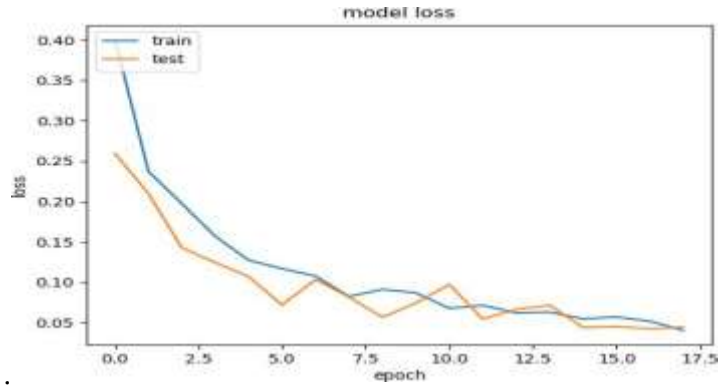


Fig. 7. Loss of the model

In this paper results of the only three of the test imageries are presented. Test images used are given in Fig. 8, Fig. 9 and Fig. 10 respectively. We used the same performance metrics mentioned on the [6]. Probability of Detection (PD) and Missed Detection Rate (MDR), False Alarm Rate (FAR) and Quality Factor (QF) are often used to quantitatively evaluate detection performance. They are defined as follows:

$$PD = N_{tt}/N_{dt} \quad (1)$$

$$FAR = N_{ft}/N_{dt} \quad (2)$$

$$MDR = N_{ut}/N_{at} \quad (3)$$

$$QF = N_{rt}/N_{at} \quad (4)$$

Here, Ntt denotes the number of targets to be detected correctly, where Nft is the number of detected error targets, Ndt denotes the number of all detected targets, Nut represents the number of targets that are not detected, Nrt represents the number of targets that are detected, and Nat represents the number of targets. Table 1 represents the performance metrics of these 3 test imageries.

Table 1. Shows the evaluation of the test images

Image	PD	FAR	MDR	QF
Image-I	0,69	0,31	0	100
Image-II	0,75	0,42	0,22	0,78
Image-III	0,65	0,35	0	100
Total	0,69	0,35	0,06	0,94

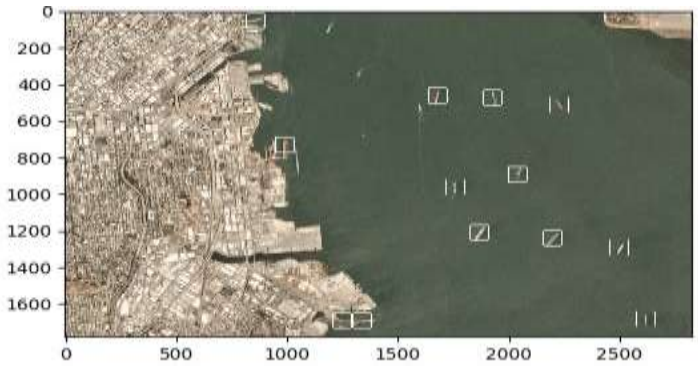


Fig. 8. Test-I image

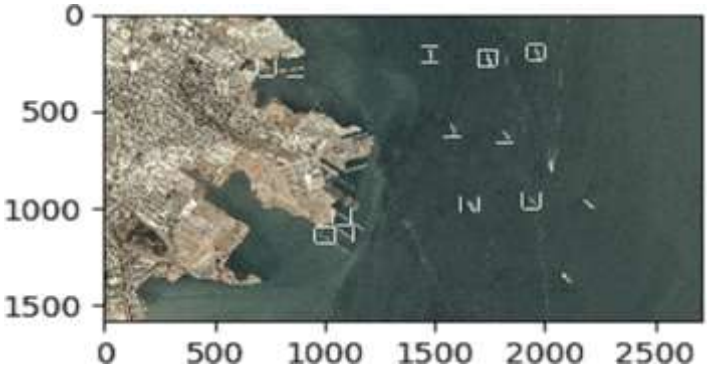


Fig. 9. Test-II image

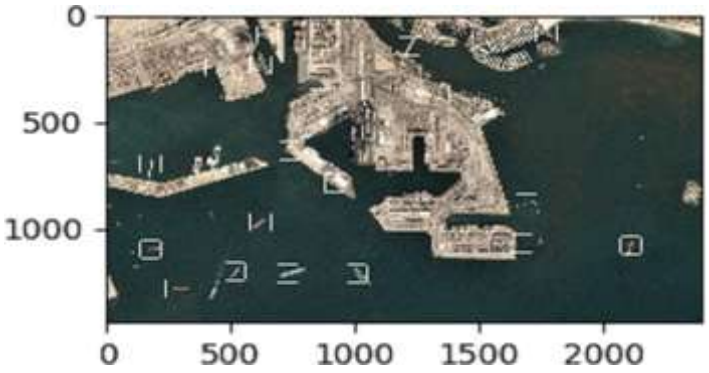


Fig. 10. Test-III image

The evaluation of our model shows that, it is very successful for detection of the ships but it creates some false positives namely it marks some non-ship object as ships. Comparison of our model with the recently proposed models on the literature shows that our model has far more better results than other studies. Generally most of the researches are using offshore imageries to detect ships which obtain very successful results. On the other hand, the most related research [6] to our work found in the literature uses very similar test imageries. However, it experiments SAR imageries instead of satellite imageries. This research average performance metrics are PD:0.36, MDR:0,64, FAR:0.28, QF:0,72. These results shows that our model has better scores for detecting ships.

The performance of the proposed Convolutional Network Model is also compared one of the most popular object detection model AlexNet [21]. AlexNet model has significant performance metrics on object detection problems, especially on images with 256x256 pixels size. As it is mentioned herein before images in our training dataset sized 80x80 pixels and ship images occupy most of these pixels. AlexNet

model has the highest accuracy rate of 98% on the epoch 16. Training the model for 16 epochs takes 6753 seconds. On the hand, our proposed model reaches the highest accuracy on the epoch 17 with the rate of 98.62% which is slightly higher. Moreover, the training of our proposed network takes 980 seconds for 17 epochs which is almost 7 (6,89) times faster than AlexNet. The accuracy rates and runtimes of the each epoch for both networks are given on Fig. 11.

4 Conclusion

Satellite imagery has a wide variety of usage area for various markets such as agriculture, defense and intelligence, energy, and finance. In this research, satellite imagery is used for ship detection on images belongs to harbour entrance which has heavy marine traffic.

Even though there are some researches about ship detection on aerial or satellite imagery only a few of the aimed to detect ships on inshore area which has a complex background. This complex background makes the problem more confusing for object detection.

Convolutional neural network is used for this research. The evaluation metrics of the proposed method is very remarkable comparing to the similar researches. The accuracy of the proposed method is very high. On the three test images which are mentioned herein before, almost all of the target objects are detected. However our model creates some false positive defections. It labels some objects like breakwater or piers as ships. We believe that with the some improvements using image processing technique, these false labels could be cleaned up.

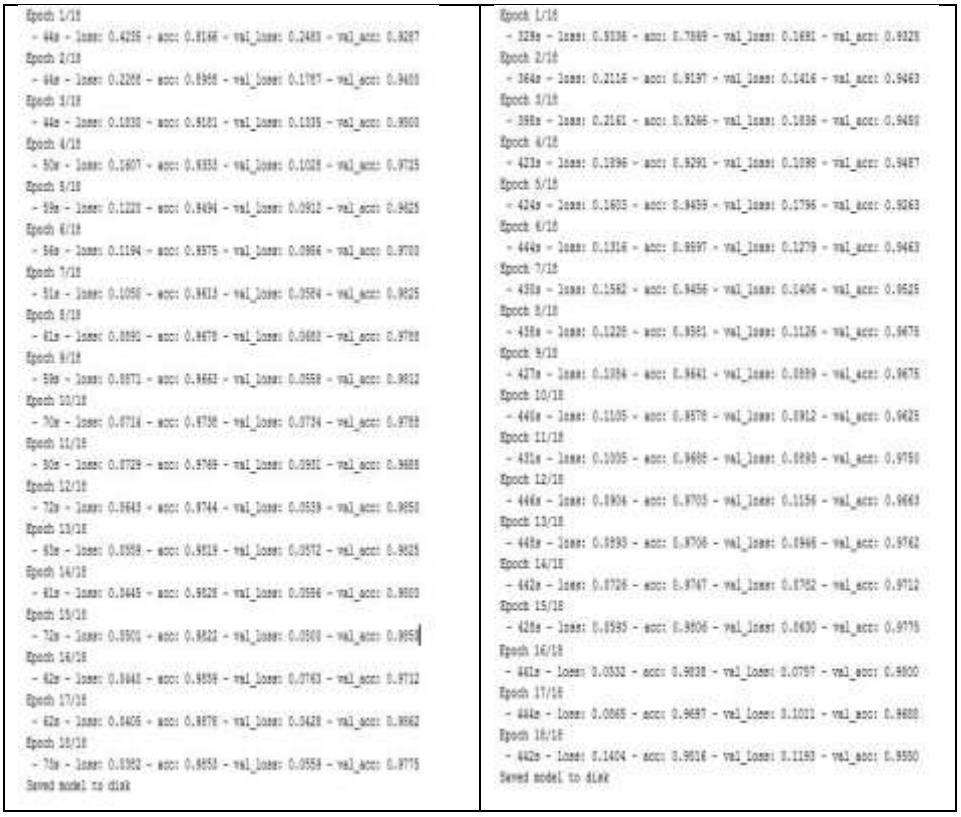


Fig. 11. Training of proposed model (on the left) and AlexNet (accuracy and runtime)

References

1. Cao L, Luo F, Chen L, Sheng Y, Wang H, Wang C, Ji R.: Weakly supervised vehicle detection in satellite images via multi-instance discriminative learning. Pattern Recognition, 64, 417-424 (2017).
2. Ishii T, SimoSerra E, Iizuka S, Mochizuki Y, Sugimoto A, Ishikawa H, Nakamura R.: Detection by classification of buildings in multispectral satellite imagery. IEEE 23rd International Conference on Pattern Recognition (ICPR), 3344-3349 (2016).
3. Qi S, Ma J, Lin J, Li Y, Tian J.: Unsupervised ship detection based on saliency and S-HOG descriptor from optical satellite images.IEEE Geoscience and Remote Sensing Letters, 21 (3), 226-234 (2015).
4. Nie G H, Zhang P, Niu X, Dou Y, Xia F.: Ship Detection Using Transfer Learned Single Shot Multi Box Detector. ITM Web of Conferences, 12, 01006 (2017).
5. Greidanus H, Alvarez M, Santamaria C, Thoorens F X, Kourti N, Argentieri P. : The SUMO ship detector algorithm for satellite radar images. Remote Sensing, 9(3), 246 (2017).
6. Liu Z, Wang H, Weng L, Yang Y.: Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds. IEEE Geoscience and Remote Sensing Letters, 13(8), 1074-1078 (2016).
7. Chen X, Xiang S, Liu C L, Pan C H.: Vehicle detection in satellite images by hybrid deep convolutional neural networks. IEEE Geoscience and remote sensing letters, 11(10), 1797-1801 (2014).
8. Wu H, Zhang H, Zhang J, Xu F.: Fast aircraft detection in satellite images based on convolutional neural networks. IEEE International Conference on Image Processing (ICIP), 4210-4214 (2015).
9. Janalipour M, Mohammadzadeh A.: Building damage detection using object-based image analysis and ANFIS from high-resolution image (case study: BAM earthquake, Iran). IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(5), 1937-1945 (2016).
10. Li W, Fu H, Yu L, Cracknell A.: Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. Remote Sensing, 9(1), 22 (2016).
11. Zhang R., Yao J., Zhang K., Feng C., Zhang J.: S-CNN Based Ship Detection From High-Resolution Remote Sensing Images, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B7 (2016).
12. Liu Z., Hu J., Weng L., Yang Y.: Rotated Region Based CNN For Ship Detection, IEEE International Conference on Image Processing (ICIP) (2017).
13. Zou Z, Shi Z.: Ship detection in spaceborne optical image with SVD networks. IEEE Transactions on Geoscience and Remote Sensing, 54(10), 5832-5845 (2016).
14. Lin Z, Ji K, Leng X, Kuang G.: Squeeze and Excitation Rank Faster R-CNN for Ship Detection in SAR Images. IEEE Geoscience and Remote Sensing Letters, (2018).
15. Kang M, Leng X, Lin Z, Ji K.: A modified faster R-CNN based on CFAR algorithm for SAR ship detection. IEEE International Workshop on Remote Sensing with Intelligent Processing (RSIP), 1-4 (2017).

16. Leng X, Ji K, Zhou S, Xing X, Zou H.: An adaptive ship detection scheme for spaceborne SAR imagery. Sensors, 16(9), 1345 (2016).
17. Planet Team. Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA, 2017. https://api.planet.com last accessed 09/052019.
18. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 818–833. Springer, Heidelberg (2014).
19. Ciresan, D., Giusti, A., Gambardella, L. M., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images, Advances in neural information processing systems 2843-2851 (20112).
20. LeCun, Y. A., Bottou, L., Orr, G. B., Müller, K. R.: Efficient backprop. In Neural networks: Tricks of the trade, 9-48 (2012).
21. Krizhevsky, A., Sutskever, I., Hinton, G. E.: Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, 1097-1105 (2012).

Generalized Cost Effective Automatic Dictionary Creation Using the Big Data on the Web

Ahmet Toprak[1] and Metin Turan[1]

[1] Istanbul Commerce University, Istanbul, Turkey
[1] Istanbul Commerce University, Istanbul, Turkey

Abstract. Domain specific dictionaries can help in order to obtain meaningful information from big data. Studies in language lexicography have been focusing on automatic dictionary creation lately. In this study, a new method is proposed to create a domain specific English dictionary by unsupervised way. Furthermore, the dictionary can be extended later without any initialization cost.

Method requires an English document (or set of documents) related to the specific domain which is a reference for initial process, called reference document. The meaningful words exists in the reference document are identified by term frequency–inverse document frequency (TF–IDF) values. A query with the meaningful words obtained from the last iteration is applied to the Azure Web Cognitive Search Interface. The meaningful words found in the new documents obtained by the query are not added directly to the dictionary, evaluated by the WordNet dictionary for the similarity of each meaningful word within the dictionary. The meaningful words with higher similarity values above a certain threshold value (experimentally learned) are added to the dictionary until the desired number of words for the dictionary is reached. This method can easily be adapted to the all languages dictionaries. Hash similarity metric is used to measure the performance of the dictionary. Dictionaries up to 75.2% hash similarity were generated by tests applied on different subjects.

Keywords: Automatic Dictionary Creation, Big Data, Hash Similarity, WordNet.

1 Introduction

There has been need to obtain meaningful information from big data lately. Domain specific dictionaries can help in order to achieve this purpose. Therefore, studies in language lexicography have been focusing on automatic dictionary creation. These dictionaries can be used for search engine optimization (Vorapatratorn et al. 2012), automatic summarizing systems (Hovy et al. 2001; Gholamrezazadeh et al. 2010), theme determination (Habacha et al. 2011) and text classification (Nguyen et al. 2013) researches.

The dictionary creation process can be done manually (Kepuska et al. 2011), semi-automatically (Koeva et al. 2016), or automatically (Ellen 1993). Hand-created dictionaries are become static, and they require continuous external intervention to increase the size of the dictionary, resulting in higher maintenance costs. It is important to automate this process in order to lower this maintenance cost and to obtain more generic applications.

Riloff (1993) described the shortcomings of manual dictionaries. Knowledge-based natural language processing (NLP) systems are often criticized although they have achieved good success with specific tasks because they are dependent on a domain-specific dictionary that requires manual engineering knowledge. This bottleneck in engineering knowledge is not practical for implementing knowledge-based NLP systems in real-world applications. Thus, these systems cannot be easily scaled or transferred to the new fields.

In this study, an algorithm was proposed and experimentally tested with the aim of creating an automatic dictionary for a specific domain. The initial step is to find meaningful words from the document (or set of documents) given by the user. This document is called reference document. The meaningful words representing the reference document were identified using term frequency–inverse document (TF–IDF) values. The first dictionary words, called seed words, are obtained from the meaningful words of the reference document. A web search is then applied using the Azure Web Cognitive Search Interface repetitively in order to construct a query using the meaningful words obtained from the last iteration. The meaningful words found aren't added directly to the dictionary. Instead, the WordNet dictionary is used to evaluate the similarity of each meaningful word within the dictionary so that adding meaningless words to the dictionary is prevented. In this dictionary, meaningful words with higher similarity values above a certain threshold value (experimentally learned) are added to the dictionary until the desired number of words in the dictionary is completed. Thus, the dictionary will continue growing automatically and continuously without any intervention from the outside.

Hash similarity was used to measure the overall similarity of the dictionary. The notion of locality-sensitive hashing was first introduced by Indyk et al. (Indyk et al. 1998). For efficient near neighbor search, the locality-sensitive hashing (LSH) technique exploits special hash functions that make buckets contain similar keys (data), yet it does not guarantee that all data in a bucket are similar to each other. The so-called locality-sensitive hash functions provide high probability for similar data to be in the

same bucket but low probability for dissimilar data to be in the same bucket (Lee et al. 2012).

In the second section of the article, prior works in the field of corpus generation are discussed. The methods and studies used for dictionary creation are explained in the third section. The data sets mentioned in the third section are explained in the fourth section. Finally, results are discussed and evaluated in the last section.

2 Prior Work

TF–IDF has been successfully applied in many different NLP applications. One of these important studies was by Paik (Paik 2013), who performed an effective ranking study in 2013 using TF–IDF. In this particular study, Paik addressed the limitations of the pivoted length normalization by exploiting new statistical factors in the Multi Aspect TF (MATF) schema. One component of the term frequency is effective for short queries, while the other performs better on long queries. The final weight is then measured by taking a weighted combination of these components, which is determined based on the length of the corresponding query. Experiments conducted on a large number of TREC news and web collections demonstrate that the proposed scheme almost always outperforms five state-of-the-art retrieval models in a remarkably significant and consistent manner.

In another TF-IDF term weighting study (Al-Talib et al. 2013), a short message service (SMS) model was developed in order to determine the statistical importance of a word for an SMS categorization. First, all SMSs are converted to text documents. After the pre-processing of the SMSs, a vector space model is prepared, and a weight is given to each term. This weighting method predicts the significance of a word for an SMS classification problem. Experiments reported in the article showed that this weighting method significantly improved the classification accuracy measured in many categorization tasks.

TF–IDF was also used in sentiment analysis studies (Das et al. 2018). Das and his team proposed a technique for text sentiment classification using TF-IDF in conjunction with Next Word Negation (NWN). They also compared the performances of a binary bag of words model, the TF-IDF, with TF-IDF with “next word negation” (TF-IDF-NWN) model for text classification. The results showed significant increases in accuracy compared to earlier methods.

In 2018, Qaiser et al. (Qaiser et al. 2018) used TF–IDF for identification of keywords in the document, and at some points pointed out the shortcomings of this method because this method yielded unexpected results. Although the words (“go”, “goes”), (“play”, “playing”), (“mark”, “marking”), (“year”, “years”) in the documents have the same word root, these words were perceived as words that altered the results. The method was also criticized for the word, coexistence, and for the control of the document in terms of dictionary rather than meaning. It has been noted that TD–IDF can be combined with other techniques, such as Naive Bayes, to obtain better results.

As stated in the previous section, the dictionary creation process can be operated in three different ways: (1) automatic; (2) semi-automatic; and (3) manual. In 1993, Riloff conducted a study (Riloff 1993) on automatic dictionary creation. In this work, Riloff developed a system called AutoSlog, a dictionary of domain-specific concepts for extracting information from a text. When AutoSlog is given text to be processed into the system, AutoSlog extracts the user-requested information from this text and creates a dictionary containing the relevant words. If the text supplied to the AutoSlog system is related to the subject of the dictionary, the success of the dictionary will be high. AutoSlog dictionary, a dictionary containing the domain of terrorist events, was then created in five person-hours. The AutoSlog dictionary was then compared to a handmade dictionary that was made by two talented graduate students, requiring approximately 1500 person-hours of effort. As a result, the AutoSlog dictionary provided 98% of the performance of the handmade dictionary.

In 1999, while at Apple Computer, Silverman et al. (Silverman et al. 1999) described the design and structure of the Victorian dictionary created to support speech synthesis research and development. The Victorian dictionary consists of five main chapters: (1) polyphony; (2) prosodic context; (3) repetitive speech; (4) function word sequences; and (5) continuous speech. This dictionary is designed to cover a particular aspect of each of the speech synthesis. The Victorian dictionary is written in United States (American) English. The purpose of the Victorian dictionary is to collect of semantic texts throughout speech. The dictionary was used in statistical estimation of the time and step models for MacinTalk 4, the next generation text-to-speech system from Apple.

Another automatic dictionary creation study (Vorapatratorn et al. 2012) was executed by Vorapatratorn, Suchato, and Punyabukkana. This study describes the method of automatic dictionary creation using a specific phonetic distribution. Usually, the system selects its data by downloading continuous text from the Internet through a web-based browser. The covetous algorithm is then applied to a given text in order to extract the appropriate words, and this process continues until the appropriate text dictionary is completed. The results of the study showed that the amount of data withdrawn from the internet can achieve the target phonetic distribution and create a telephone coverage area of 99.13%. This text dictionary could then be used to efficiently produce the speech dictionary.

Grabar, Claveau, and Dalloux (Grabar et al. 2018) created a semi-automatic corpus in French that provided medical data close to those produced in the clinical context and included descriptions of clinical cases and their discussions. Overall, the corpus currently contains more than 397,000 word-occurrences, excluding punctuation marks. The corpus is currently annotated with several layers of information: (1) linguistic (PoS-tagging, lemmas) and (2) semantic (the Unified Medical Language System [UMLS] concepts, uncertainty, negation, and their scopes). The corpus is continuously enriched as more clinical cases are published.

One example research of automatic dictionary creation was MirasText (Sabeti et al. 2018). MirasText is an automatically generated text corpus for the Persian language. In this study, over 250 Persian web sites were crawled and several fields, such as content,

description, keywords, title, and others were extracted to generate MirasText. The generated corpus contains more than 2.8 million documents and more than 1.4 billion content words. MirasText is the largest Persian text corpus available that can be used for a variety of natural language processing (NLP) applications, such as language modeling, automatic summarization, keyword extraction, and title generation.

In the specific study achieved by Kepuska and Rojanasthien (Kepuska et al. 2011), a data collection system was created to produce speech dictionary from movies, TV series, and DVDs. The dictionary provides a lower cost solution compared with the traditional speech dictionary acquisition method. In addition, it was stated that the collection of data and processing the dictionary was shorter.

Studies on creating a language dictionary using NLP techniques can be applied to different languages. Although studies referring to the English language (Ellen 1993; Silverman et al. 1999) are more numerous, there are also dictionary generation works in other languages, such as Tigrinya (Abera et al. 2018), Thai (Vorapatratorn et al. 2012), French (Grabar et al. 2018), Arabic (Alansary et al. 2014; Alansary et al. 2008), and Turkish (Aktas et al. 2016). In the latest version (Aktas et al. 2016), a detailed WordNet literature research was conducted. As a result of this research, a dictionary containing Turkish informatics terms was created. The words in the dictionary are grouped according to their relationship, similarity, and class. It was stated that the resulting dictionary was larger and more comprehensive than other information dictionaries created up to date.

One of the latest works was completed by Vijay and his team (Vijay et al. 2018). They created a Hindi-English code-mixed dictionary using tweets published online for the last eight years. In order to create this dictionary, tweets were taken from Twitter, using the Twitter Python API, which primarily uses the advanced search option in Twitter. All information, such as received tweets, timestamp, Uniform Resource Locator (URL), text, user, retweets, answers, full name, identities, and likes were converted to the JavaScript Object Notation (JSON) format. A comprehensive semi-automated procedure was performed to remove all noisy tweets. After completion of these steps, a dictionary of 2866 words was created, and dictionary words were classified as happiness, sadness, anger, surprise, hate, and multiple emotions. Emotional analysis was performed for the tweets published online, and a 58.2% accuracy was obtained.

3 Method

In order to identify the meaningful words of the documents given to the system, the documents are pre-processed initially. After meaningful terms are determined, they are now candidates for the dictionary and web search. The general system schema is given by Fig. 1.

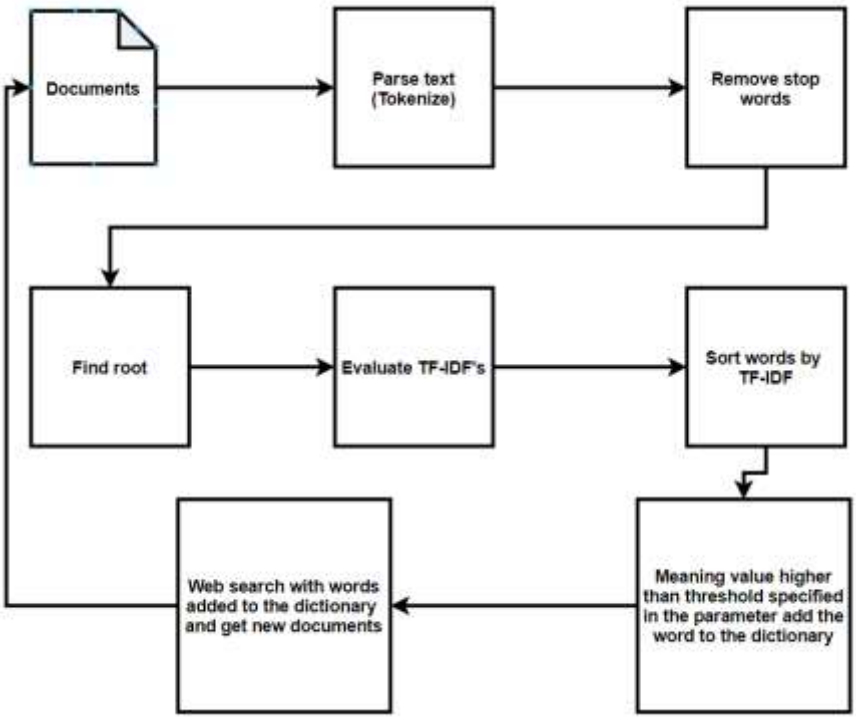


Fig. 1. General structure of automatic dictionary creation

3.1 Pre-processing

Text pre-processing is one of the most important preliminary step of NLP and is often troublesome but affects the success of the algorithm. In pre-processing, it is necessary to remove prepositions, conjunctions, exclamations, letters, and words with no categorical meaning. These words, called stop words, must be filtered in the first step of text pre-processing. Text pre-processing is applied to each document, and it is used for noise removal, lexicon normalization, and object standardization.

After this process, the document words are converted to lower-case format. It is then necessary to separate the words from the suffixes and find their root forms. The main purpose of stemming is to reduce different grammatical forms/word forms of a selected word, such as its noun, adjective, verb, or adverb, to its root form. The most common rule-based stemming algorithm for the English language was developed by Porter, and it is still most commonly used method to date. The most commonly known and easy-to-use Porter Stemmer algorithm (Porter 2001) was used in this project. Examples successfully applied to stemming are given by Jivani (2011). Fig. 2 lists the stemming algorithms in NLP.

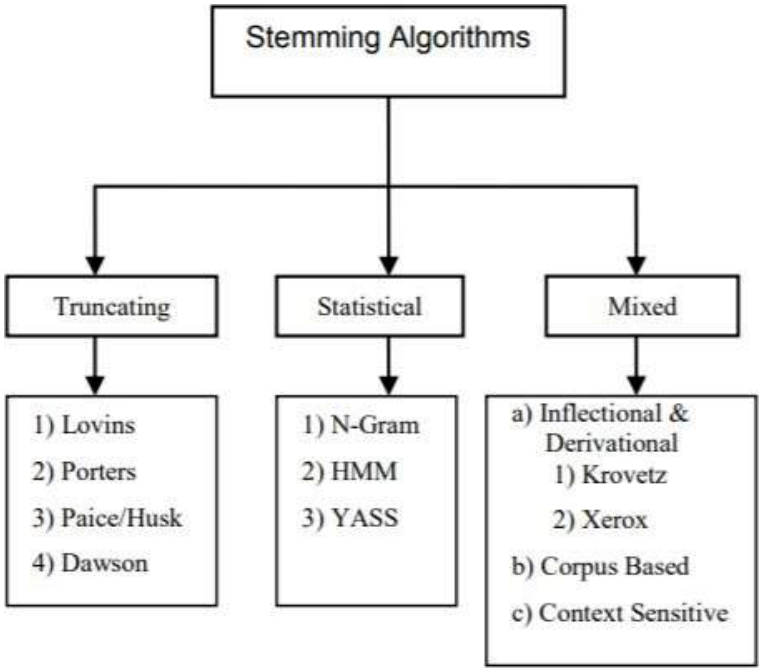


Fig. 2. Types of stemming algorithms

3.2 Selection of Keywords

At this stage, the main system task is to find meaningful words from both the initial given documents and the documents returned from the web search. In this study, the TF–IDF was applied to find the meaning values of words. TF–IDF consists of combination of two different concepts: (1) term frequency and (2) inverse document frequency. The TF–IDF is commonly used to weight each word in the text document according to its level of uniqueness. In other words, the TF–IDF captures the relationship between words, text documents, and specific categories (Deepshikha et al. 2011).

Suppose we have a document, "T1", containing 5000 words and the word "alpha" is present in the document exactly 10 times. It is a very well-known fact that the total length of documents can vary from very small to large, so it is highly possible that a term may occur more frequently in large documents in comparison to small documents. Thus, in order to normalize document length, the occurrence of any term in a document is divided by the total terms present in that document to yield the term's frequency. So in this case, the term frequency of the word "alpha" in the document "T1" is

$$TF = 10/5000 = 0.002$$

Inverse document frequency (IDF) assigns less weight to common words and more weight to less common words. For example, if we have 10 documents and the term "technology", is available in five of these documents, the IDF value is calculated:

$IDF = \log_e (10/5) = 0.3010$

The TF–IDF calculation is equal to the product of the TF and IDF values:

TF-IDF = 0.002 * 0.3010 = 0.000602

First, all word frequencies in the document set are recorded in the database. Using the formulas mentioned above, TF, IDF and TF–IDF values of each word are then calculated. An example of the metric values for the given document set is shown in Table 1. After calculating these values, the words with a meaning value above the threshold are added to the dictionary. The resulting dictionary hash similarity values obtained by adding different threshold meaning values are presented in detail in section 4.

Table 1. Term-Frequency–Inverse Document Frequency (TF–IDF) values of words

DocumentId	Word	TF	IDF	TFIDF
1	season	0.0422	1	0.0422
1	yard	0.0578	0.6931	0.0400
1	pass	0.0533	0.6931	0.0370
1	player	0.0222	1	0.0222
1	game	0.0200	0.6931	0.0139
1	touchdown	0.0200	0.6931	0.0139
1	football	0.0133	1	0.0133

3.3 Creating the First Dictionary Words from The Starting Document

When the document related to the subject of the desired dictionary is initially given to the system, the meaningful words of these seed document are determined using TF–IDF. Only the highest TF–IDF value is added to the dictionary.

The web search is then started with this word. However, the rule of determining meaningful candidate words for insertion into the dictionary through the web search is not only the most meaningful word but also the words in which the TF–IDF values are above 0.03.

3.4 Determination of Similarity Value

When candidate meaningful words are selected after a web search, a similarity value using WordNet similarity is then calculated for each word with respect to the dictionary.

Finally, the words with similarities above the similarity threshold value are added to the dictionary. For example, after finding meaningful words for web search results, suppose that three words with the highest TF–IDF values will be added to the dictionary. Suppose these three words are “student", "math", and "lesson". Let the words in the dictionary be “team” and “ball”. A loop is then used to calculate the WordNet similarity of each meaningful word with the words in the dictionary. If we take the word “football”:

WordNet Similarity (“football”, “team”) = 0.35
WordNet Similarity (“football”, “ball”) = 0.30
Total WordNet Similarity (“football”) = 0.35 + 0.30 = 0.65 is the result.

The average WordNet similarity value of the word is then calculated using formula 1.

$$\text{Average WordNet Similarity} = \frac{\text{Total WordNet Similarity}}{\text{Dictionary Word Count}} \quad (1)$$

According to formula 1,

Average WordNet Similarity = 0.65 / 2 = 0.325 is then calculated.

If the average WordNet similarity parameter value is defined as 0.35 in the application, then the word is not added to the dictionary; on the other hand, if the similarity parameter value is 0.30, the word is added to the dictionary.

WordNet is the English dictionary database in the Cognitive Science Laboratory of Princeton University. Names, verbs, adjectives, and adverbs are grouped into sets of cognitive synonyms (synsets), each of which expresses a separate concept. The hypernym feature, a more abstract and general meaning of a word, is used in this database. For example, by looking for the word hypernym for the word cat in the WordNet database, the concept of this word can be determined as animal. WordNet is free and can be downloaded publicly. The structure of WordNet is therefore a useful tool for computational linguistics and NLP. Fig. 3 shows the WordNet hierarchy (Wei et al. 2015).

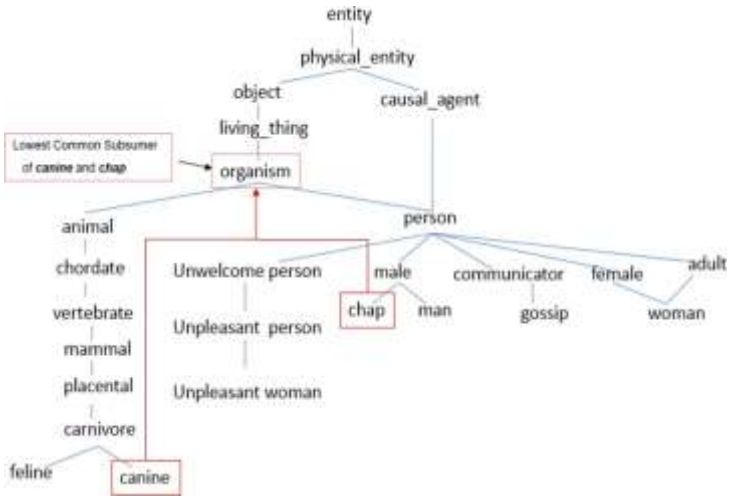


Fig. 3. WordNet hierarchy sample

3.5 Web Search Process

The most meaningful word inserted into dictionary for the last iteration is used in the next web search together with the synonym words obtained from WordNet. For example, if the word added to the dictionary is “student”, the synonym for “student” is “pupil”. In the web search, “student” and “pupil” are used together.

The Azure Cognitive Web Search Interface is used to search on the web. The words to be used on the web search are combined with the space character and given as a parameter to the Azure Web Search Service. For example, the words “Ronaldinho”, “Barcelona”, and “el classic” are listed consecutively, and these words are used together as “ronaldinho barcelona el classico” in the web search. The web search with these words will result in more than one document. Among these documents, a single document that has not been previously processed is selected. Pre-processing steps are then applied to this document, and new meaningful words are obtained. Thus, the dictionary is constantly expanding with new documents, and the dictionary can be prevented from repeating itself. This process continues until the dictionary reaches the number of dictionary words in which a parameter value is initially given by the user to the system. Thus, a continuous cycle is provided in the system.

3.6 Determination of General Similarity Value of Dictionary

In this step, the similarity of the dictionary is determined using the SimHash algorithm. The SimHash algorithm is an algorithm used to find similarity between files or web sites in applications, such as a search engine, especially in text processing. The SimHash algorithm views two files as vectors and tries to find the cosine link between these vectors.

SimHash is a hash function, and the more similar its property is to the text input, the smaller the Hamming distance of the hash values (Hamming distance is the number of locations in which the corresponding symbols are different). The algorithm works by dividing the text into pieces and combining each piece with a selected function. Each mixed set is represented as a binary vector, and the bit value is converted to +1 or –1 depending on whether the bit value is 1 or 0. In order to obtain SimHash, all bit vectors as bitwise are collected. Finally, if the total is negative, the bits that result as 1 are otherwise are set to 0.

The SimHash algorithm has been used in many different studies for the purpose of similarity detection. For example, in studies by Jiang and Pi (Jiang et. al 2011; Pi et al. 2009), the SimHash algorithm is used to obtain document similarity. Fig. 4 illustrates the working procedure of the SimHash algorithm (Ho et al. 2014).

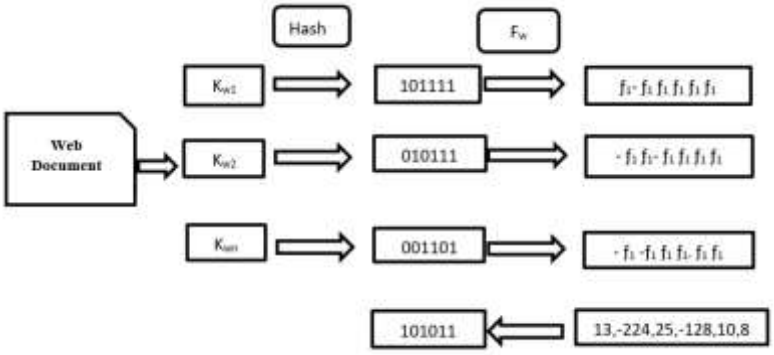


Fig. 4. Working procedure of the SimHash algorithm

4 Experiments

For all dictionaries, documents belonging to the sports data and containing HTML content used in the study (Turan et al. 2018) were given to the system as a starting point. First, the number of dictionary words was set at 25, 50, and 100. The effect of the meaning value of a word calculated by the TF–IDF was then evaluated for candidate word selection for dictionary. Finally, the effect of the number of documents given initially was used in the experiment. For each experiment, the dictionary similarity values were calculated and discussed.

4.1 Experiment I

In the Table 2 below, the hash similarity value for 25 words limited dictionary initialized by only 1 sport (badminton) document is given.

Table 2. Hash similarity value for the 25 words limited dictionary initialized by only one sports document

Parameter Name	Parameter Value
Dictionary Similarity Threshold of the Word	0.03
Dictionary Maximum Number of Words	25
Number of Starting Documents	1
Dictionary Similarity Value %	**75.2**

The similarity value of the 25 words dictionary was shown to be higher than the dictionaries with 50 and 100 words. The most important reason for achieving this similarity value was that the new documents obtained from the web causes deviation from the original reference document topic. Fig. 5 lists the words obtained for parameters given in the Table 2.

academy	badminton	bwf	committe	court
debut	entry	event	exercise	federation
game	injury	medal	meter	olympic
people	player	racket	sear	select
set	shoe	shop	sport	world

Fig. 5. Twenty-five words limited dictionary initialized with only one sports document

In Table 3, the hash similarity value for the 50 words limited dictionary initialized by only 1 sport (badminton) document is given.

Table 3. Hash similarity value for the 50 words limited dictionary initialized by only one sports document

Parameter Name	Parameter Value
Dictionary Similarity Threshold of the Word	0.03
Dictionary Maximum Number of Words	50
Number of Starting Documents	1
Dictionary Similarity Value %	**73.8**

The similarity value of the 50 words dictionary was higher than the dictionary with 100 words. Obviously, when all parameters were kept constant, and only the number of dictionary words was increased, the dictionary similarity value decreased. Fig. 6 lists the words obtained for the parameters given in Table 3.

academy	arcadia	badminton	blais	box
brand	bwf	committe	company	cooki
country	court	cutter	dark	debut
dimense	enjoy	entry	equip	event
excel	feder	game	injury	league
ligament	medal	member	meter	olympic
player	policy	postage	racket	school
sear	select	service	set	ship
shoe	shop	sport	techniqu	tenni
weight	wayn	world	wrist	yonex

Fig. 6. Fifty words limited dictionary initialized with only one sports document

In Table 4, the hash similarity value for the 100 words limited dictionary initialized by only one sport (badminton) document is given.

Table 4. Hash similarity value for the 100 words limited dictionary initialized by only one sports document

Parameter Name	Parameter Value
Dictionary Similarity Threshold of the Word	0.03
Dictionary Maximum Number of Words	100
Number of Starting Documents	1
Dictionary Similarity Value %	**68.7**

The similarity value of the 100 words dictionary was lower than the similarity values of the 25 and 50 words limited dictionaries. Using this result, we could make an inference: When the maximum number of words in the dictionary increases, deviations oc-

cur in the dictionary after a certain number of words. The factor that causes the deviation originates from the web search that was done with this word after adding an unmeaningful word to the dictionary. In order to prevent deviations in the dictionary, the dictionary should be checked periodically, and words relating to this document should be added to the dictionary based on the initial reference document. Fig. 7 lists the words obtained for the parameters given in Table 4.

academi	address	alberta	answer	associ
backhand	badminton	ball	basic	bird
box	brand	build	bwf	canada
card	championship	club	coach	committe
company	cooki	country	court	cutter
dan	debut	double	draw	entry
equip	event	excel	experiment	fact
fast-paced	fat	federation	format	game
group	guest	har	hotel	image
Inform	ingen	injury	job	knee
language	league	ligament	lin	match
medal	member	merge	meter	ntrue
och	olymp	oppon	pain	partner
player	point	policy	procedure	program
question	racket	read	sear	season
select	service	set	shoe	shop
shuttlecock	sign	singapore	skill	sport
team	technique	thesen	tournament	use
vatn	video	way	weight	wib
world	wrist	year	yonex	zone

Fig. 7. One-hundred words limited dictionary initialized with only one sports document

4.2 Experiment II

In this experiment, the effect of the meaning threshold value of a word was examined. If the meaning threshold value was selected to be 0.03, words with meaning values greater than this meaning threshold value became candidates for the dictionary. If this threshold was applied through selection of dictionary words, higher dictionary similarity values were obtained. On the other hand, when words with a meaning value less than 0.03 are added to the dictionary, lower dictionary similarity values are obtained. In the following experiments, the threshold value for the TF–IDF was selected 0.02, and 25 and 50 words limited dictionaries were created in order.

In Table 5 below, the hash similarity value for the 25 words limited dictionary initialized with only one sports (badminton) document is given.

Table 5. Hash similarity value for the 25 words limited dictionary initialized with only one sports document

Parameter Name	Parameter Value
Dictionary Similarity Threshold of the Word	0.02
Dictionary Maximum Number of Words	25
Number of Starting Documents	1
Dictionary Similarity Value %	**62.5**

One sports document (badminton) was again given to the system, and as a result of the calculation with the TF–IDF, a 25 words limited dictionary was produced using the words with a meaning value equal or lower than 0.02. The overall similarity value of final dictionary was calculated using the hash similarity method, and a 62.5% similarity was obtained. This similarity value was lower than the similarity value of all of the dictionaries obtained by adding the words with the meaning value over 0.03. It was noted that addition of words lower than the meaningful threshold value to the dictionary caused deviations and decreased the similarity of dictionary words. Fig. 8 includes dictionary words obtained in this experiment.

badminton	ball	bwf	city	coach
committe	court	day	dimens	exercise
feder	game	languag	meter	nn
olympic	player	point	quarterfin	sign
sport	state	street	tennis	world

Fig. 8. Twenty-five words limited dictionary initialized with only one sports document

In the Table 6, the hash similarity value for the 50 words limited dictionary initialized by only one sports (badminton) document is given.

Table 6. Hash similarity value for the 50 words limited dictionary initialized with only one sports document

Parameter Name	Parameter Value
Dictionary Similarity Threshold of the Word	0.02
Dictionary Maximum Number of Words	50
Number of Starting Documents	1
Dictionary Similarity Value %	**59.2**

The overall similarity of this dictionary was calculated using the Hash similarity method and was equal to 59.2%. This similarity value is lower than the expected similarity value of the 25 words limited dictionary described in Experiment I. Fig. 9 includes the dictionary words obtained in this experiment.

academy	article	assist	associ	badminton
book	box	bwf	champion	championship
committe	competition	cooki	court	cup
cutter	descript	draw	editor	entry
event	feder	feedback	game	group
klcc	list	medal	member	meter
nn	olympic	play	player	point
policy	racket	read	select	service
set	shoe	side	singapor	sport
team	tennis	tournament	world	year

Fig. 9. Fifty words limited dictionary initialized with only one sports document

4.3 Experiment III

In this experiment, the effect of the number of documents fed to the system initially was studied. In the following experiments, two different dictionaries of 25 words were created by giving two and three initial documents to the system.

In the Table 7, the hash similarity value for the 25 words limited dictionary initialized by two sports (badminton)-related documents are given.

Table 7. Hash similarity value for the 25 words limited dictionary initialized with two sport documents

Parameter Name	Parameter Value
Dictionary Similarity Threshold of the Word	0.03
Dictionary Maximum Number of Words	25
Number of Starting Documents	2
Dictionary Similarity Value %	**62.7**

When all parameters were kept constant, however the number of documents initially fed to the system were changed, the similarity value of the dictionary for a small number of initial documents then becomes higher. This finding can be explained by deviations in the dictionary words based on the initial given documents. Fig. 10 lists the words obtained for the parameters given in the Table 7.

announce	associate	badminton	bwf	canada
championship	contact	court	dimense	double
entry	event	game	inform	line
person	rule	service	side	sport
team	ticket	video	yonex	zone

Fig. 10. Twenty-five words limited dictionary initialized with two sport documents

In Table 8, the hash similarity value for the 25 words limited dictionary initialized by three sports (badminton) documents is given.

Table 8. Hash similarity value for the 25 words limited dictionary initialized by three sports documents

Parameter Name	Parameter Value
Dictionary Similarity Threshold of the Word	0.03
Dictionary Maximum Number of Words	25
Number of Starting Documents	3
Dictionary Similarity Value %	**60.3**

The similarity value of the 25 words limited dictionary created with three initial documents was the lowest dictionary similarity value, including almost the dictionaries shown in experiments 1 and 2. It should be noted that even if the TF–IDF meaning value is set to 0.03, dictionaries with a high success rate cannot be obtained when too many initial documents are given. Fig. 11 lists the words obtained for the parameters given in Table 8.

associate	august	badminton	championship	change
contact	court	date	dog	dont
entry	event	game	hour	inform
partner	player	point	score	service
side	time	world	year	weight

Fig. 11. Twenty-five words limited dictionary initialized with three sports documents

5 Results

In this study, meaningful words of the seed documents were calculated using the TF–IDF, and a specific domain dictionary was automatically created. Through this study, the effect of establishing a similarity threshold between a candidate word to be inserted into dictionary and the dictionary, dictionary size, and the effect of initial number of documents were researched. As a result, we can conclude three findings from these experiments:

(1) The first one indicates that when the dictionary size grows, the similarity of dictionary then decreases intuitively as expected. This can be explained by less deviation in the dictionary because the number of dictionary words was small;

(2) The second one is to establish bigger threshold for the meaningful value of a word in order to add to the dictionary enhances the dictionary similarity. This is a result of selecting more meaningful (similar) words; and

(3) The third one indicates that when the number of initial documents given increases, dictionary similarity significantly decreases as a result of the differences of the contents of the given documents.

The work was compared with similar studies, and it is obvious that the results of the study are quite successful in terms of average results, dictionary creation, and growth rate. The study (Vijay et al. 2018) mentioned in the literature review section achieved successful results in automatic dictionary creation in two studies. However, unlike Vijay et al. (2018), the web search section was included in this work. This work is more valuable than the study of Vijay et al. (2018) in terms of working on instant and current data. Moreover, the WordNet similarity calculation is used in this work in order to determine whether the words to be added to the dictionary are related to the subject to be created. In general, the following conclusions can be drawn from the results of the work.

- The set of documents should be related to the dictionary you want to create. If the system is presented with a set of documents that are not associated with the desired dictionary, the resulting dictionary will contain meaningless words.
- Once meaningful words are found, the words above dictionary similarity threshold value are added to the dictionary. When similarity threshold value is increased, the speed of dictionary creation decreases, while Hash similarity of the final dictionary increases.
- Web search is made with meaningful words added to the dictionary. When searching the web with a certain percentage of these words, not all the meaningful words, the success rate will change positively.
- Querying with the last words added to the dictionary for processing with different data continuously during the web search process will prevent duplication of the dictionary. It is therefore useful to mark the last words in the dictionary.

6 Future Works

The following studies, which are thought to contribute to the literature, will be discussed.

- Since the suggested system operates in a continuous growth manner, the number of iterations in which the system approaches saturation can be determined.
- A WordNet similarity method was used to calculate the similarity value between the words in the dictionary. According to the WordNet similarity result, it is decided whether or not to add the word to the dictionary. Different similarity techniques could also be examined. For example, Word2vec similarity method and WordNet similarity method could be compared. Structural differences between the two dictionaries could then be revealed.
- Once it has been determined to add a word into the dictionary, all synonyms of this word could be added to the dictionary together. SimHash algorithm could be calculated by comparing the results.

References

1. Abera, H., Mariam, S. (2018). Design of a Tigrinya Language Speech Corpus for Speech Recognition. Association for Computational Linguistics,78–82.
2. Aktas, Y., Ince, E.Y., Cakır, A., Kutlu, A. (2016). Wordnet ve Bilgisayar Ag Terimleri Sözlügünün Olusturulması. Akademik Bilişim 2016, 7(2), 1-9.
3. Alansary, S., & Nagi, M. (2014). The International Corpus of Arabic: Compilation, Analysis and Evaluation. Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP), https://www.doi.org/10.3115/v1/W14-3602
4. Alansary, S., Nagi, M., Adly, N., Alexandrina, B. (2008). Building an International Corpus of Arabic (ICA): Progress of Compilation Stage. Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP), https://www.doi.org/10.3115/v1/W14-3602
5. Al-Talib, G., A., & Hassan, H.S. (2013). A Study on Analysis of SMS Classification Using TF-IDF Weighting. International Journal of Computer Networks and Communications Security, 1(5), 189-194.
6. Das, B., & Chakraborty, S. (2018). An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation. ArXiv, abs/1806.06407.
7. Ellen, R. (1993). Automatically Constructing a Dictionary for Information Extraction Tasks. Proceedings of the Eleventh National Conference on Artificial Intelligence, 811–816.
8. Gholamrezazadeh, S., Salehi, M., Gholamzadeh, B. (2010). A Comprehensive Survey on Text Summarization Systems. 2009 2nd International Conference on Computer Science and its Applications, https://www.doi.org/10.1109/CSA.2009.5404226
9. Grabar, N., Claveau, V., Dalloux, C. (2018). CAS: French Corpus with Clinical Cases. LOUHI 2018 - The Ninth International Workshop on Health Text Mining and Information Analysis, https://www.doi.org/10.18653/v1/W18-5614
10. Habacha, A., Kboubi, F., Ahmed, M. (2011). Thematic Analysis and Visualization of Textual Corpus. International Journal of Computer Science and Engineering Survey, https://doi.org/10.5121/ijcses.2011.2402
11. Ho, P., Kim, S. (2014). Fingerprint-Based Near-Duplicate Document Detection with Applications to SNS Spam Detection. International Journal of Distributed Sensor Networks, https://www.doi.org/10.1155/2014/612970
12. Hovy, E., & Lin, C. (2001). Automated text summarization and the SUMMARIST system. Proceedings of a Workshop on Held at Baltimore, https://www.doi.org/10.3115/1119089.1119121
13. Indyk, P., & Motwani, R. (1998). Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality. Conference Proceedings of the Annual ACM Symposium on Theory of Computing, https://www.doi.org/10.1145/276698.276876
14. Jiang, Q., & Sun, M. (2011). Semi-Supervised SimHash for Efficient Document Similarity Search. ACL, 12(1), 93-101.
15. Jivani, A.G. (2011). A Comparative Study of Stemming Algorithms. Int. J. Comp. Tech. Appl, 2 (6), 1930-1938.
16. Këpuska, V. Z., Rojanasthien, P. (2011). Speech Corpus Generation from DVDs of Movies and TV Series. Journal of International Technology and Information Management, 20(1) 49-82.
17. Koeva, S., Stoyanova, I., Todorova, M., Leseva, S. (2016). Semi-automatic Compilation of the Dictionary of Bulgarian Multiword Expressions. Proceedings of GLOBALEX 2016: Lexicographic Resources for Human Language Technology, 86-95.

18. Lee, K. M. (2012). Locality-Sensitive Hashing Techniques for Nearest Neighbor Search. International Journal of Fuzzy Logic and Intelligent Systems, https://www.doi.org/10.5391/IJFIS.2012.12.4.300
19. Nguyen, Thien., & Shirai, Kiyoaki. (2013). Text Classification of Technical Papers Based on Text Segmentation. International Conference on Application of Natural Language to Information System NLDB 2013, https://doi.org/10.1007/978-3-642-38824-8_25
20. Paik, J.H. (2013). A novel TF-IDF weighting scheme for effective ranking. Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, https://www.doi.org/10.1145/2484028.2484070
21. Patel, D., & Bhavnagar, M. (2011). Mobile SMS Classification. International Journal of Soft Computing and Engineering (IJSCE), 1(1), 47-49.
22. Pi, B., Fu, S., Wang, W., Han, S. (2009). SimHash-based Effective and Efficient Detecting of Near-Duplicate Short Messages. Proceedings of the Second Symposium International Computer Science and Computational Technology (ISCSCT '09), 20-25.
23. Porter, M. F. (2001). Snowball: A language for stemming algorithms. Published online. http://snowball.tartarus.org/texts/introduction.html last accessed 11 March 2008.
24. Qaiser, S., & Ali, R. (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. International Journal of Computer Applications, https://doi.org/10.5120/ijca2018917395
25. Sabeti, B., Abedi, H., Choobbasti, A.J., Najafabadi, M., Vaheb, A. (2018). MirasText An Automatically Generated Text Corpus for Persian. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018),1174-1177.
26. Silverman, K.E., Anderson, V., Bellegarda, J.R., Lenzo, K.A., Naik, D. Design and collection of a corpus of polyphones and prosodic contexts for speech synthesis research and development. Sixth European Conference on Speech Communication and Technology (EUROSPEECH 1999), 2707-2708.
27. Turan, M., Ogtelik, S. (2018). Ingilizce Dokumanlarda Tema ve Alt Kavramlar Tespit Modeli. Düzce Üniversitesi Bilim ve Teknoloji Dergisi, https://www.doi.org/10.29130/dubited.420104
28. Vijay, D., Bohra, A., Singh, V., Akhtar, S.S., Shrivastava, M. (2018). Corpus Creation and Emotion Prediction for Hindi-English Code-Mixed Social Media Text. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop, https://www.doi.org/10.18653/v1/N18-4018
29. Vorapatratorn, S., Suchato, A., Punyabukkana, P. (2012). Automatic online text selection for constructing text corpus with custom phonetic distribution. Ninth International Conference on Computer Science and Software Engineering (JCSSE), https://www.doi.org/10.1109/JCSSE.2012.6261916
30. Wei, T., & Chang, H. (2015). Measuring Word Semantic Relatedness Using WordNet-Based Approach. Journal of Computers, https://www.doi.org/10.17706/jcp.10.4.252-259

Convolutional Autoencoder Model for Reproducing Fingerprint

Şafak Kayıkçı[0000-0002-3325-4731]

Bolu Abant Izzet Baysal University, Computer Engineering Department, Bolu, Turkey
safak.kayikci@ibu.edu.tr

Abstract. Fingerprint recognition is usually based on comparing the feature points in the fingerprint and their parameters. Therefore, it is important for fingerprint recognition systems to clean and improve the fingerprint image, to correctly identify the feature points and parameters to be used in fingerprint recognition systems, and to perform the comparison process correctly. Fingerprint technology was never foreseen to be used to unlock mobile phones and authenticate payments, or to arrest criminals. Autocoders are neural networks that usually obtain a lower dimensional representation of the data and try to produce the same data as output using this representation. With this feature, the training of self-coding, which is an example of demonstration learning, is realized through unattended learning. In this paper, convolutional auto encoders are used to reconstruct latent or damaged fingerprints is proposed.

Keywords: Autoencoder, Convolutional Neural Networks, Fingerprint

1 Introduction

Biometrics is the identification of individuals using behavioral or physiological features. Because they use behavioral or physiological characteristics that are different in each person, they are more reliable than traditional knowledge-based and key-based approaches in distinguishing empowered individuals and fraudsters. Besides, they require the person to be identified to be at the point of identification. The biometric meets the security requirements needed in the electronic communication environment and will be the dominant automated person recognition method soon. A biometric system is essentially a pattern recognition system that makes personal identification using the physiological or behavioral characteristics of users [1]. Logically, it can be divided into two units, the recording unit, and the identification unit.

The registration unit does the job of registering different people to the biometric system. During the registration process, a biometric characteristic of a person is read through a biometric reader, and a raw digital equivalent of this characteristic is obtained. To facilitate the matching process, the raw digital response is processed by the feature extractor to obtain a small size but meaningful template containing the features of this digital equivalent. According to the application, the template is saved on a magnetic card or smart card, which will be given to a central database or person.

The identification unit comes into play at the stage of granting access to persons. In this process, the raw digital equivalents of the biometric characteristics taken from individuals with biometric readers are processed by the feature extractor and the response obtained in the previous stage is created. This response is given to the feature mapper. The feature mapper compares and identifies the templates it obtains from the template database. An identification system is essentially a database access system. In addition to the level of security applied to the processed identity entry, two other measures of accuracy should be provided, such as the sensitivity and recall that characterize the access accuracy of the database access system, to determine the adequacy of the system. Sensitivity can be defined as the ratio of the number of real records from templates called by the identification system from the database to the total number of templates called from the database. Recall, on the other hand, is the ratio of the number of real records from the templates called from the database by the identification system to the real records in the database. In addition to accuracy, verification/identification speed is another important measure of performance. There is an only one-to-one comparison in a verification system, and the speed performance is related to the response time of the verification (and feature extraction) algorithm, that is, more precisely, the computational complexity of the algorithm used. It is generally easy to ensure that a verification system is at the desired speed. But in an identification system, especially in systems that use a database of millions of templates, many comparisons need to be made to identify an individual. Factors such as speed performance, response time, throughput, computational complexity, scalability affect.

2 Fingerprint Recognition Systems

Fingerprint recognition is usually based on comparing the feature points in the fingerprint and their parameters. Therefore, it is important for fingerprint recognition systems to clean and improve the fingerprint image, to correctly identify the feature points and parameters to be used in fingerprint recognition systems, and to perform the comparison process correctly. Studies on the detection of fingerprints were first performed by Galton and Henry [2]. These features, called Galton characteristics, are called Endpoints, Surrounded Points, Fork Points, and Islets. To achieve successful results, it is enough to use tip and fork points. The terminologies used for fingerprints are explained below.

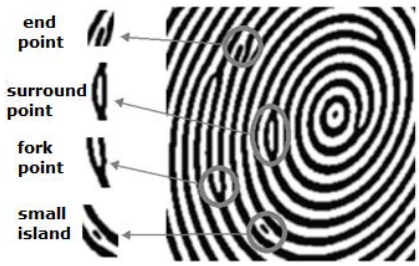


Fig. 1. Characteristic Points of Galton

2.1 Line, endpoint and fork point

The line is a single circular line on the fingerprint. The sum of the line lines reveals the fingerprint pattern. End and fork dots used in fingerprint recognition systems are formed by sudden termination and bifurcation of line lines.

2.2 Direction maps

Direction maps show the direction of the line lines. When creating a direction map, the picture is split into small pieces. Local routing is applied on each track. The unit perpendicular to the slope determines the orientation on the vector part. The average of all points in the part is calculated. The resulting vector determines the direction for this part. The direction map is a low-resolution general representation of the fingerprint. Direction map can be used for the classification of fingerprints, finding reference points, identification with the artificial neural network, and fingerprint recognition systems.

2.3 Delta and midpoints

When categorizing fingerprints, only part of the fingerprint's design area is dealt with. The pattern area of a fingerprint is the area surrounded by two lines called the innermost type line, which come close together and separate from each other to look for the central part of that fingerprint. The pattern areas of the fingerprints in a loop or spiral type include two kinds of singular points.

We can define the delta point, also called the outer stop, as the point of the line closest to the center of the separation points of the type lines, opposite this center. This point may be a point line, a short line, a bifurcation point of a bifurcating line, a line end or the line that extends between two types of lines, directly opposite the separation point. The middle point (navel), also called the inner stop, can be defined as the special point on the innermost curved lines. Due to the wide variety of curved lines, the rules for determining the navel are quite complex.

To compare the two fingerprints, the coordinates of the feature points, the number of line lines passing through them and the number of line lines passing between them, and if the size of the fingerprint image does not change, such as the distance between the reference points and the reference points to each other, the local direction and type of each feature point many parameters can be used. Another important concept in fingerprint classification and fingerprint matching is the number of line lines cut by an imaginary line drawn between the delta and the midpoint (navel point). Due to the complexity of the line shapes, it is difficult to make an exact definition of the number of line lines. The number of line lines can be briefly defined as the number of lines cut by an imaginary line drawn between two feature points.

2.4 Feature Point Extraction

Feature point extraction is to obtain characteristics that represent a fingerprint named feature point from fingerprint images taken as input. In automatic fingerprint mapping, the input fingerprint images must be represented remarkably and conveniently.

A good feature point extraction algorithm will be both reliable and effective. Reliability means that the feature point extraction algorithm does not extract false feature points, does not miss existing feature points, and is accurate in calculating the location and orientation of the feature point [3]. Extracting reliable feature points from fingerprint images is a difficult process. If the fingerprint images are of good quality, the lines and valleys flowing locally in the fixed directions of the fingerprint are well determined and can be easily separated from each other. In such cases, the line terminations and line bifurcations, which are the irregular places of the lines, can be easily determined and their positions can be removed from the binary line images.

Depending on its quality, a poor fingerprint image can be rejected or strengthened before removing features. A very weak fingerprint image, in which the line structures are completely broken, should be rejected, and a weak fingerprint image that is at a level that the line structures are visible should be strengthened before removing the feature. A good feature extraction algorithm should be able to ignore distorted line structures to a certain extent. Fingerprint correction and preprocessing algorithms occupy an important place in fingerprint recognition systems. The removal of unnecessary parts of the fingerprint image from noise and noise provides some advantages in finding feature points. Thanks to the high-quality improvement of the image, other processing steps use this improved image as input.

Although there are many features used in fingerprint recognition, the two most common features are endpoints and forks. These feature points are formed by cutting the normal flow direction of a line suddenly (endpoint) or changing it in two (fork point). In identification, the types of these feature points, their coordinates, angles, distances, and the number of line lines passing between them are used. It is difficult to get feature points accurately from some low-quality fingerprint images due to fingerprints, scars, and noise. While there are feature points in a fingerprint, the first thing to decide is whether a pixel belongs to a line. For this, noise reduction techniques are used. Then, thresholding and thinning processes are applied. Once the image's refining and enhancement are complete, the image is ready for the discovery of feature points (Fork and endpoint). Each feature point is defined by its coordinate information and angle.

3 Autoencoders

Autoencoder networks feed several hidden layers of neural networks forward. Those networks attempt to restructure the output layer input data. The target data in the output layer is identical to the input data so the output layer size is the same as the input layer. It is trained using the gradient descend method in Autoencoder, just like the backpropagation algorithm. Since the size of the hidden layer is smaller than the size of the input data in an autoencoder, the size of the input data is resized to a smaller region, limiting it to the hidden layer. Data from the hidden layer is then converted back to the original data in the output layer. It can render bidirectional mapping between the area of data and area of code. Using large numbers of hidden layers in an model, input data of large dimensions can be restricted to a much less code area.

However, training a neural network with more than one hidden layer can be quite annoying and also has difficulty in achieving good results. This is because hidden layers at very low levels are very difficult to optimize [4].

Autoencoder is built with feedforward, this neural network may have one or more secret layers. It tries to recreate input values within the output layer of the data. The key difference between autoencoder and traditional neural artificial networks is the output layer thickness. For an autoencoder the output layer size is the same as the input layer size. However, it is still less than the number of input layers and the number of output layers irrespective of the number of hidden layers in an Autoencoder.

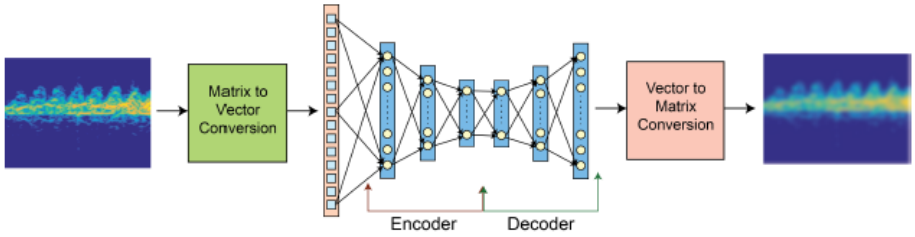


Fig. 2. Diagram showing the architectural structure of Autoencoder.

An autoencoder consists of two sections called encoder and decoder. While the "Encoder" part reduces the multi-dimensional input information to small dimensions, the "Decoder" part returns the data reduced to small dimensions by the other part to its original structure. Autoencoder uses sigmoid activation functions for nonlinear mapping. Autoencoder can behave similarly to principal component analysis when using the sigmoid activation functions [5]. Networks can be equipped to minimize the difference in error between original data and reconstructed data on average. By applying the Chain Rule to backpropagate error derivatives, you can easily obtain the desired gradient value, first going through the decoder and then the encoder pieces. The encoder calculates the nonlinear mapping of the input for a given input vector X as follows.

$$E = \sigma (WX + b) \tag{1}$$

Here, s represents the nonlinear activation function, W network weights, and b represents the constant term. Attributes learned by the encoder then go through the following process to rebuild input X by a decoder

$$Z = \sigma (\widetilde{W}E + \tilde{b}) \tag{2}$$

Here $\widetilde{W}$ and $\tilde{b}$ represent the weight and constant term of the decoder. In unattended pre-training, it tries to minimize the following cost function for each x_i value by setting the network weight and constant term $\theta = [W, b, \widetilde{W}, \tilde{b}\ ]$.

$$J(\theta) = \frac{1}{N}\sum_{i=1}^{N}(x_i - z_i)^2 \tag{3}$$

Here, various restrictions must be added to Equation 3 and the network itself to prevent the network from learning the unit matrix directly. The first is to keep the number of neurons in the hidden layer less than the size of the vector given as input. This will force the network to recreate the input in the upper space by nonlinear operations in a low dimensional space. Besides, a rarity parameter must be added to Equation 3. This parameter forces the network to learn the correlations between the given input vectors. After adding the Rarity parameter, Equation 3 becomes the following.

$$argmin_{\theta}\, J(\theta) = \frac{1}{N}\sum_{i=1}^{N}(x_i - z_i)^2 + \beta \sum_{j=1}^{h} KL(p||p_j) \tag{4}$$

Here, h indicates the number of neurons in the hidden layer, b is the rarity rate, and $\beta \sum_{j=1}^{h} KL(p||p_j)$ is the Kullback-Leibler (KL) divergence between the p and p_j averaged Bernoulli random variables. KL divergence between two random variables is given as follows.

$$KL(p||p_j) = plog\left(\frac{p}{p_j}\right) + (1-p)\, log\left(\frac{1-p}{1-p_j}\right) \tag{5}$$

3.1 Convolutional Autoencoders

Convolutional autoencoders are the cutting edge methods for convolutional filters. After these filters have been trained to extract features, they can be applied to any content. Instead, these features can be used to perform any function requiring a compact representation of the data, such as classification [6].

The main difference between CNN and CAE's traditional understanding is that the former are trained end-to-end to learn filters and combine features to identify their data. CNN's are also commonly referred to as supervised learning algorithms. Instead the latter are only equipped to learn filters capable of extracting features that can be used to recreate the data.

CAEs scale well to realistic-sized high-dimensional images due to their evolutionary existence, since the number of parameters needed to generate an activation map, is always the same, no matter what the size of the data [7]. Hence, CAEs are extractors of general-purpose features different from AEs that disregard the 2D image structure. Also, the image must be unrolled into a single vector in AEs, and the network must be constructed according to the number of inputs constraint. In other words, AEs incorporate consistency in the parameters, requiring each function to be universal (i.e., to cover the entire field of vision), while CAEs do not.

Encode

It's easy to understand that a single convolutional filter cannot learn how to remove the broad range of patterns that make up a picture. For this reason, each convolutional layer is composed of convolutional filters n (hyper-parameter), each with depth D (the depth of the input).

Therefore, a convolution among an input volume I={I1,⋯, ID} and a set of n convolutional filters $\{F_1^{(1)}, \dots, F_1^{(n)}\}$ each with depth D, produces a set of n activation maps, or equivalently, a volume of activations maps with depth n:

$$O_m(i,j) = a\left(\sum_{d=1}^{D}\sum_{u=-2k-1}^{2k+1}\sum_{v=-2k-1}^{2k+1} F_{m_d}^{(1)}(u,v)(i-u,j-v)\right)\ m = 1,\dots,n \qquad (6)$$

To improve the network's generalization capabilities, each conversion is wrapped in a non-linear function (activation), so the training process can learn to represent input combining non-linear functions:

$$z_m = O_m = a\left(I * F_m^{(1)} + b_m^{(1)}\right)\ m = 1,\dots,m \qquad (7)$$

Where $b_m^{(1)}$ is the bias for the m^{th} function map (single real value for each activation map).The word z_m was used to use the same name as the variable for the latent variable used in the autoencoders. The activation maps generated are the encoding of the input *I* in a low dimensional space; a dimension that is not *O* dimension (width and height) but the number of parameters used to construct each feature map O_m, that is, the number of parameters to be learned [8]. Since the aim to reconstruct the input *I* from the feature maps generated, we would like a decoding operation capable of doing this. Convolutional autoencoders are entirely convolutional networks so the decoding process is a convolution again.

Decode

Reconstructing the input image *I* from this reduced representation, the generated *n* function maps $z_m=1...n$ serial latent representations will be used as input to the decoder [9]. The decoding convolution hyper-parameters are set by the encoding architecture. Filters volume $F^{(2)}$ with dimensions *(2k+1, 2k+1, n)* since the convolution will cover every feature map and generate a volume with the same spatial extent *I* and number of filters to learn *D* as we are interested in reconstructing the D-depth input image. Hence the reconstructed image $\tilde{I}$ is the product of the convolution between the volume of feature maps $Z = \{z_i=1\}^n$ and the volume $F^{(2)}$ of this evolutionary filter.

$$\tilde{I} = a(Z * F_m^{(2)} + b^{(2)}) \qquad (8)$$

4 Material and Method

4.1 Dataset

FVC2002 fingerprint dataset [10] is used for the model. The dataset has four sensor fingerprints. These are Low-cost Optical Sensor, Low-cost Capacitive Sensor, Optical Sensor, and Synthetic Generator. It has 4x80 = 320 images in total.

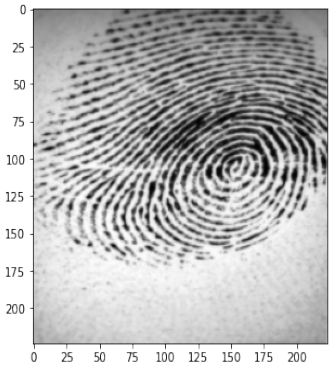


Fig. 3. An Example from Dataset

4.2 Convolutional Autoencoder Model

The images are transformed into 224 x 224 x 1 or 50,176 dimensional matrix. The matrix is normalized into an array of 0 and 1 as input. The batch size is chosen 128. This contributes significantly to the determination of the learning parameters and affects prediction accuracy. There are 32 filters with 3x3 in the first layer of an encoder. The down sampling layer uses max-pooling. The next layer has 64 scales 3 x 3 filters with another down sampling layer. The end encoder layer will have 128 scales 3 x 3 filters. A decoder has the opposite layer architecture. As a loss function, mean squared error (MSE) is used after every batch pixel by pixel. After 300 epochs training, the auto encoder model has a 0.86183 score. The lack of recognition and the lack of preparation all are in line. This indicates that the model isn't overfitting.

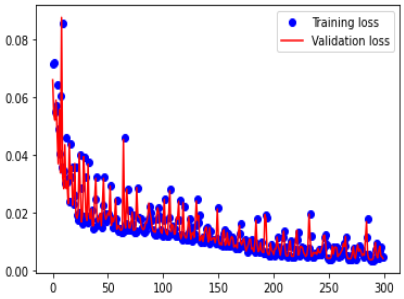


Fig. 4. Training and Validation Loss

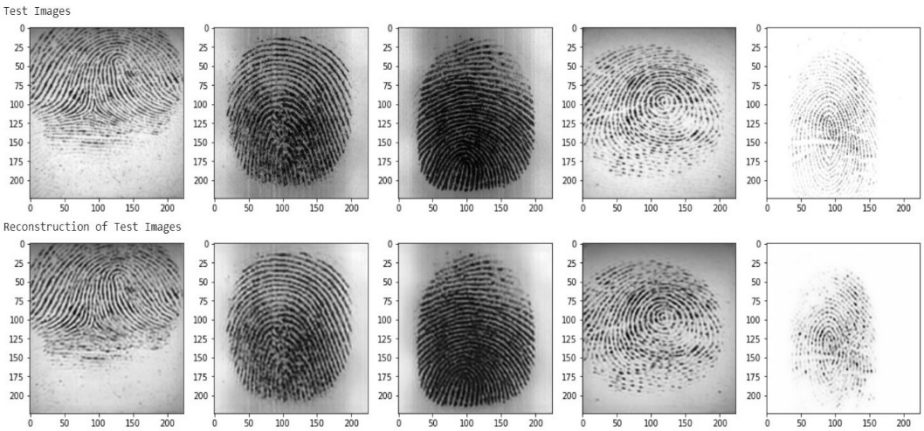


Fig. 5. Test Images and Reconstructed Images

5 Results and Conclusion

In terms of features, fingerprints are applied in various authentication systems. Generally, these authentication systems are equivalent to all of the information and symbol-based authentication, and they all work with data storage in a database. However, in fingerprint-based authentication systems, there must be additional modules to process the data. Despite their complexity, fingerprint authentication systems are systems where users do not experience problems such as password forgotten or loss of key that may occur in information and symbol based authentication systems, respectively. Among the common methods used extensively in image processing tasks are traditional auto encoders for image denoising and painting. The possibility of reconstructing latent or impaired fingerprints using the convolutional auto encoders is proposed in this paper. The generated model has a 0.86183 accuracy score. The developed auto encoder model detects the feature points used in fingerprint recognition. The biggest advantage of using this model instead of classical approaches is that it can be easily concluded without any derivative calculations with high performance and thus it is possible to avoid complex calculations. It is observed that using convolutional auto encoder based system to find feature points in the pictures of fingerprints gives successful results.

References

1. Jain, Anil, Karthik Nandakumar, and Arun Ross. "Score normalization in multimodal biometric systems." *Pattern recognition* 38.12 2270-2285 (2005)
2. Stigler, Stephen M. "Galton and identification by fingerprints." *Genetics* 140.3 857 (1995)
3. Lee, Chih-Jen, and Sheng-De Wang. "Fingerprint feature extraction using Gabor filters." *Electronics Letters* 35.4 288-290 (1999):

4. Tan, Chun Chet, and Chikkannan Eswaran. "Performance comparison of three types of autoencoder neural networks." *2008 Second Asia International Conference on Modelling & Simulation (AMS)*. IEEE (2008)
5. Makhzani, Alireza, et al. "Adversarial autoencoders." *arXiv preprint arXiv:1511.05644* (2015).
6. Masci, Jonathan, et al. "Stacked convolutional auto-encoders for hierarchical feature extraction." *International conference on artificial neural networks*. Springer, Berlin, Heidelberg (2011)
7. Mao, Xiao-Jiao, Chunhua Shen, and Yu-Bin Yang. "Image restoration using convolutional auto-encoders with symmetric skip connections." *arXiv preprint arXiv:1606.08921* (2016).
8. Chen, Junjie, and Xinghua Shi. "A Sparse Convolutional Predictor with Denoising Autoencoders for Phenotype Prediction." Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. (2019)
9. Makhzani, Alireza, and Brendan Frey. "K-sparse autoencoders." *arXiv preprint arXiv:1312.5663* (2013).
10. Maio, Dario, et al. "FVC2004: Third fingerprint verification competition." *International conference on biometric authentication*. Springer, Berlin, Heidelberg, (2004)

Ship Detection from Göktürk-2 Satellite Images using Convolutional neural network

SOHAIB K. M. ABUJAYYAB[1], I.R. KARAS[2], C. AKBULUT[3], V.E. TELLIER[4]

[1] Karabuk University, Karabuk/Turkey, s.jayyab@hotmail.com
[2] Karabuk University, Karabuk/Turkey, irkaras@gmail.com
[3] Karabuk University, Karabuk/Turkey, akbuluttceren@gmail.com
[4] Karabuk University, Karabuk/Turkey, vemretellier@gmail.com

Abstract. Ship detection from satellite images involves multiple steps of processing. Ship detection from satellite images is useful for controlling of ship traffic and determining the ship when necessary. Ship detection also used in solving several problems such as marine field violations and illegal hunting. Nowadays, optical satellite images and remote sensing methods are being used for ship detection. This project constitutes a prototype model to provide data to the government, coast guard units, general maritime, ship traffic services centers, and national defense. The aim of this study is to utilize the existing Turkish satellite data to perform ship detection. Göktürk-2 satellite images used in turkey for observation and were selected for the analysis. It is possible to distinguish between ships and all other objects in satellite images. Deep Convolution Neural Network (DCNN) model was proposed to obtain results. Some feature-based corrections have been made for satellite images. The shapes of the ships were detected in squares and their numbers were specified. This model has obtained a sufficient accuracy value by handling a pre-processed satellite image. Graphical results show that the proposed model provides an efficient detection process with an accuracy of 89.60%. The designed CNN model was preferred in a simple and easy to apply form, especially based on the Göktürk-2 satellite images.

Keywords: Deep convolutional neural networks, ship detection, remote sensing, satellite imagery.

1. Introduction

Several methods including Computer vision and artificial intelligence for solving varied problems in our life for long time without the need for human power [1]. Lately, numerous approaches in many areas have been replaced by deep learning methods due to the required high computing power in hardware, amount of data, and the rapid development of algorithms [2]. As in any case, these methods are of interest in the analysis of satellite images and their use is becoming widespread.

Analysis of satellite images plays a crucial role in many areas such as the determination of forest areas and fires, the monitoring of cultivated areas in agricultural land, city and road planning, security and military surveillance, disaster and crisis management [3]. Considering the available satellite images that can cover square kilometers,

former methods are very costly and time consuming [3], [4]. In addition, in order to obtain correct results, the people who will be performing the analysis tasks need to have experience in their fields. In the analysis of satellite images, computer vision solutions are categorized under three headings: classification, segmentation, and object detection. Convolutional neural networks (CNN) with deep learning architecture are used in all these analyzes, making use of the positional features on the image [1].

Currently, there are six Turkish satellites existing, three of them are used for communication and three are utilized for observations. Türksat 3A, Türksat 4A, and Türksat 4B satellites are used for communication. RASAT, Göktürk-1, and Göktürk-2 satellites meet the needs of observation purposes. Since the images of the ships in the seas were used in the project, images were taken from Göktürk-2, the observation satellite. Those images with high resolution are suitable for model training.

In this paper, ship detection using Convolutional neural networks was made from the satellite images. The aim is to propose a prototype model using the available Turkish satellite data to help in sea traffic management. The project can help institutions such as coast guard, defense industry, statistical studies, national defense, TAF, TURKSTAT, and SSB to obtain the relevant data.

2. Methods

In this section, tools development and satellite image processing methods implemented using ArcMap and MATLAB platforms. The general purpose of using MATLAB program is the CNN model development and training, ArcMap platform utilized to process the satellite images.

Figure 1. Manual selecting of ships from Göktürk-2 satellite images for training dataset.

2.1. Satellite images collection

A petition sent to "Ankara Keşif Uydu Komutanlığı" obtain "Göktürk-2" satellite images. As a result, 300 GB of satellite images were collected which are very useful for model training. According to the data exploring several images selected for the image extraction stage. The selected optical Göktürk-2 images were having 5-meter spatial resolution. All images were selected from Turkey in the data set. There are some images along the Mediterranean and Marmara Regions. (e.g., Kumluova-Armutveren, Fethiye-Zvezdets, İbiller-Bandırma, Manisa-Demirci Keslik, Bayraklı-Topcukoy, Boğazcık-Ayvacık, Dalaman-Karaevli). All the images that mostly cloudy, containing strips of no data were excluded.

2.2. Dataset Preprocessing

In this stage, the satellite images were processes, and ships were extracted. The satellite images were processed to generate the colored RGB images. Then, the ships were manually defined by sets of squares. Figure 1. Illustrate the manual selecting of ships from Göktürk-2 satellite images for the training dataset. Then, a model builder in Arcmap generated to extract the images for each ship from the images as illustrated in Figure 2. As a result of these processes, 94 RGB images of ships and non-ship with 63x62 24 bit depth were extracted. Examples of these images were presented in Figure 3.

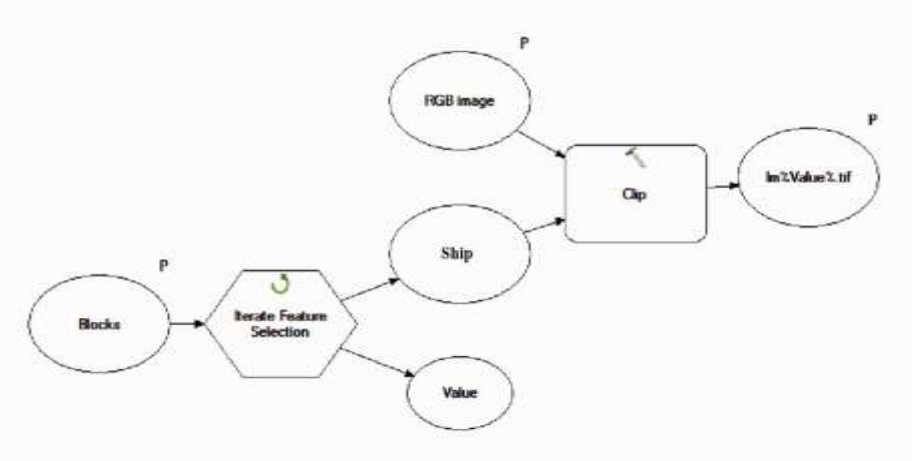
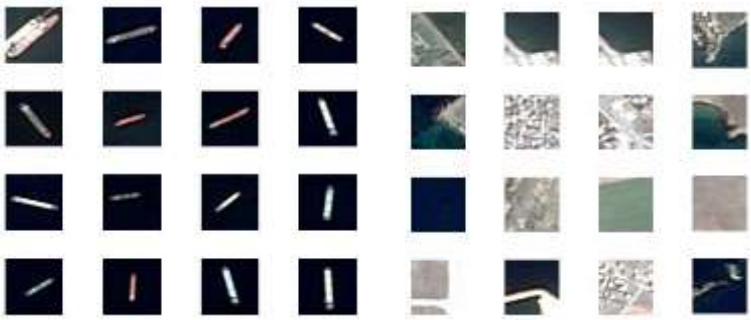


Figure 2. Model builder in Arcmap for extracting ships images from Göktürk-2 satellite images for CNN training.



(A) (B)

Figure 3. Examples of ships (A) and non-ships (B) extracted RGB images.

2.3. Convolutional neural network CNN

The main functionality of convolutional neural network is to perform images recognition, object detections and images classifications. CNN takes an input picture, processes and categorizes it. The input image as an array of pixels and depends on the image resolution. It refers to pixel values h x w x d (h = Height, w = Width, d = Dimension). CNN looks like a mixture of sciences such as biology and computers, but it is a very effective system used for image recognition. CNN detects low-level features such as curves and edges in an image and create concepts that are more abstract [5].

CNN uses the standard Neural Network to solve the classification problem, but uses other layers to identify information and identify some features. CNN consisting several layers of processing such as Convolutional Layer, Non-Linearity Layer, Pooling (Downsampling) Layer, Flattening Layer, Fully-Connected Layer. The CNN layers presented in Figure 4.

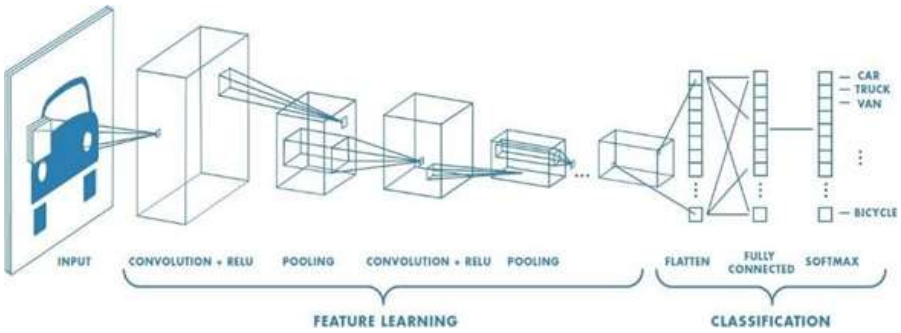


Figure 4. The layers of image processing in convolutional neural network

At this stage, deep learning was implemented based on convolutional neural networks to generate the model for ship detection. Firstly, the images of the ships were imported in Matlab. Then the CNN architecture defined (62 x 63 x 3 pixels). The train-ing options specified. The structure of this architecture was created from 8 layers that are related to

each other. These layers are Image Input Layer, Convolutional Layer, Batch Normalization Layer, ReLU Layer, Max Pooling Layer, Fully Connected Layer, Softmax Layer, Classification Layer. Specifying Training Options: After defining the network structure, training options were specified. SGDM (stochastic gradient descent with momentum) optimization process was applied in the network education with an initial learning rate of 0.01. The maximum Epoch value is set to 40. The epoch value is a complete training cycle across the entire training data set. Data was mixed in each cycle. Verification data and frequency were specified as 8. A CNN architecture created by adding a ReLU layer after each. Then, the CNN model generated and trained. The performance accuracy was calculated.

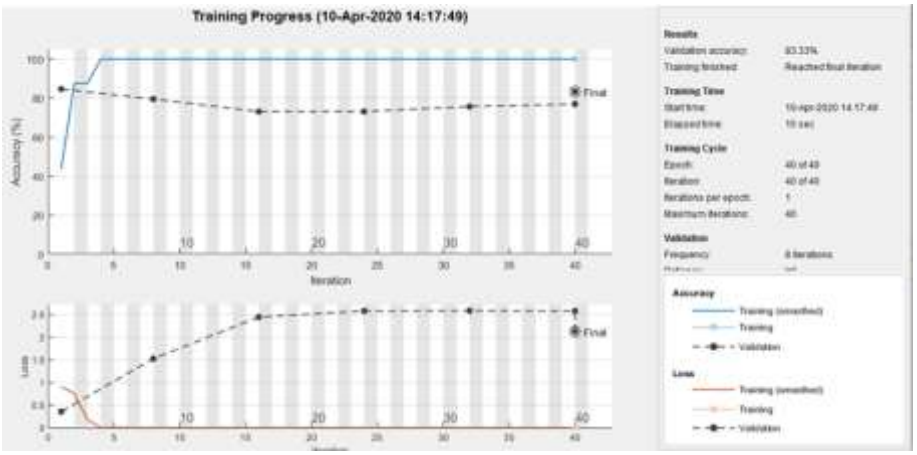


Figure 5. Performance accuracy of CNN based on the Göktürk-2 satellite images for ship detection

3. Results and conclusion

In this article, ship detection was made using the CNN model in Göktürk-2 satellite images. Detection was made using layers that are fully linked with the convolutional neural networks. Satellite images used effectively for ship detection. The progress of the model education was graphically monitored. The proposed model provides both accuracy and other performance evaluation parameters. The labels of verification data were estimated using the trained network and the final verification accuracy was calculated. Accuracy is the proportion of images that correctly classified by the network. At the end of the study, the accuracy value was calculated as 83.33%. Im-ages with the feature of ship classification and detection from the satellite can be used in future studies, with a larger dataset, to improve the model learning process. The obtained accuracy value as a result of this study is shown in Figure 5.

References

[1] Y. A. Kara, Ö. K. Uçarer, and B. Gündoğdu, "Automatic Warship Recognition System : Dataset, Feature Representation and Classification Analysis," in *2019 27th Signal Processing and Communications Applications Conference (SIU)*, 2019, pp. 1–4, doi: 10.1109/SIU.2019.8806462.

[2] F. UCAR and D. KORKMAZ, "Ship Target Classification in Satellite Images using Deep Convolutional Neural Networks," *Sak. Univ. J. Sci.*, vol. 24, no. 1, pp. 197–202, Feb. 2020, doi: 10.16984/saufenbilder.587731.

[3] Y. Liu, H.-Y. Cui, Z. Kuang, and G.-Q. Li, "Ship Detection and Classification on Optical Remote Sensing Images Using Deep Learning," *ITM Web Conf.*, vol. 12, 2017.

[4] B. Jiang, X. Ma, Y. Lu, Y. Li, L. Feng, and Z. Shi, "Ship detection in spaceborne infrared images based on Convolutional Neural Networks and synthetic targets," *Infrared Phys. Technol.*, vol. 97, pp. 229–234, Mar. 2019, doi: 10.1016/j.infrared.2018.12.040.

[5] Ö. Kutlu, Ö. Demir, and B. Doğan, "Analysis Of Images Obtained By Unmanned Aerial Vehicle By Deep Learning Methods," in *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, 2019, pp. 1–4, doi: 10.1109/UBMYK48245.2019.8965587.

Doc2vec Approach for Text Document Clustering

Eray YELMEN[1] and Nevcihan DURU[1]

[1] Kocaeli University, Kocaeli, Turkey
erayyelmen@hotmail.com, nduru@kocaeli.edu.tr

Abstract. Manuel categorization of text documents according to their similarities are very difficult and take long time. Therefore, highly evolved means and methods are needed to deal with this challenge. Clustering is an important method used in many applications of business and data sciences. Document clustering separates text data into several groups, where documents in each group are specified by measure of proximity or similarity. To achieve a high success rate in document clustering assumes a main role in helping its users to correctly explore, condense, and sort out the data. In this paper, we propose a word embedding approach to cluster question based text documents. In the proposed approach, doc2vec word embedding method gives the better result on gaussian mixture clustering algorithms instead of other word embedding and clustering methods.

Keywords: Document Clustering, Word Embedding, Doc2Vec, Gaussian Mixture.

1 Introduction

In the past decade, software developers have been sharing their questions with each other through Question and Answer (Q&A) websites. Consequently, these websites have become significant information repositories, covering many topics related to specific programming languages [23]. StackOverflow.com is one of the most preferred Q&A website focused on questions related to programming languages.

Software developers can post their issues directly via StackOverflow, which causes the same or similar questions to occur on the site. In addition, users may have difficulty in finding the answers to the issues that are related to each other. Therefore, there is a need for an automatic cluster of huge data in order for users to find answers to their problems quickly and accurately.

One of the most beneficial solutions to manage this big amount of data is to cluster them automatically according to the similarities. Clustering is an important data mining method in categorizing, classifying and organizing text documents [14]. The high applicability of document clustering has led towards a number of other area such as information retrieval, web mining and search engines [24]. The Challenge is to discover similar documents from the set of unstructured documents.

Until this time, many traditional algorithms and methods have been applied for the document clustering. However, with the rapid development of storage systems for data

storage and information exchange, the challenge for the existing clustering algorithms has also been increased. The reality behind the challenging issue of algorithms may be because of data driven approach of clustering algorithms. The data contains structure format as well as unstructured format at the same time. Besides, natural language contents have even expanded the complexity factor of data [25]. Therefore, in order to cluster documents based on either content-based grouping or their topic model requires automated and powerful machine learning techniques to cope with complexity of data.

Taking advantage of the fast development of natural language processing area, many neural network language models have applied to address representation problem in document clustering. Word embeddings are commonly used as simple building blocks for text representation trained by Word2vec [12] and Glove [13]. Documents can be explicit by these word embeddings methods with weights. However, these word embeddings are ignore the polysemy of words and also they are uncontextualized.

This paper proposes categorization methods of unsupervised document utilizing deep embedding clustering, to automate question based document clustering. The method that we propose is as follows. The word embedding vector representation method is applied to embed the documents into the vector space. By applying the deep embedding clustering method, the proposed method enables that a distance between different types of documents in a space is further distanced and the distance between similar documents is made closer to improve the success rate of clustering. By applying doc2vec word embedding method, we achieved higher document clustering success rate instead of other word embedding techniques.

The rest of this paper is organized as follows. In Section 2, the related work on document clustering with the three document representation methods (Fasttext, Word2Vec, Doc2Vec) and clustering algorithms is discussed. In Section 3, the details of dataset are presented. The data preprocessing, feature representation, clustering and experimental works are presented in Section 4, 5, 6 and 7, respectively. Finally, in Section 8, we conclude the current study with some future research directions.

2 Related Work

Recently, the document clustering has become a popular research area in the machine learning. It has a wide of usage such as in information retrieval, topic modelling and document organization. The document clustering is the task of separating the documents according to their meaning similarity. It focuses on splitting the documents into separate clusters therefore the documents belonging to one cluster are much similar to each other and different from in other clusters of documents [1].

Certain researchers have investigated the utility of linguistic features to group documents according to their similarity. They have applied various techniques both vector space and word embedding with clustering algorithms. Vector Space Model (VSM) is a popular document representation approach. It transform unstructured documents into a high dimensional vector. In spite of the popularity, this model has two main weaknesses. It loses the ordering of the words and cannot establish a semantic relationship between words [9].

In a study, for document clustering a similarity measure space model was suggested. The suggested model derives low dimensional semantic subspace of text documents corresponding to the same semantic between the documents in the local patches and outside these patches respectively by maximizing and minimizing the correlation [2].

In 2010, Shin-Jye Lee et al, recommended clustering-based scheme to recognize the fuzzy system. To start the mission, is tried to proposed a hybrid clustering based modular method. After, finding the number and position of clusters seemed the prime concerns for developing such a model. So, taking input, output, specialization and generalization a HCA has been designed. This three-part enter production clustering method accept lots of clustering characteristics all together to identify the issue [11].

In another study, a correlated application domain of mining, e-mails are group by using structural, and domain-specific features. K-means, Bisecting K-means and EM clustering algorithms were used in this study [10].

A method is presented in [3] to learn an inter-document similarity measure on human reasoning. In the proposed method, a document dataset with similarities rated by humans was used to train and calculate the methods and measure of machine learning. It is called as HE50 [4]. The measure is uses concepts from Wikipedia [5] and WordNet which is based on the BOC model. It is obtained from the experimental results that Wikipedia is more effective for this application than WordNet. To classify and cluster documents from four datasets the measure has applied. For the clustering task with hierarchical and k-means algorithms, the studied measure has performed better than cosine.

In a research, despite both [6,7] are including heterogeneous text document, when clustering documents consisting of text the authors has evaluated their algorithms. They selected a subset of the document collection from 20 Newsgroups and showed that their algorithms perform better than DBSCAN and k-means.

Doc2Vec is an extension of the word-to-vector (Word2Vec) representation which is the newest among the three document representation schemes. The word2Vec assumption is that the item values of a word are affected by the values of other words surrounding the target word. This assumption is encoded as a neural network structure, such as skip-gram, the network weights and continuous bag-of-words are adjusted by learning observed examples [8]. Doc2Vec extends Word2Vec because sentences can also be considered as documents [9].

In this study, we have proposed a deep embedding based approach to cluster unstructured question based documents. After NLP preprocessing, three different word embedding techniques which are doc2vec, word2vec and fastText have applied for training respectively. Then k-means++, k-medoids and gaussian mixture have applied on the trained data to cluster the documents according to their similarities. As a result, applying gaussian mixture clustering algorithm on doc2vec has outperformed other word embedding techniques and clustering algorithms.

3 Dataset

In the data collection phase, a total of 617,347 sample data was collected from between January and October 2019 using the stackexchange API via the StackOverflow website.

The top 10 keywords were identified by separating the collected data (documents). It was determined that all of these 10 keywords were programming languages. Distribution of questions according to programming languages are shown in Fig. 1. Documents related to the first 2 most asked programming languages were prepared as a dataset. Detailed information of the used data in analysis is shown in Table 1.

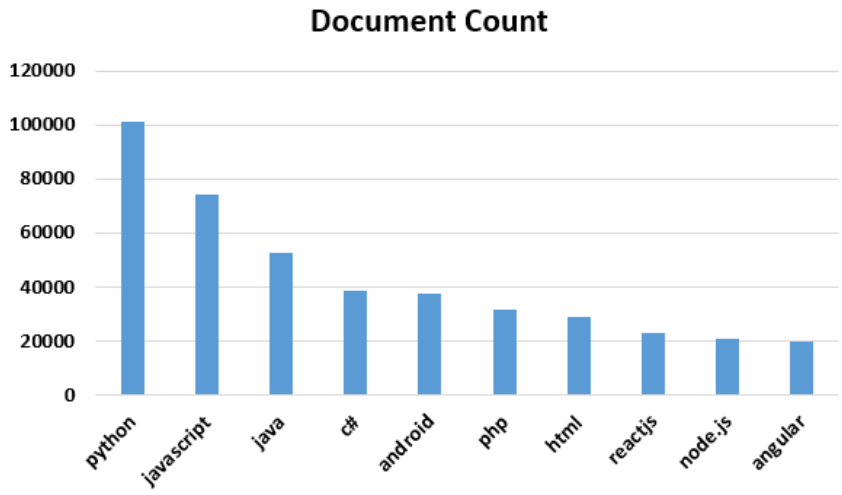


Fig. 1. Top 10 Keywords.

Table 1. Dataset.

Keyword	Number of Documents
Python	101.111
Javascript	74.357

4 Data Preprocessing

During the data preprocessing phase, links, numeric characters, punctuation marks, stop words, blockquote and code tags were removed in documents. In addition, all words were converted to lower case. Next, using the symspell library in python the word spell correction process was performed on the data that had completed the normalization

process. Lastly, the words were broken down to their roots in order to increase the success rate of the model. Since rooting the misspelled words would be wrong, stemming was applied with the program previously prepared using the snowballstemmer library after the spell correction step. All data preprocessing steps are shown in Fig. 2.

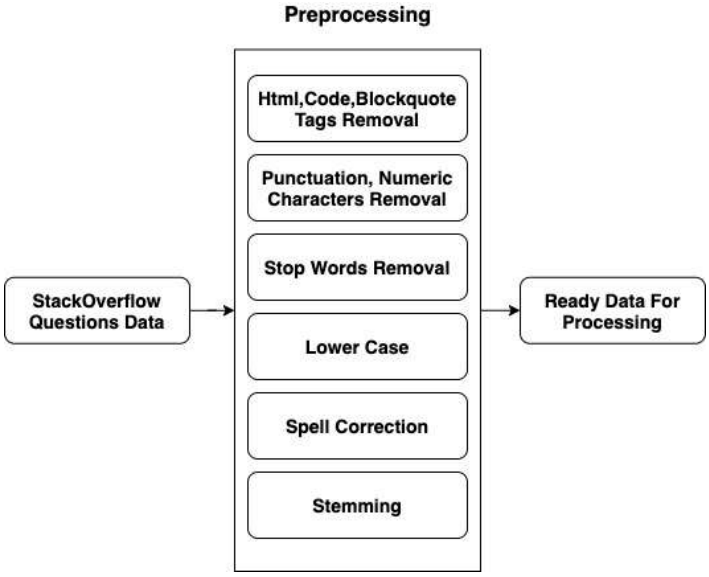


Fig. 2. Preprocessing steps.

5 Feature Representation

In this work, the performance of three word embedding methods was evaluated to create feature representation for documents combined with three clustering algorithms.

A word2vec model is trained to create a dense vector with fixed dimension for every token in a corpus. Although a pre-trained word2vec model is available for social media text data [27], the authors found that it did not perform well in the social media data sets used in the work. One issue was occurred that many tokens in the data were except the trained model's vocabulary, and also the semantic relationships between words may be very different on different data sets. On the other hand, there was no pre-trained model for large amounts of Reddit data. Moreover, these models contain many hyper parameters, so finding an ideal value set for different data sets has been a useful addition. For these reasons, they have trained their own word and document embedding models [28].

The word2vec models was trained with the continuous bag of words (CBOW) technique [8] with 100 epoch, 200 vector size, a context window of size 5 and minimum word count of 5. Variations of hyper-parameters was tested with including context window sizes ranging from 2 to 10, higher vector size and minimum counts of word. We found that, if the vector size and epoch are increased then silhouette score was

increased. On the other hand when we tested epoch values between 25 and 250 in increments of 25 we obtained the optimal epoch value was 100.

FastText is a word embedding model that uses a combination of methods to consider distributed word representation and word order [30]. It calculates embeddings in a similar way as the skip gram model, but with the label as the middle word and instead a bag of words, which captures some information about the word order. The algorithm works as supervised and unsupervised. In supervised mode, the documents are converted to vectors by averaging the embeddings. Besides, fastText simply generates word embeddings for general purposes in the unsupervised mode, then not taking classes into account [17].

In the experiment, we tested variations of hyper-parameters with 100 vector size, 50 epoch, a context window of size 5 and minimum word count of 5. We found that, if the epoch is decreased then silhouette score was increased. On the other hand when we tested epoch values between 25 and 250 in increments of 25, we obtained the optimal epoch value was 50.

Doc2vec is a neural network based word embedding method for representing documents as a vector and is a generalization of the word2vec method. The doc2vec models in doc2vec matrix for each document were trained with 200 vector size with 25 epoch using the distributed memory (dm), a context window of size 8. Distributed memory perform as a memory which remembers what is missing is the topic of the paragraph or from the current context. As a result of the experiments we found that, if the epoch is decreased then silhouette score was increased. On the other hand when we tested epoch values between 25 and 250 in increments of 25 we obtained the optimal epoch value was 25. The experiments have applied on 2 different dataset and parameters are shown in Table 2.

Table 2. Model Training Parameter.

Dataset	Model	Epoch	Window Size	Vector Size
Questions related with Javascript	Doc2Vec	25	8	200
	Word2Vec	100	5	200
	fastText	50	5	100
Questions related with Python	Doc2Vec	25	8	200
	Word2Vec	100	5	200
	fastText	50	5	100

6 Clustering

The main problem of a clustering is that we have a limited training set $X = \{ x_i : x_i \in R^n , i = 1 , \ldots , m \}$ and we have to determinate k clusters separating this set on the basis of well-established criteria [18]. Clustering method can be defined as follows according with [19]:

Definition 1. Given a set of objects X, a clustering $C^* = \{ C_i : C_i \subseteq X, i = 1 , \ldots , k \}$ is a partition of X such that

$$U_{C_i \in \mathbb{C}} C_i = V \quad (1)$$

$$\forall C_1, C_j \in \mathbb{C}: C_i \cap C_j = \emptyset \; for \; i \neq j \quad (2)$$

Arthur and Vassilvitskii [29] defined K-means++, an algorithm to seed the initial cluster centers which improves clustering. The K-means++ algorithm selects randomly the first centroid. The subsequent centroids are then selected according to the minimum probable distance. The K-means++ algorithm selects one point in each iteration according to a non-uniform distribution and uses k iterations [16].

Another commonly used partitioning method is K-medoids algorithm. Some features of this algorithm are similar with k-means's, kernel of this algorithm is denifinitely different from k-means. K-medoids gets the centre points as candidate by sampling objects at random. The approach adopted to improve the quality of centers is to select random objects as candidate centers that are better than existing centers, replaces the existing centers. So, cost of implementation is very high.

K-medoids algorithm's input and output are similar with k-means. Aim of the algorithm is to find final centres. Algorithm gets the final centres means that gets the final clustering results. There are two issues should be done. first one is, how to measure an object is one of the final centres; second one is, how to deal with the objects being affected if one subcluster's centre is changed [20].

Gaussian Mixture is a model to solving the clustering problem which is very effective in certain cases but it might not be the fastest [21]. The Gaussian Mixture divides a specific data set into k number of clusters, depending on the parameters of each set including variance, mean and previous probability. Each of these parameters are determined by Expectation Maximization (EM) algorithm. The algorithm steps are as follows: If a new data point is given, the probability that it belongs to a particular set is calculated and the data point is assigned to the highest probability cluster [22].

In order to compare the results obtained with 2 different data sets, we have selected three different clustering algorithms. First, we used the k-means ++ clustering algorithm using Euclidean metric and maximum 100 iterations. The algorithm was run over the data as multiple times with varying random seeds. Then we applied k-medoids and gaussian mixture algorithms respectively with a maximum of 100 iterations.

Our aim in this work is to cluster the similar documents by using word embedding and clustering algorithms together. We have experimented 3 word embedding methods with three clustering algorithms: K-means++, K-medoids and Gaussian Mixture.

7 Experiments and Evaluation

Before we start to experiment, the raw data collected from Stackoverflow was first of all refined from unnecessary expressions and stop words. Then, to obtain the high score in document clustering phase, spell correction and stemming processes are applied in preprocessing phase.

An important hyper parameter for training neural network models is the number of epochs. Too many epochs and the model may not fit to the data, too few and poor

performance [28]. Firstly, we explored performance change of the mean word2vec, doc2vec and fasttext models with the number of epochs. We applied k-means++, k-medoids and gaussian mixture clustering for the clustering method to find the best results for the embedding representations. For each epoch value between 25 and 250, with increments of 25. This was done for all two data sets.

A better performance is obtained with a maximum value 100 in word2vec. For the doc2vec, we set the optimal number of epochs to be 25 for the doc2vec and 50 for the fastText methods.

When Word2Vec method and K-Means ++ algorithm are used together on Python related dataset, it gives much better results than other 2 clustering algorithms. In Javascript related data set, the scores of K-Means ++ and Gaussian Mixture algorithms are close to each other and Gaussian Mixture gave higher results. The lowest success score was obtained with K-Medoids algorithm. It was determined that the number of data increased, and the score of the Gaussian Mixture algorithm increased in both data sets.

The combination of the fastText method and the K-Means ++ algorithm gave the highest result, and the K-Medoids algorithm yielded higher results than Gaussian Mixture. When we interpret the results, it can be said that the consistency in the cluster in the K-Means ++ algorithm is better than other algorithms. It has been observed that the 3 clustering algorithm gives similar results on dataset related with python and javascript.

It was determined that the score obtained with the use of Doc2Vec method together with the K-Means ++ clustering algorithm yielded higher results than the K-Medoids algorithm. However, the Gaussian Mixture algorithm yielded much higher results than the other 2 clustering algorithms. When we interpret the results, it can be said that the consistency in the cluster in the Gaussian Mixture algorithm is better than other algorithms. It has been observed that the 3 clustering algorithm gives similar results on dataset related with python and javascript.

One of the important parameters for the success of the model is the vector size. The vector maps the document to a point in an n-dimensional space, and more distinction can be made between documents as the dimensions grow. As the vector size increases for Doc2Vec and Word2Vec, the success rate of the model increases, while the success of the model in fastText is higher in the lower vector size. The optimum value is 200 for Doc2Vec and Word2Vec, while it is 100 for fastText.

It is clear from Table 3 and Table 4 that the doc2vec feature representation model with gaussian mixture clustering algorithm outperformed the other approaches. This is because the Doc2Vec model is based on contextual similarity. In other models, the similarity is based only on words.

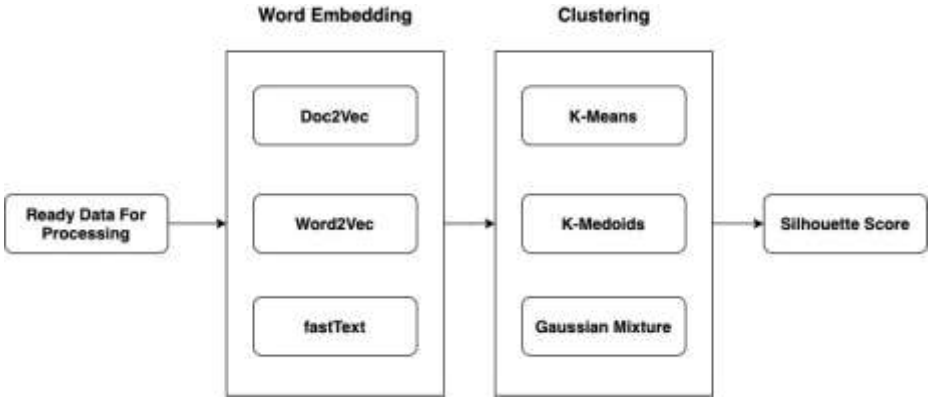


Fig. 3. Proposed System Architecture

Table 3. Clustering Scores (Python).

Feature Representation	Clustering Algorithm	Silhouette Score
Doc2Vec	K-Means++	0.051210795
	K-Medoids	0.032164106
	Gaussian Mixture	0.114848055
Word2Vec	K-Means++	0.040494468
	K-Medoids	0.008342793
	Gaussian Mixture	0.014674502
fastText	K-Means++	0.021525037
	K-Medoids	0.0056469887
	Gaussian Mixture	-0.021828536

Table 4. Clustering Scores (Javascript).

Feature Representation	Clustering Algorithm	Silhouette Score
Doc2Vec	K-Means++	0.052172665
	K-Medoids	0.034015682
	Gaussian Mixture	0.09895941
Word2Vec	K-Means++	0.030102726
	K-Medoids	0.0032026912
	Gaussian Mixture	0.03829827
fastText	K-Means++	0.03905312
	K-Medoids	0.012441033
	Gaussian Mixture	-0.022544546

8 Conclusion and Feature Work

This paper presents an automatic clustering method of posted questions in stackoverflow. Our results showed that word embedding representations of question-based data can be effectively used as the basis for document clustering. Doc2vec word embedding approach outperformed word2vec and fastText techniques. In Doc2Vec and fastText models, the success rate was higher when the number of epoch was lower, while the Silhouette score was higher as the number of epoch increased in the Word2Vec approach. We also demonstrated that gaussian mixture clustering provided the best silhouette score with doc2vec model.

As future work, first of all, feature extraction will be done on vectors created by using word embedding methods. Since feature extraction will narrow the search space, it is aimed to increase the success of the model and obtain more meaningful information by doing this operation correctly. Secondly, clustering of similar documents will be carried out experimental studies with fuzzy c-means clustering algorithm and also different metrics will be used to evaluate clustering scores.

References

1. Klinczak, M., and Kaestner, C.: Comparison of clustering algorithms for the identification of topics on twitter. Latin American Journal of Computing - LAJC, 3, 19–26 (2016).
2. Zhang, T., Tang, Y. Y., Fang, B. and Xiang, Y.: Document clustering in correlation similarity measure space, Knowledge and Data Engineering, IEEE Transactions 24(6), 1002–1013 (2012).
3. Huang, L., Milne, D, Frank, E.,Witten, IH.: Learning a concept-based document similarity measure. JAm Soc Inf Sci Technol 63(8), 1593–1608 (2012).
4. Lee, MD., Welsh, M.: An empirical evaluation of models of text document similarity. In: In CogSci2005, Erlbaum, pp. 1254–1259. (2005).
5. Hu J, Fang L, Cao Y, Zeng HJ, Li H, Yang Q, Chen Z.: Enhancing text clustering by leveraging wikipedia semantics. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,ACM, NewYork,NY, USA, SIGIR '08, pp. 179–186. (2008).
6. Meng L, Tan AH, Xu D.: Semi-supervised heterogeneous fusion for multimedia data co-clustering. IEEE Trans Knowl Data Eng 26(9), 2293–2306 (2014).
7. Meng L, Tan AH,Wunsch DC.: Adaptive scaling of cluster boundaries for large-scale social media data clustering. IEEE Trans Neural Netw Learn Syst 27(12), 2656–2669 (2015).
8. Mikolov, T., Chen, K., Corrado, G. and Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013).
9. Le, Q. and Mikolov, T.: Distributed representations of sentences and documents. In International conference on machine learning, pp. 1188-1196. (2014).
10. Huang, G.Y., Liang, D.P., Hu, C.Z. and Ren, J.D.: An algorithm for clustering heterogeneous data streams with uncertainty. In 2010 International Conference on Machine Learning and Cybernetics Vol. 4, pp. 2059-2064. IEEE (2010).
11. Lee, S.J. and Zeng, X.J.: A three-part input-output clustering-based approach to fuzzy system identification. In 2010 10th International Conference on Intelligent Systems Design and Applications, pp. 55-60. IEEE (2010).

12. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J.: Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pp. 3111-3119. (2013).
13. Pennington, J., Socher, R. and Manning, C.D.: Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532-1543. (2014).
14. Sherkat, E., Nourashrafeddin, S., Milios, E.E. and Minghim, R.: Interactive document clustering revisited: A visual analytics approach. In 23rd International Conference on Intelligent User Interfaces, pp. 281-292. (2018).
15. Abualigah, L.M., Khader, A.T., Al-Betar, M.A. and Alomari, O.A.: Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering. Expert Systems with Applications, 84, pp. 24-36. (2017).
16. Bahmani, B., Moseley, B., Vattani, A., Kumar, R. and Vassilvitskii, S.: Scalable k-means++. arXiv preprint arXiv, 1203.6402 (2012).
17. Stein, R. A., Jaques, P. A., and Valiati, J. F.: An analysis of hierarchical text classification using word embeddings. Information Sciences, 471, 216-232 (2019).
18. Schwenker, F. and Trentin, E.: Pattern classification and clustering: A review of partially supervised learning approaches. Pattern Recognition Letters, 37, pp. 4-14. (2014).
19. Stein, B. and Eißen, S.M.: December. Document categorization with MajorClust. In Proc. 12th Workshop on Information Technology and Systems. Citeseer. (2002).
20. Qiao, S., Geng, X., and Wu, M.: An improved method for K_medoids algorithm. In 2011 International Conference on Business Computing and Global Informatization, pp. 440-444. IEEE (2011).
21. Day, WH. and Edelsbrunner, H.: Efficient algorithms for agglomerative Gaussian Mixture Model clustering methods. Journal of classification, 1(1), 7–24 (1984).
22. Virupakshappa, K., and Oruklu, E.: Unsupervised Machine Learning for Ultrasonic Flaw Detection using Gaussian Mixture Modeling, K-Means Clustering and Mean Shift Clustering. In 2019 IEEE International Ultrasonics Symposium (IUS), pp. 647-649. IEEE (2019).
23. Bazelli, B., Hindle, A. and Stroulia, E.: On the personality traits of stackoverflow users. In 2013 IEEE international conference on software maintenance, pp. 460-463. IEEE (2013).
24. Shah, N. and Mahajan, S.: Document clustering: a detailed review. International Journal of Applied Information Systems, 4(5), pp. 30-38. (2012).
25. Jajoo, Pankaj. Document clustering. IIT Kharagpur, Thesis (2008).
26. Afzali, M. and Kumar, S.: Text Document Clustering: Issues and Challenges. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), pp. 263-268. IEEE (2019).
27. Godin, F., Vandersmissen, B., De Neve, W., and Van de Walle, R.: Named entity recognition for twitter microposts using distributed word representations. Proceedings of the workshop on noisy user-generated text. Association for Computational Linguistics 146–153 (2015).
28. Curiskis, S. A., Drake, B., Osborn, T. R., and Kennedy, P. J.: An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. Information Processing & Management, 57(2), 102034 (2020).
29. Arthur, D., and Vassilvitskii, S.: k-means++: The Advantages of Careful Seeding, SODA '07 Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (2007).
30. Joulin, A., Grave, E., Bojanowski, P. and Mikolov, T.: Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759 (2016).

Detection of Chronic Kidney Disease Stages by Data Mining Methods

E. KABULLAR[1], F. ATASOY[2] and N. ÖZKAN SEVENCAN[3]

[1]Karabuk University, Karabuk/Turkey, elifkabullar@karabuk.edu.tr
[2]Karabuk University, Karabuk/Turkey, ferhatatasoy@karabuk.edu.tr
[3]Karabuk University, Karabuk/Turkey, nurhayatozkan@karabuk.edu.tr

Abstract. Data mining has created a new perspective in the use of health data, in addition to finding answers to problems related to large amounts of data and it has become a method that the usage prevalence continues to increase rapidly.

Chronic kidney disease is an important public health problem in our country as all over the world. If chronic kidney disease is detected early, it can be prevented, or its progression can be prevented. In this study, a system has been developed using data mining to determine the stage of the disease of patients with kidney failure. The data in 2018 of Karabuk University Training and Research Hospital were used by obtaining the necessary ethics committee permissions as a data set.In practice, performance comparison was also made between these two algorithms using logistic regression and artificial neural networks. After the comparison, artificial neural networks were more successful in terms of model success criteria.

Keywords: Chronic kidney disease, Data Mining, Artificial Neural Network, Logistic Regression.

1. Introduction

Medical field is one of the most important branches in which data mining is used because of the data in that field is vital. This makes knowledge discovery to be collected from medical data important. It is thought that data mining used in the medical field will provide a new viewpoint by taking part in researches where there are some objectionable circumstances in terms of both economic and human health. Interdisciplinary studies in academic informatics are inevitable. Especially, considering the shortcomings of our country in the processing of medical data, it should be encouraged that the fields of medicine and informatics come together. Medical data should be available as an open source within the academic framework within the required limits. As long as human beings exist with an inference from the studies, the data mining studies in the medical world are increasingly important and will increase day by day. Because the data sets stored in this area are vital. With this study, we wanted to show the feasibility of successful academic informatics studies with medicine. When the literature is reviewed, it is seen that there have been many studies.

A. Kusiak et al [1] take part in a study on lung diseases. The aim of the study is to determine *whether the tumor* is benign or malignant. With the invasive test data that

were collected at different times and studies performed, the accuracy rate in the diagnosis was 100%.

Negar Zıasabounchi tried to design a specialist system in her thesis to determine the patient's degree of heart disease [2]. In her study, Cleveland was used the heart disease data set. Also, fuzzy systems and artificial neural networks were used together. The accuracy of the system was detected as 92.3%.

In the study by Lin, the classification and regression tree (CART) and case-based reasoning (CBR) techniques have been used appropriately to improve the diagnostic accuracy of liver disease and establish a clever diagnostic model aimed at providing a comprehensive analytical structure [3].The results indicate that the CART rate of accuracy is 92.94%. In the second phase, CBR is developed to diagnose the type of liver disease, and the new case triggers the CBR system to retrieve the most similar case from the case base in order to support the treatment of liver disease. The new case is supported by a similarity ratio, and the CBR diagnostic accuracy rate is 90.00%. Actual implementation shows that the intelligent diagnosis model is capable of integrating CART and CBR techniques to examine liver diseases with considerable accuracy.

The treatment and follow-up expenses of patients with kidney failure increase day by day. Today, it is estimated that the global cost of chronic kidney disease treatments worldwide is over $ 1 trillion. 6 percent of the health budget in the USA, 5 percent in Japan and 8 percent in Taiwan are spent on the treatment of patients with end-stage kidney disease[4]. Chronic kidney disease is a worldwide health crisis. For example, in the year 2005, there were approximately 58 million deaths worldwide, with 35 million attributed to chronic disease, according to the World Health Organization [5].When we consider the budgets allocated by countries, it is seen that early diagnosis has gained importance.

In this study, a system was developed using data mining to determine the stage of the disease of patients with kidney failure. The data in 2018 of Karabuk University Training and Research Hospital were used. There are 9 different characteristics of approximately 240 000 individuals on a 1-year data set. These features are age, gender, creatinine, potassium, parathormone, urea, uric acid and calcium values.

The aim in our study is to improve the quality of life of the patient, as well as to provide an effective and cost-effective solution by making use of routine tests made in hospitals. *Early detection of* chronic *kidney disease* can be determined by screening for estimated glomerular filtration rate (GFR) measurement, urine analysis (proteinuria) and the presence of microalbuminuria in diabetic patients. [6]

Kidney disease consists of 5 stages, as shown in Table 1, based on the glomerular filtration rate (GFR) which is presented by the Turkish Nephrology Association in the last full paper of 2018 [6]. GFR in a healthy individual is approximately 120-130 ml/min/1.73 m^2. When this rate drops below 15 ml, the patient also has end-stage renal failure (ESRD) and the patient enters the 5th stage. In practice, a training was carried out using logistics regression and artificial neural networks. artificial neural networks were more successful in terms of model success criteria. [6,7] (see Table1)

Table 1. Stages in chronic kidney disease

Phases	Definition	**GFR (ml/dk/1.73 m^2)**
1	**Presence of kidney damage with normal or increased GFR**	Normal or GFR>=90+presence of evidence of concomitant kidney damage *
2	**Mild disorder**	60-89 + presence of evidence of concomitant kidney damage *
3	**Moderate disorder**	30-59
4	**Serious disorder**	15-29
5	**End-stage renal failure**	<15

** Microalbuminuria in diabetic patient, persistent proteinide, persistent kidney-induced hematuria, structural anomaly in kidneys shown by USG or other imaging methods, presence of chronic kidney disease diagnosed by biopsy*

2. DataSet Description

Our data set consists of data which belongs to 2018, from KBU Training and Research Hospital. There are 9 different features of 240,000 individuals on 1 year data set. These features are age, gender, creatinine, potassium, parathormone, urea, uric acid and calcium.

GFR value is an important reference range in the diagnosis of the stage of the disease. There are two commonly used methods to calculate GFR. In our study, Modification of Diet in Renal Diseases (MDRD) formula was used, because it is more compatible for our data set.

Since creatine value is the main criterion in MDRD calculation, 240 thousand patients whose creatine analysis existed, were chosen. Data with incomplete creatine information were excluded. As a result of the elimination, 240 thousand patients were reduced to 150 thousand individuals whose MSDR value can be calculated. Our data preprocessing section also includes the assignment of values between 0 and 1 as a result of min-max normalization of the data as well as the reducing of missing data. A new feature called GFR has been created for kidney disease. According to the calculated GFR values, a new column named “Disease Evre” was created by taking the data in Table 1 as reference. This column is the determining factor for the classification process of disease stages. This column is the determining factor for the classification process of disease stages.

2.1 GFR Factor

GFR measurement takes an important place in the diagnosis of kidney failure. In this way, the filtering ability of the kidney is determined. Serum creatinine level varies according to patient characteristics (age, gender, race, size, muscle mass, diet etc.) [6]. Estimated GFR measurement corrected for age, gender, race, and body weight is a reliable method. In normal young adults, GFR is approximately; It is 120-130 ml /min/ 1.73 m^2 [7]. Generally, two methods have been used for the determination of GFR based on measurement of serum creatinine level. These are Cockcroft-Gault and MDRD formulas [7]. Equations are shown in (1) and (2)

2.1.1 Cockcroft-Gault Formula

Cockcroft-Gault formula is given in (1). The method uses serum creatinine value to calculate GFR. However a special analysis should be done for serum creatinine value. Since physicians want the analysis in extraordinary situations and it is not effective to perform the analysis routinely, (1) is not suitable for normal cases and our data set.

$$GFR = \frac{(140-age)xBody\ Weight\ (kg)}{72xSerum\ Creatinine\ \left(\frac{mg}{dl}\right)} \quad (1)$$

2.1.2 MDRD Formula

The MDRD formula is based on the patient's age, gender, and measured creatine value. In this study, MDRD formula was used since the data set is compatible with it. Since the population consists of patients in Karabük and its surroundings, while calculating the GFR value, the condition of the patient being black is neglected [6].

$$GFR = 186 \times (Serum\ Creatinine)^{-1.154} \times (age)^{-0.20} \times gender\ coefficient \times race\ coefficient \quad (2)$$

In Equation (2):
Gender coefficient is 0.742 for women and 1.0 for men.
Race coefficient is 1.272 for black race and 1.0 for white race.

3. Methodology

Since there is no determinant class information such as disease outcome in our data set, a new column named Disease Stage was created according to GFR value. Numbers from 1 to 5 are given from the beginning to the critical value of disease stage in the created column. Therefore, since the class numbers and the classes of test data are known, supervised learning algorithms became available as data set has target data. In the classification process, performance of two different methods (logistic regression and artificial neural network) were compared.

3.1 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. (see Table2)

Table 2. Confusion matrix [9].

Projected Class	Class 1	Class 2
Class 1	**a**	**b**
Class 0	**c**	**d**

a: TP (True Positive) *b: FN (False Negative)*
c: FP (False Positive) *d: TN (True Negative)*

3.1.1 Accuracy - Error rate

Accuracy rate is the most prominent and simple method for measuring model performance. It is defined as 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 defined as 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 (3) and (4)

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

$$\text{Error rate} = \frac{FP + FN}{TP + FP + FN + TN} \tag{4}$$

3.1.2 Precision

The precision is defined as the ratio of correctly predicted examples over the total number of positive examples. It is calculated by using (5).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

3.1.3 Sensitivity

Sensitivity given in (6), defines the ratio of correctly classified positive samples over the total number of positive samples:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{6}$$

3.1.4 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 (7).

$$\text{F} - \text{criteria} = \frac{2 \times \text{Sensitivity} \times \text{Precision}}{\text{Sensitivity} + \text{Precision}} \tag{7}$$

3.1.5 Support

Support is the number of examples of the real answer in that class.

3.2 Artificial Neural Network (ANN)

ANN is the logical software developed to perform the basic functions of the brain such as learning, remembering, and generalizing the information by imitating the working mechanism of the human brain. ANN are synthetic structures that mimic biological neural networks.

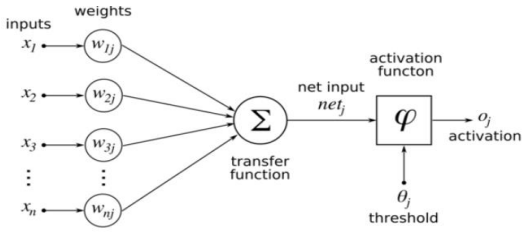


Fig. 1. Model of ANN [10].

Artificial neural networks are composed of inputs, weights, aggregation function, and actuation function. Figure 1 shows the ANN model [10].

Inputs: Information from an outside world into an artificial cell. These are

determinedby the examples that the network wants to learn.
Weights: It shows the importance of the information coming to an artificial cell and its effect on the cell.
Transfer Function: This function calculates the net input to a cell. Different functions are used for this. The most common is the weighted sum.
Activation Function: This function determines the output that the cell will produce in response to this input by processing the net input to the cell. The activation function is usually chosen as a nonlinear function.
Output of the Cell: It is the output value determined by the activation function. The produced output is sent to the outside world or another cell. The cell can also send its output as input to it.

3.3 Logistic Regression (LR)

Logistic Regression is a data analysis method to classify binary dependent variables (Yes / No, Male / Female, Fat / Thin, etc.) which has relations one or more independent variables. It is used to classify categorical or numerical data [11].

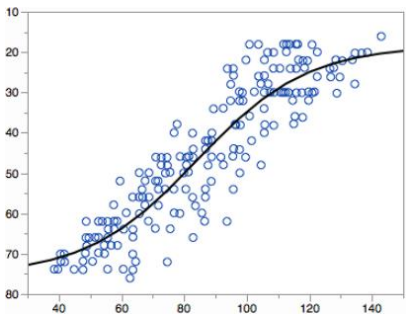


Fig. 2. Classification with LR [11].

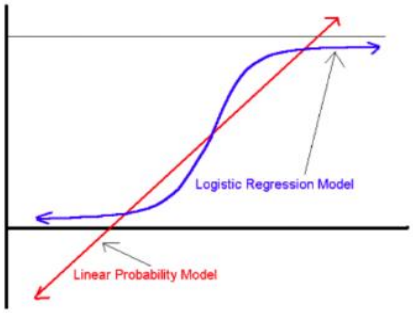


Fig. 3. Comparison of linear versus logistic regression [11]

Figure 2 represents how LR algorithm classifies entry visually. It is widely used in linear classification problems. Therefore, it is very similar to linear regression. Figure 3 shows the comparison graphic.

4. Results

The study was written in Python Language in Jupyter Notebook. Python's libraries scikitlearn, numpy, pandas, seaborn are used as development environment.Results of the study are given in Table 3 and Table 4 in terms of model performance criteria. LR and ANN methods show an average of over 90 percent success. The results shown that ANN is more successful than LR in terms of precision, F1-score ,recall, precision and support. Detailed algorithm comparison was made in Table 3 and Table4. Confusion matrixes both algorithms are given in Fig4 and Fig5 respectively.

Table 3. Results of Logistic Regression

Phases	Precision	Recall	F1- score	Support
1	0.96	0.99	0.97	111
2	0.90	0.78	0.84	46
3	0.81	0.83	0.82	30
4	0.89	0.92	0.90	36
5	0.96	0.96	0.96	57

Table 4. Results of Artificial Neural Network

Phases	Precision	Recall	F1 – score	Support
1	0.99	0.97	0.98	110
2	0.94	0.96	0.95	53
3	0.97	0.97	0.97	34
4	0.97	0.97	0.97	30
5	0.98	1.00	0.99	53

Fig. 4. Confusion Matrix of LR

Fig. 5. Confusion Matrix of ANN

References

1. A. Kusiak, K. H. Kernstine, J. A. Kern, K. A. Mclaughlin, and T. L. Tseng, "Data Mining : Medical and Engineering Case Studies The University of Iowa," pp. 1–7, 2000.
2. N. ZIASABOUNCHI, "Bulanık Uzman Sistemler Kullanılarak Tıpta Hastalık Teşhisi," 2014.
3. "Lin, R., "An Intelligent Model For Liver Disease Diagnosis", Artificial Intelligence in Medicine, 53-62 (2009)"
4. Couser WG, Remuzzi G, Mendis S, Tonelli M. The contribution of chronic kidney disease to the global burden of major noncommunicable diseases. *Kidney Int.* Dec 2011;80(12):1258-1270. (4)
5. Levey AS, Atkins R, Coresh J, et al. Chronic kidney disease as a global public health problem: approaches and init initiatives - a position statement from Kidney Disease Improving Global Outcomes. *Kidney Int.* Aug 2007;72(3):247-259. (5)
6. *Türkiye böbrek hastaliklari önleme ve kontrol programi (2018-2023)*. 2018.
7. T.Arınsoy,Ö.Güngör,İ.Kocyiğit"Böbrek Fizyopatolojisi".
8. İ.Ergün,"Birinci Basamakta Böbrek Fonksiyon Bozukluğu Olan Hastanın Değerlendirilmesi ve Sevk"
9. E.ŞİRİN,"http://www.datascience.istanbul/2017/07/07/hatamatrisi-confusion-matrix-python-uygulama/. last accessed 2019/04/02
10. "https://en.wikipedia.org/wiki/Artificial_neural_network" last accessed 2020/05/02.
11. M.Nandu,"Introduction to Logistic Regression",https://blog.goodaudience.com/machine-learning-using-logistic-regression-in-python-with-code-ab3c7f5f3bed.last accessed 2019/04/26.
12. 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.

Self-Management Technology in Cellular Communication Systems*

Rozhgar Dhyab[1] and Muhammed Talo[2]

[1] Faculty of Technology, Software Engineering, Firat University, Elazig, Turkey
[2] Department of Software Engineering, Firat University, Elazig, Turkey
rozhgar@gmail.com

Abstract. The stations of the cellular networks distributed along the service areas are expose to adverse weather conditions. The risk of failure is high and services are likely to be out of service especially in rural areas due to unfavorable weather conditions. In this paper, we have used a self-healing mechanism to ensure the availability of the service during a failure. It is important to re-establish the network in a short time using the available resources. We have also taken care to keep a low volume data exchanged by the base stations to prevent network overload. If a cell or station fails, self-healing processes try to fill gaps as much as possible. The self-management of a network is a promising way to reduce cost and sophistication in a network infrastructure.

Keywords: Self-management, Operations and Maintenance, Operations Support Systems.

1 Introduction

Mobile companies offer new services and increase the capacity of networks through base stations to reach more costumers. Therefore, the latest technologies and methods such as signal processing, multi-carrier modulation, multiple antennas, and error correction algorithms have been adopted to provide the best performance [1].

Self-management of a network excludes the human element outside the network control loop by leaving the task of managing the network itself using the automatic manager [2]. The Autonomic Manager is a self-management component that contributes to the management of a network and maintains its services. It can also detect the internal and external environment and act locally on its assigned components [3]. The Autonomic Manager can accomplish self-management using the help of management and control tools. This process forms the Autonomic Control Loop (ACL) for the network resources [4].

The main characteristics of self-management as described by IBM are self-configuration, self-optimization, self-healing, and self-protecting [5]. Self-configuration is the ability of the system to change its settings to a high level in changing environments.

* This article was produced from Rozhgar Dhyab's Master's thesis.

Self-optimization aims to maintain the effectiveness of system operations, even in an event of an unexpected environmental change. Therefore, the system should be aware of its ideal performance and be able to measure the current performance. The self-healing process aims to detect and fix underlying faults and ensures that the system is working properly. This is accomplished by forecasting the problems that may occur and taking preventive actions to avoid failure or reduce its impact if it occurs. Self- protecting is the ability of the system to protect itself from external attacks or unwanted failures that prevents achieving its goals.

The concept of healing (recovery) within cellular telecommunications consists of outage prediction, outage detection, and compensation for loss through the adjacent cells [6]. A cell is considered to be out of service, when user equipment (UE) within the considered cell area, cannot establish or maintain all or a partial set of holders due to a hardware or software failure of a base station. This defect may be at the radio level or the transport level. The effective management of cellular is described in three stages to explain the different cases of out of service cells [7]. The schematic representation of these stages is shown in Fig. 1.

- Predicting the stage of the cells going out of service.
- The detection phase of inactive cells.
- The stage of compensation for the impact of the loss.

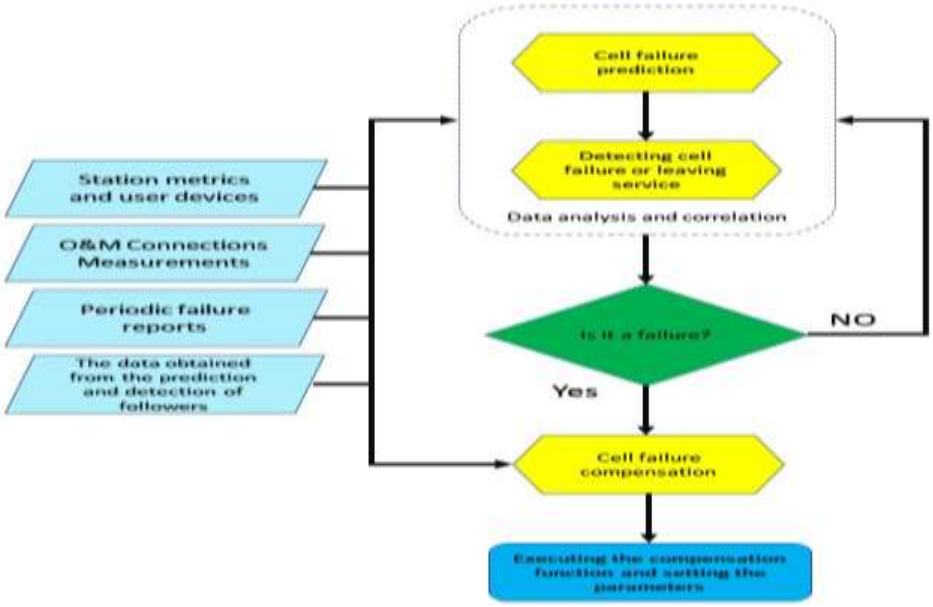


Fig. 1. The flowchart of the cell efficiency management function.

Predicting out of service cells provides early warning and speeds up the actual detection of cell loss. Detecting cell departures give a starting signal to take appropriate action

for compensation processes. The forecasting and detection stage makes decisions based on the analyzes and correlations of the different measurements from the network. In the compensation phase, any cell of the adjacent cells is counted as the ideal candidate to compensate for the loss. The settings are changed to carry out the compensation process. The prediction and detection functions are included in the current network management systems [7].

The compensation is done automatically by adjusting network parameters to improve the coverage and performance to meet operator and business requirements such as the number of subscribers, capacity, data rate, etc. The automated restructuring results from cell loss mitigate either completely or to the extent possible. There is an upper limit to the degree of compensation that is already achieved within the capabilities of neighboring cells. The resetting is necessary to reduce the poor performance of radio carriers. This must happen without delaying the permissible period of compensation to reduce the impact of the coverage gap [8].

The compensation process should be examined in detail including the frequency plan, statistical distribution, and the ability of each of the cells considered to be compensated for out-of-cell. The aim is to ensure that adjusting some parameters is not increasing interference within the network area. After that, the compensation process is carried out according to the privacy of each generation of communication systems.

2 A Practical Procedure

In this study, we have work in an area where the connection is supplied by a mobile network (Turktelekom). We have studied the local cell network stations with a rounded distribution of cells (see Fig. 2). If Site-X fails, the Site-1 and Site-2 stations adjacent to Site-X through cells A1 and B2 exit from service. They cannot fully compensate for the loss of Site-X, especially during prime time. Therefore, we should try to reset the minimum number of parameters for both A1 and B2 to alleviate the loss. Observing the lowest possible number of KPIs in real-time generalizes all cells in the region and reduces the computational burden.

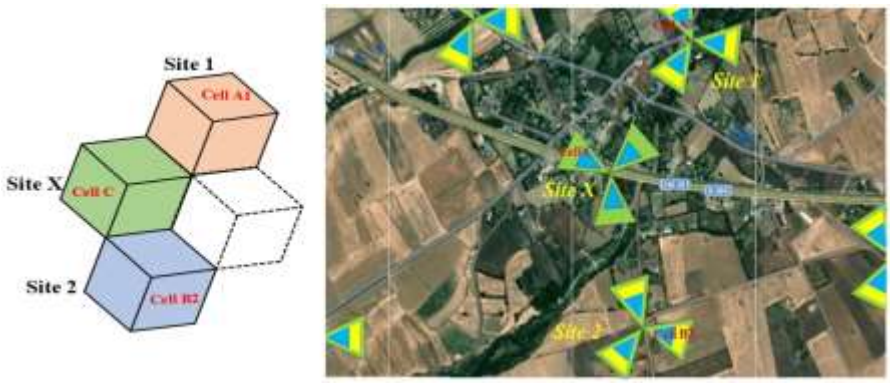


Fig. 2. The studied area is on the ground with a rounded distribution of cells.

We have carried out the experiment in two stages as follows:

1. **The first stage:** The Site-X is out of service and compensates through cells A1 and B2: The system must decide that the Site-X station is out of service and will not be servicing soon to avoid fluctuation during the regulation. Here, the coverage zone of both cells A1 and B2 must be expanded to partly compensate for cell X leaving from services. The coverage extension zone is shown in Fig. 3.

Fig. 3. The First Stage: The action is taken when the Station-X goes out of service.

2. **The second stage:** The return of Station-X to the service: At this stage, it should be ensured that the cell is not taken out of service again, that is, it does not suffer as in the first stage. Besides, the parameters of cells adjacent to the Station-X must be reset to original settings to reduce the interference effect and prevent any loss as shown in Fig. 4.

Fig. 4. The second stage: The action is taken when the X station returns from service.

After confirming that the station, which is not planned to return to service in a short time (it is not working), then the compensation step begins. The stations that accept the new settings are reset and put the control in some way to compensate. A reset is performed to ensure that the station which is out of service will not return to work again. When the station receives the new settings, it continues to work on the current settings until the station returns to operation, or until the cell indicators running in the new settings are unable to serve more users as a result of the reset. Fig.5 shows the organization of the compensation algorithm.

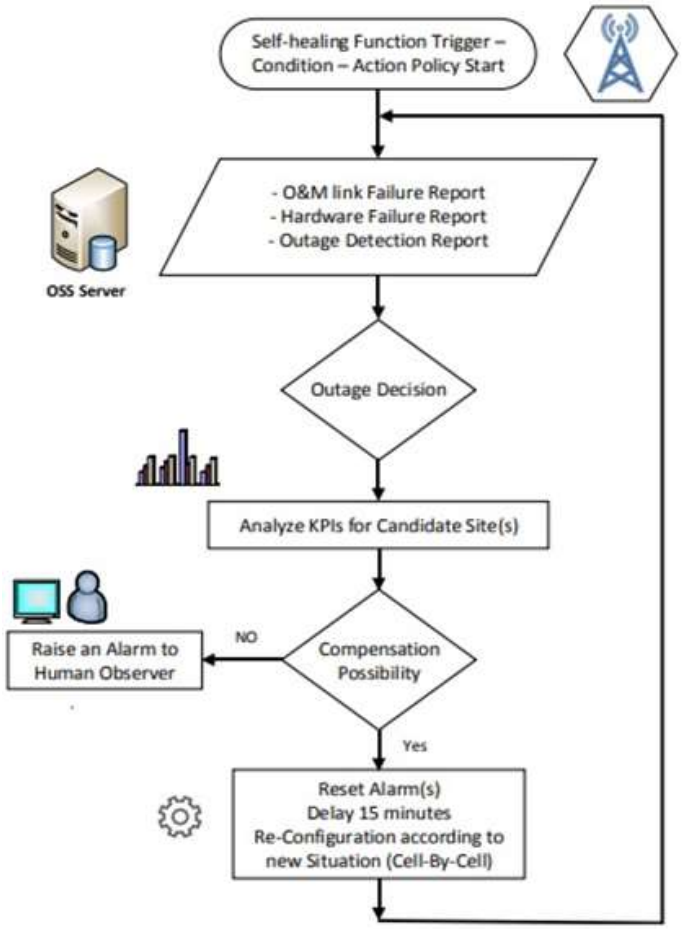


Fig. 5. The structure of the reset algorithm.

Each station has two connections that hold and control user communication and signal information. This is called the Monitoring and Operation (M&O) link. Using the M&O connection, station parameters can be adjusted and monitor performances. When any station is out of service, it can be controlled via this link. While all of the information

needed by the network cannot be imported, some can be brought in, the remaining largest part is obtained as a result of the experiments applied to the network.

The privacy of each network is implemented by taking into account the area it covers and the country it serves. The user behavior determines planning and development. Fig.6 shows 24-hour user behavior for one of the local cellular telephones to indicate privacy depending on call traffic.

Fig. 6. The typical 24-hour local area network call adapter (TCH) performance indicator for the BSC controller.

It can be seen that the load of call traffic increases gradually from the early morning hours and a decrease in the afternoon. It then decreases significantly during evening hours. Finally, it returns to the lowest level. The peak period is the range of [12:00 - 14:00]. There are also some minor differences regarding the movement of the load for a weekend as shown in Fig. 7.

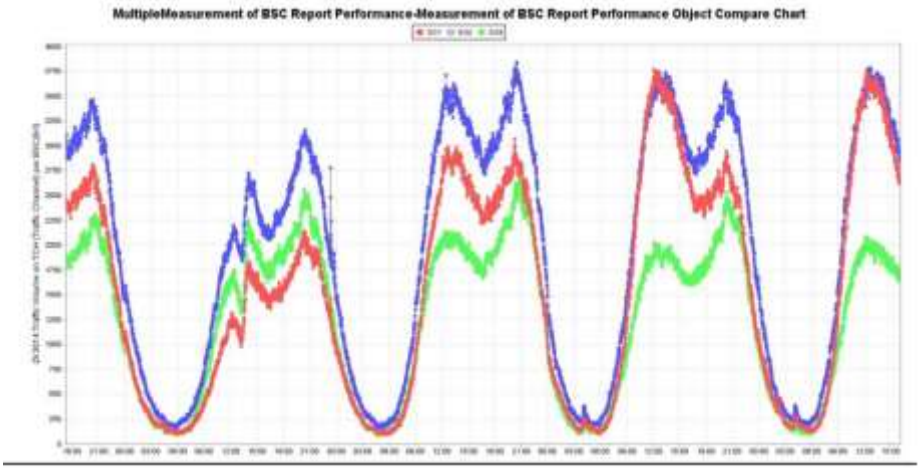


Fig. 7. A special case of TCH over 72 Index for multiple BSC controllers.

The network behavior display in Fig. 6 and Fig. 7 is important for network adjustments. Because the operations to expand the central network or to reset some parts of it may require service outages. Therefore, the network adjustments must occur when the load is at the bottom level. Hence, building a self-system requires more than benchmarking and implementing required operations using the policies. Moreover, the addition of other features or tasks to the work of the network also requires increasing the number of performance indicators. As a result, we have figured out that the burden of the self-management network processing is adding to the current network operations little by little. Monitoring and control operations must take place through the management and control servers linked to the network.

These servers are located within the Operations Support Systems (OSS) section to communicate with all network elements. This section works with the Business Support System (BSS) to define and support various types of communication services [9].

The self-management component can be placed easily within the section of the network. Hence it is supervising the monitoring information using the function of prediction, detection, and compensation in the case of self-healing. Although the general structure of cellular networks has undergone many changes within successive generations, the OSS division still uses the same features. The disadvantages of this study are as follows:

- Monitoring information and performance indicators for all components of the network need a great time to collect and analyze them before making decisions.
- A large amount of data is a major burden on OSS servers, as long as self-control processes require continuous monitoring, analysis, and planning.
- The cellular networks in Turkey have their own way of controlling the equipment. Because of the diversity of cellular network installations (base stations, central network equipment, transport equipment, etc.), the different control methods and performance indicators are used for the cellular networks (Turktelekom, Turkcell, and Vodafone).

3 Conclusion

Building a self-system needs to build a system of analysis and correlation of data following the performance indicators for the network and the behavior of users. In this paper, we have used the adaptation processes to achieve the goals of service availability and quality through self-healing. If a cell or station fails, the self-healing processes work to fill the resulting gaps as much as possible. We have also discussed the essential implications of self-managers to achieve recovery.

References

1. Weyns, Danny, Gowri Sankar Ramachandran, and Ritesh Kumar Singh. "Self-managing internet of things." In *International Conference on Current Trends in Theory and Practice of Informatics*, pp. 67-84. Edizioni della Normale, Cham, (2018).
2. Herrmann, Klaus, Gero Muhl, and Kurt Geihs. "Self management: the solution to complexity or just another problem?." IEEE distributed systems online 6, no. 1 (2005).
3. Huebscher, Markus C., and Julie A. McCann. "A survey of autonomic computing—degrees, models, and applications." ACM Computing Surveys (CSUR) 40, no. 3 (2008): 1-28.
4. Computing, Autonomic. "An architectural blueprint for autonomic computing." IBM White Paper 31, no. 2006 (2006): 1-6. Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016).
5. Qin, Hang, and Li Zhu. "Subject oriented autonomic cloud data center networks model." Journal of Data Analysis and Information Processing 5, no. 3 (2017): 87-95. Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010).
6. Osifeko, Martins O., Olufunmilayo I. Sanusi, Olaolu Folorunsho, and Raphael O. Abolade. "A Genetic Algorithm-based Approach for Cell Outage Compensation in Self-Healing Networks."
7. Shaik, Altaf, Ravishankar Borgaonkar, Shinjo Park, and Jean-Pierre Seifert. "On the impact of rogue base stations in 4g/lte self organizing networks." In Proceedings of the 11th ACM Conference on Security & Privacy in Wireless and Mobile Networks, pp. 75-86. 2018.
8. Jiang, Dajie, and Guangyi Liu. "An overview of 5G requirements." In 5G Mobile Communications, pp. 3-26. Springer, Cham, 2017.
9. Barona Lopez, Lorena Isabel, Angel Leonardo Valdivieso Caraguay, Marco Antonio Sotelo Monge, and Luis Javier García Villalba. "Key technologies in the context of future networks: operational and management requirements." Future Internet 9, no. 1 (2017): 1.

Investigation of Tower Grounding Resistance Effect for Lightning Overvoltage in 154 kV Transmission Line using ATP-EMTP

Mustafa ŞEKER[1][0000-0002-3793-8786]

[1] Sivas Cumhuriyet University, Hafik Kamer Ornek Vocational School of Higher Education, Department of Transportation Services, Rail Systems Management Program, Sivas/Turkey
mustafaseker@cumhuriyet.edu.tr

Abstract. Lightning overvoltage caused by atmospheric phenomena leads to significant damage and long-term outages in the electrical power systems. In this study, overvoltage occurring on the insulator as a result of lightning strikes on the pylon towers, which are commonly used in 154 kV power transmission lines, were investigated by using the Alternative Transient Program (ATP). In the analysis, the multistory transmission tower model was used to define the tower model. The lightning waveform was modeled as a Heidler function with a lightning impulse of 8/20 μs. The voltage levels on the insulators in the event of lightning strikes on the tower were evaluated based on various positions and lengths of the tower grounding electrode.

Keywords: Lightning overvoltage, Alternative Transient Program (ATP), Tower earth resistance, Pylon Tower.

1 Introduction

Lightning strikes occur when the voltage between the cloud and the earth exceeds the dielectric strength due to atmospheric phenomena [1]. Overvoltage caused by lightning strikes on electrical systems causes significant malfunctions and long-term power outages in system equipment [2]. Therefore, evaluating the effects of lightning overvoltage on electrical power systems is very important in terms of ensuring the stability of the system. Lightning strikes on transmission lines are common events in power systems. Lightning strikes on the transmission lines occur by a lightning strike on the phase conductor, protection conductor or tower. The most common situation is the lightning strike on the tower or the protective conductor [3].
Lightning waveforms can be measured using special tower systems. However, the data measured from these towers contain noise, reflection, and resolution (accuracy) problems [4]. To overcome these problems, an artificial waveform is used to define the

lightning waveform. The artificial waveform can be generated in high voltage laboratories. However, the construction of the system that generates an artificial waveform involves high costs [5]. Today, advanced simulation programs allow modeling lightning overvoltage in the digital environment. Alternative Transient Program (ATP), which is based on the study of Dommel, has been widely used in modeling electrical systems and lightning waveform [6].

In the modeling of the power towers, firstly, the geometric structure of the tower was evaluated considering the electromagnetic field effect. A mathematical model was designed using geometric shapes.

In his experimental studies, Kawai has established the relationship between the conductor's characteristic impedance and wave propagation speed [7]. Similarly, Chislom and Chow also conducted studies on the characteristic impedance and propagation constant of conductors [8]. On the other hand, Ishii developed the Multistory Transmission Tower Model using numerical methods considering that the protection conductor should also be taken into account in the modeling of the tower structure [9].

In this study, the effects of overvoltage occurring in the event of a lightning strike on the pylon tower used in 154 kV power transmission lines were evaluated taking into account the tower grounding resistance and grounding position. The lightning current waveform was modeled as a Heidler function with a lightning impulse of 8/20 μs and an amplitude of 10 kA. The characteristic of the transmission line was defined using the JMarti frequency-dependent model [10]. The non-uniform Multistory Transmission Tower Model was used for modeling the pylon tower. The voltage levels generated by the lightning overvoltage on the isolator were evaluated depending on the position and length of the grounding electrode used in tower grounding.

2. Modeling of The Electrical System

System equipment should be modeled accurately to analyze the effects of lightning overvoltage using a simulation program. Therefore, the parameters such as the characteristic feature of the transmission line, geometric structure of the tower, grounding resistance, the lightning current waveform should be taken into account. The determination of circuit parameters used in the simulation in this study is explained in detail in this part.

2.1. Characteristic features of the Transmission Line

The characteristic features of the transmission line examined in this study are presented in Table 1. The geometry of the transmission line is defined for the tower structure shown in Figure 1 by taking into account the heights of conductors above the ground and the distance between the conductors.

Table 1: Characteristic features of the Transmission Line.

Conductor	Type	Diameter (mm)	Resistance (Ω/km)
Phase conductor	Cardinal	30.35	0.0586
Protection Conductor	7N8	9.78	1.4625

2.2. Identification of Equivalent Circuit Model for Pylon Tower

The structure of the multistory transmission tower model of the pylon tower used in the 154 kV transmission system is shown in Figure 1. The parameters used in the dimensions of the pylon tower are presented in Table 2.

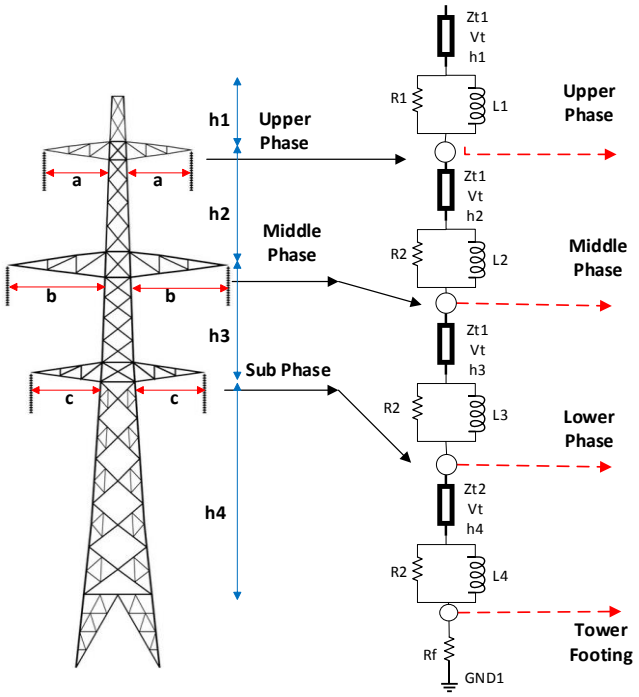


Fig. 1. Geometric structure of double circuit pylon tower and Multistory Transmission Tower Model

Table 2. Dimensions of the pylon tower (m)

Parameters	a	b	c	h1	h2	h3	h4
Length (m)	3.20	4.10	3.50	3.15	4.15	4.15	17.90

The equations (1-4) are used to calculate the parameters of the multistory transmission tower model [9, 11, 12].

$$R_i = \frac{-2.Z_{t1}.ln\sqrt{\gamma}}{h_1 + h_2 + h_3}.h_i \quad (i = 1-3) \tag{1}$$

$$R_4 = -2.Z_{t2}.ln\sqrt{\gamma} \tag{2}$$

$$L_i = \alpha . R_i . \frac{2.H}{V_t} \qquad (i = 1 - 4) \tag{3}$$

$$H = h_1 + h_2 + h_3 + h_4 \tag{4}$$

where;
Z_t : Tower surge impedance,
V_t : Surge propagation velocity,
γ : Attenuation coefficient,
α : Damping coefficient,
R : Damping resistance,
L : Damping inductance.

Surge propagation velocity seen in the lightning waveform is considered to be approximately equal to the speed of light (Vt = 300.000 km/s).

2.3. Modeling of the lightning current waveform

Lightning current waveforms were modeled using lightning current waveforms of 1.2/50 µs, 10/350 µs, and 8/20 µs which were defined in international standards [13]. The lightning waveform of 8/20 µs was used in this study. The maximum amplitude value of the lightning current was chosen to be 10 kA, which has been statistically the most frequent situation. The lightning current waveform can be modeled using a current source and a resistance parallel to the current source, as shown in Figure 2 [14]. The parallel resistance value of 400 Ω was used in the model [15].

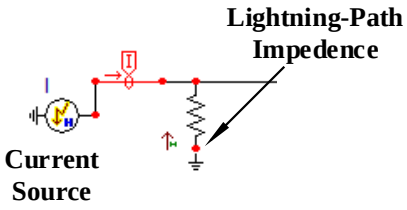


Fig. 2. Lightning Strike Heidler Model.

The mathematical expression of the Heidler model used in determining lightning current waveform is defined by the equation (5),

$$i(t) = \frac{I_0}{\eta} . \frac{(\frac{t}{\tau_1})^n}{(\frac{t}{\tau_2})^n + 1} . e^{\frac{-t}{\tau_2}} \tag{5}$$

where I_0 denotes the peak value of the lightning current, τ_1 denotes the constant of time depending on the current rise time, τ_2 denotes the constant of time depending on the delay time of the current, n denotes the steepness factor of the current. The steepness

factor is determined by the equation (6) and determines the time for the lightning current to reach the maximum value. The lightning current waveform of 8/20 μs with a maximum amplitude of 10 kA is shown in Figure 3.

$$\eta = e^{\left[-\left(\frac{\tau_1}{\tau_2}\right).\left(n.\frac{\tau_2}{\tau_1}\right)\right]^{\frac{1}{n}}} \tag{6}$$

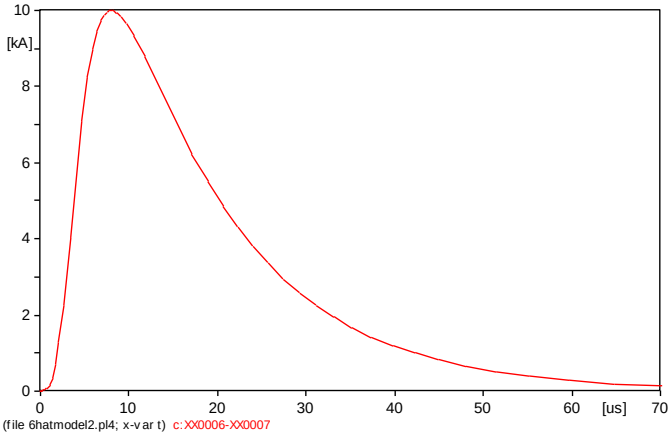


Fig. 3. The lightning current waveform of 8/20 μs.

2.4. Determination of Grounding Resistance

The towers carrying the power transmission lines are grounded using electrodes at the bottom of the tower. In the case of low frequency, the ground resistance varies depending on whether the electrode is in a horizontal or vertical position. Due to the high current caused by lightning overvoltage, the ground resistance is defined by the lumped circuit high frequency [16].

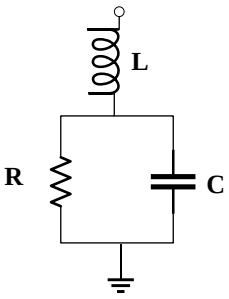


Fig. 4. Lumped circuit high-frequency model for ground resistance

In case the grounding electrode is horizontal or vertical, the lumped circuit high-frequency circuit models are shown as in Figure 5.

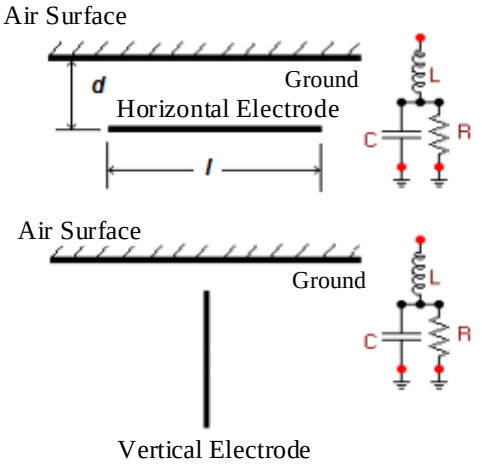


Fig. 5. The lumped circuit high-frequency model with horizontal and vertical electrode structure for ground resistance.

According to Dwight, the mathematical expression of the resistance is defined by equation (7) in the case of using the ground electrode vertically [15]. According to Sunde, if the ground electrodes are placed horizontally, the resistance is defined by the equation (8) [17].

$$R = \frac{\rho}{2.\pi.l}\left[ln\left(\frac{4l}{a}\right) - 1\right] \tag{7}$$

$$R = \frac{\rho}{\pi.l}\left[ln\left(\frac{2l}{\sqrt{2ad}}\right) - 1\right] \tag{8}$$

In these equations, ρ denotes the ground, l denotes the length of the ground resistance, a denotes the radius of the ground electrode, d denotes the depth of the electrodes under the ground, and $l > a$ and $l > d$. The capacitance value presented for the ground resistance in the lumped circuit high-frequency model is defined by the equation (9) depending on the resistance of R [16].

$$C = \frac{\rho.\varepsilon}{R} \tag{9}$$

In equation (9), ε is the permeability of the soil. In the lumped circuit high-frequency model given for the ground resistance, L, which is expressed using the same mathematical equation both for the horizontal and vertical grounding electrodes, is defined by the equation (10) [18].

$$L = \frac{\mu . l}{2\pi}\left[ln\left(\frac{2l}{a}\right) - 1\right] \quad (10)$$

μ in the equation (10) stands for the permeability of the soil.

3. ATP Model of 154 Kv Pylon Towers

The ATP model of the power transmission line for the pylon tower designed by considering the system parameters presented in Part 2 is presented in Figure 6. In the model, the length of the transmission line was assumed to be one kilometer. In the analysis, the capacity value of 80 pF, which is the equivalent of the insulator chain, was used for the insulators [19].

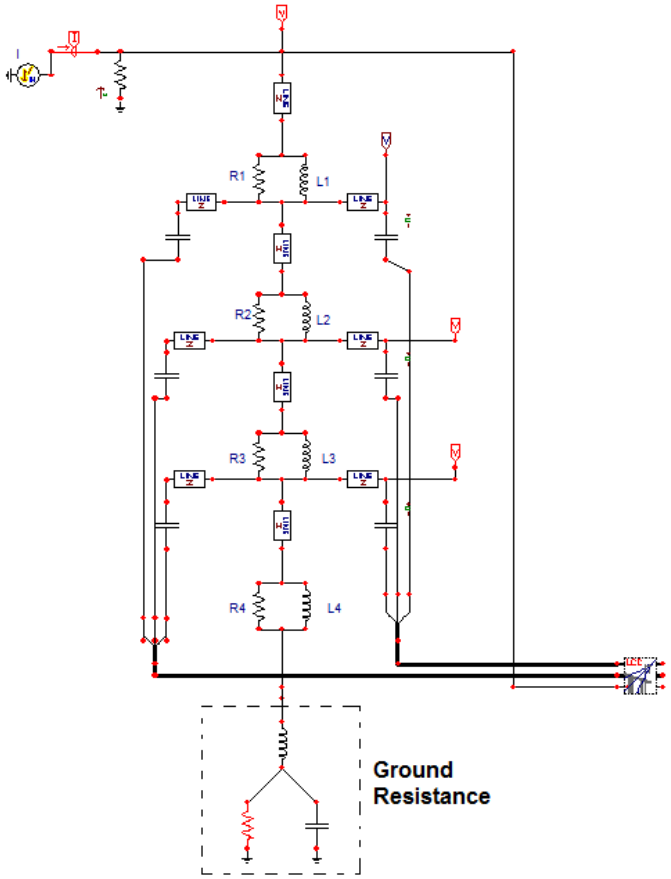


Fig. 6. ATP model of 154 kV transmission line with pylon tower.

The following parameters were used in the calculations: electrode diameter = 2.5 mm^2, d=0.8 m, ρ=50, and $\mu=10^{-3}$. The circuit parameters of the lumped circuit high-frequency model calculated based on the various electrode lengths are given in Table 3.

Table 3. The circuit parameters of the lumped circuit high-frequency model calculated based on the various electrode lengths for horizontal and vertical electrode positions.

Electrode Length (m)	R(Ω) Horizontal	C (F) Horizontal	R (Ω) Vertical	C (F) (Vertical)	L (H)
1	32.4292	0.0154	20.7313	0.0241	0.0054
2.5	15.8883	0.0315	14.1258	0.0354	0.0171
5	9.0473	0.0553	9.2693	0.0539	0.0397

4. Simulation Results

In the transmission line model of the pylon tower presented in Figure 6, when a lightning current impulse of 10 kA is applied to the top of the tower, insulator voltages for different electrode lengths and electrode positions are presented in Figure 7, Figure 8, and Figure 9, respectively.

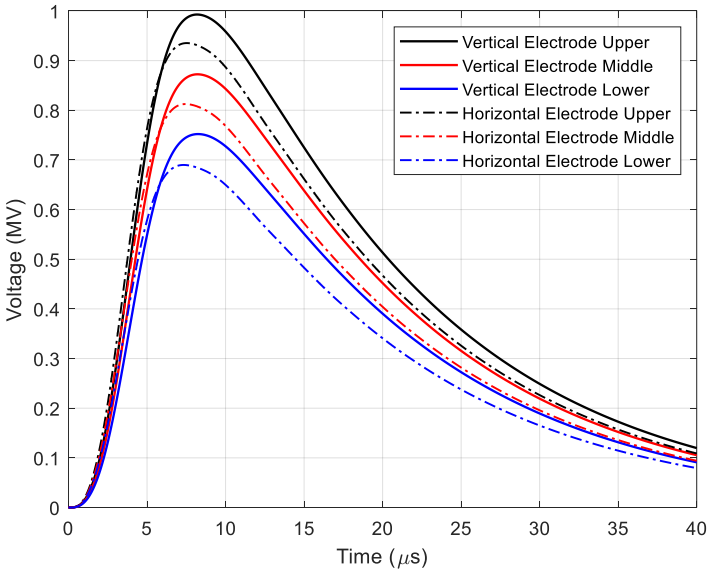


Fig. 7. Isolator voltage levels caused by the lightning current of 10 kA striking the tower in case of using horizontal and vertical grounding electrode of 1 meter.

The charts obtained as the result of the analysis reveals that horizontal grounding seems to reduce the voltage levels that occur on the insulator when the grounding electrode of 1 meter is used in pylon towers. In the case of using the grounding electrode longer than 2.5 meters, the isolator voltage levels seem to be equal for horizontal and vertical grounding. Therefore, horizontal grounding is recommended for pylon towers grounded using electrodes shorter than 2.5 meters.

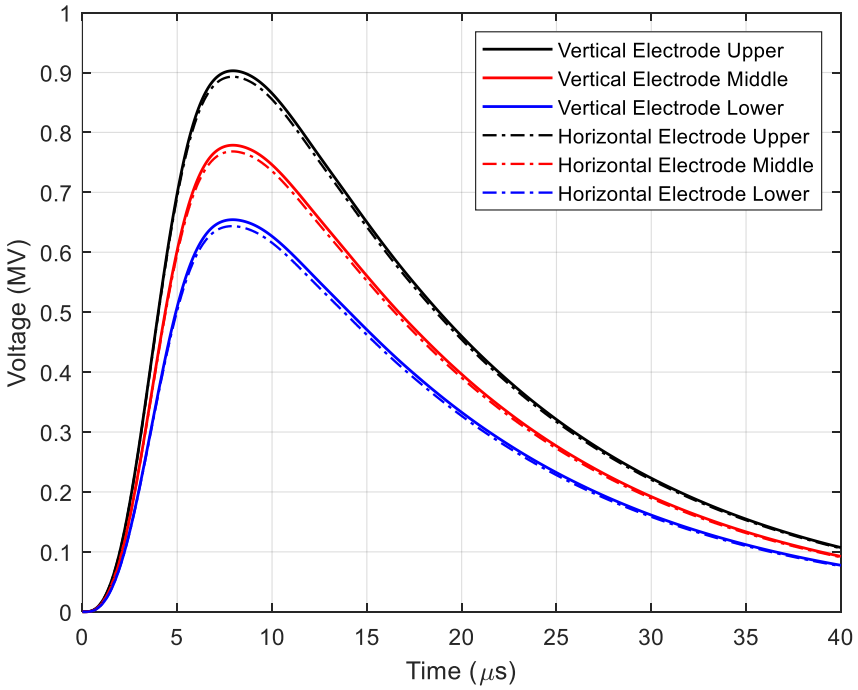


Fig. 8. Isolator voltage levels caused by the lightning current of 10 kA striking the tower in case of using horizontal and vertical grounding electrode of 2.5 meters.

5. Conclusion

Lightning overvoltage caused by atmospheric phenomena cause outages in electrical systems and significant damages to system equipment. In this study, overvoltage that may occur on the insulator in case of the pylon towers used in the 154 kV transmission system which is exposed to a lightning strike of 10 kA was analyzed using the Alternative Transient Program (ATP) considering the tower ground resistance and its position. The ground resistance was modeled using the lumped circuit high-frequency model and the pylon tower was modeled using the multistory transmission tower model. The results of the study reveal that horizontal grounding reduces the effect of lightning overvoltage on the isolators in the case of grounding using a ground electrode shorter than 2.5 meters. Horizontal grounding is recommended when grounding electrodes shorter than 2.5 meters are used in the transmission lines with pylon towers.

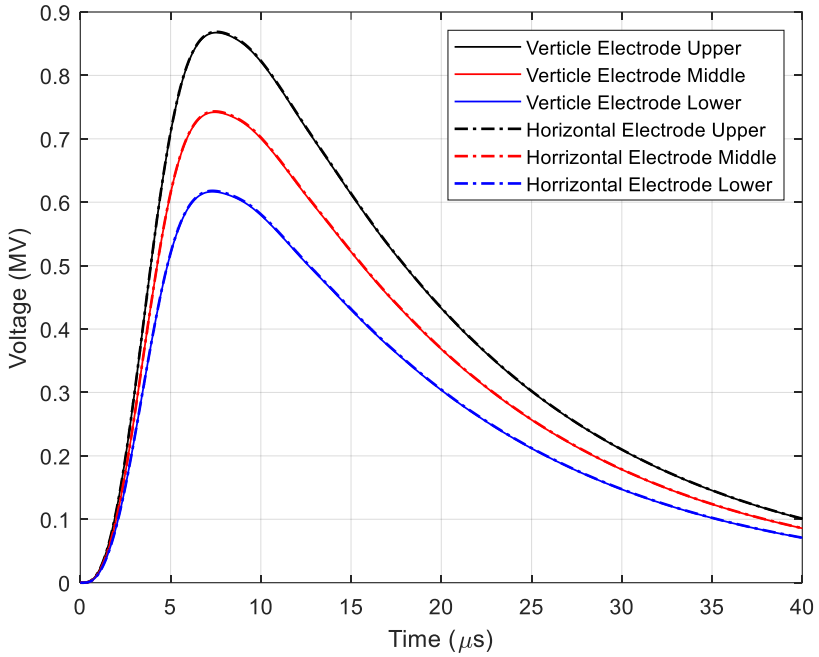


Fig. 9. Isolator voltage levels caused by the lightning current of 10 kA striking the tower in case of using horizontal and vertical grounding electrode of 5 meters.

References

1. Minnaar, U.J., Gaunt, C.T., Nicolls, T.F.: Characterisation of power system events on South African transmission power lines. Electrical Power System Research, Vol. 88, pp. 25-32, 2012.
2. Kaygusuz, A., Mamiş, M.S., Akin, E.: Yıldırım düşmesi nedeni ile izolatörler üzerinde oluşan aşırı gerilimler, Elektrik-Elektronik-Bilgisayar Mühendisliği 10. Ulusal Kongresi, Istanbul, 2003.
3. Ueda, T., Ito, T., Watanabe, H., Funabashi, T., Ametani, A.: A Comporasion between Two Tower Models for Lightning Surge Analysis of 77 kV System, International Conference on Power System Technology (PowerCon), 4-7 December, Perth, WA, Australia, 2000
4. Greenwood, A.: Electrical Transients in Power System. *New York: John Willey&Sons*, ISBN:978-0471620587, 1991.
5. Dommel, H.W.: EMTP Theory Book. *Bonneville Power Administration, Conversion into Electronic Format by Canadian/American EMTP user group*, 1995.
6. Kawai, M.: Studies of Surge Responce on a Transmission Line Tower, IEEE Transaction, PAS-83, pp. 30-34, 1964.
7. Chislom, W.A., Chow, Y.L., Strivastova, K.D.: Lightning Surge Response of Transmission Towers, *IEEE Transaction, PAS-102, pp. 3232-3242*, 1991.
8. Ishii, M., Kawamura, T., Kauno, T., Ohsaki, E., Shikawa, K., Murotani, K., Higuchi, T.: Multistory Transmission Tower Model for Lightning Surge Analysis. IEEE Transaction on Power Delivery, Vol. 6, Issue. 3, July, 1991.
9. Orlando, P.H.: Alternative Transient Program: Comparison of Transmission Line Models. Revista Latinoamericana del ATP, Vol. 1, pp. 1-5, 1999.

10. Nur, Z., Imran, J., Faizuhar, M.: Modeling of 132 kV Overhead Transmission Lines by using ATP/EMTP for Shielding Failure Pattern Recognition. *Procedia Enginering, Vol 53, pp. 278-287, December*, 2013.
11. Takamitsu, I.., Ueda, T., Watanabe, H., Funabashi, T., Ametani. A.: Lightning flashover on 77-kV systems: observed voltage bias effects and analysis. IEEE Transactions on Power Delivery, Vol: 18, Issue: 2, pp. 545 550, 2003.
12. International Electrotechnical Commission (IEC) Standarts, 61643-11.
13. Ametani, A., Kawamura, T.: A Method of a Lightning Surge Analysis Recommended in Japan using EMTP. IEEE Transaction on Power Delivery, (20), pp. 867-875, 2005.
14. Bewly, B.V.: Travelling Wawes on Transmission Systems. New York Dower, 1963.
15. Dwight, H.B.: Calculation of the resistance to ground. Electrical Engineering, (55), pp. 1319-1328, 1936.
16. Sunde, E.D.: Earth Conduction Effects in Transmission Systems. 2nd edition, New York Dover, 1968.
17. Bourg, S., Sacepe, B., Debu, T.: Deep earth electrodes in highly resistive ground: frequency behaivour. IEEE International Electromagnetic Compat, in Proc., pp. 584-589, 1995.
18. Shaida, N., Jamoshid, B.T.: Lightning Simulation Study on Line Surge Arresters and Protection Design of Simple Structure, Degree of Mater of Engineering (Electric-Power) Faculty of Electrical Engineering Technology, Malaysia, 2008.

Prediction of Solar Irradiance and Photovoltaic Generation Using Artificial Neural Network

Mustafa Seker[1][0000-0002-3793-8786]

[1] Sivas Cumhuriyet University, Hafik Kamer Ornek Vocational School of Higher Education, Department of Transportation Services, Rail Systems Management Program, Sivas/Turkey
mustafaseker@cumhuriyet.edu.tr

Abstract. A prediction approach to predict the amount of solar irradiance that photovoltaic (PV) system is exposed to and its output power using Artificial Neural Network (ANN) is presented in this study. Levenberg-Marquardt Backpropagation algorithm was chosen as the training algorithm in the ANN model. The real-time winter dataset served as an open dataset by IEEE PES for Petrolina FV1/Enerq/USP system (Southeast of Brazil) for the period of 1-10 June 2014 was used as PV generation data in the analysis. The results of Feed Forward MultiLayer Perception (FFMLP) ANN topology revealed that the ANN model could be used effectively to predict solar irradiance and PV power output by considering the panel and ambient temperature, wind speed data, and time series together.

Keywords: Solar Irradiance, Photovoltaic generation, Artificial Neural Network (ANN).

1. Introduction

Due to the developments in technology, industry, and urbanization, demand for electricity as an energy source has been increasing each passing day. Most of this energy demanded is obtained from fossil fuels. We have a limited reserve for fossil fuels. Moreover, CO and CO_2 gases emitted after the consumption of fossil fuels have negative effects on living things and the ozone layer [1]. Considering all these disadvantages, the countries have been seeking alternative energy sources such as wind, solar, and geothermal energy to overcome their energy crisis and preserve the environment. Photovoltaic (PV) systems have an important advantage over the other renewable energy methods in terms of wider installation areas, having no noise pollution and gas emissions, and using solar energy as an unlimited energy source [2-5].

Despite the technological advances today, the efficiency of PV cells is around 11-28% [6]. Therefore, PV systems should be designed to achieve maximum efficiency. Several environmental factors such as solar irradiance, the indoor and outdoor temperature of the panel, shading and clouding effect, humidity, wind are effective in the efficiency of PV systems. A reliable prediction methodology is required to evaluate the performance of PV modules and ensure effective planning and energy management

taking into account environmental factors [7]. The Artificial Neural Network (ANN) method has been very popular in prediction applications due to its high processing speed, simplicity of application, and low cost [8].

Various methods based on ANN applications for solar irradiance have been studied by many researchers. Bora et al. reported that the outputs of PV modules can be predicted using ANN [9]. Ceylan revealed that the ANN application can be used to determine the module temperature in photovoltaic panels [10]. Priya predicted global solar radiation using ANN in his study [11]. On the other hand, the researchers discussed the use of ANN methods to predict the amount of daily global solar radiation (GSR) [12-13]. Apart from these studies, many researchers conducted studies on modeling solar irradiance using ANN [14-21].

This study presents an ANN-based approach to predict solar irradiance and PV output power PV systems using meteorological data. The winter dataset served as an open dataset by IEEE PES for Petrolina FV1/Enerq/USP system (Southeast of Brazil) for the period of 1-10 June 2014 was used as PV generation data in the analysis. Day, month, hour, minute, outside temperature, panel temperature, and wind speed data in the dataset have been defined as training sets in modeling the artificial neural network. The solar irradiance and PV generation data calculated using the Feed Forward ANN (FFANN) model were assessed by statistically comparing them with the actual values; thus, the proposed model was tested.

2. Artificial Neural Network for Predicting Solar Radiation and Solar Output

ANN was developed inspired by the working principle of the human brain. It is a data processing technique that has emerged as a result of artificially imitating the nerve cells in the human brain and applying them to computer systems to solve complex problems. ANN performs the learning process using the examples provided to it [22-23].

The general structure of the artificial neurons that make up the main structural unit of ANN consists of five main parts: inputs (x_i), weights (w_i), summation function (combining function), activation (transfer) function, and output. Products obtained by multiplying each input by weight are simply summed up with the threshold value and the activation function is used to create the result, thus the output is determined. The learning ability of an artificial neuron depends on the appropriate adjustment of the weights within the selected learning algorithm [24].

ANN analyses presented in this study were carried out using Matlab software. Three major climate variables, such as outside temperature, panel temperature, and wind speed, and four date-time variables such as the month, day, hour, and minute data were defined as input variables in the structure of the model. The parameters of solar irradiance and PV output power were defined as the output variables in the model. Consisting of 7 input parameters, 10 hidden layers, and 2 output parameters, FFMLP ANN topology was defined using the Feed Forward MultiLayer Perception (FFMLP) network structure as shown in Figure 1.

The proper transfer function for neurons was defined as a logistic sigmoid function -$f(z_i)$ as in Equation 1. The z_i term defined in the second equation denotes the total

weight of the inputs. The x_j term denotes the signal coming directly from the jth neuron. The w_{ij} term denotes the weights directly combined from jth neuron to ith neuron.

$$f(z_i) = \frac{1}{1 + e^{-z_i}} \tag{1}$$

$$z_i = \sum_{j=1}^{7} w_{i,j} x_j + \beta_i \tag{2}$$

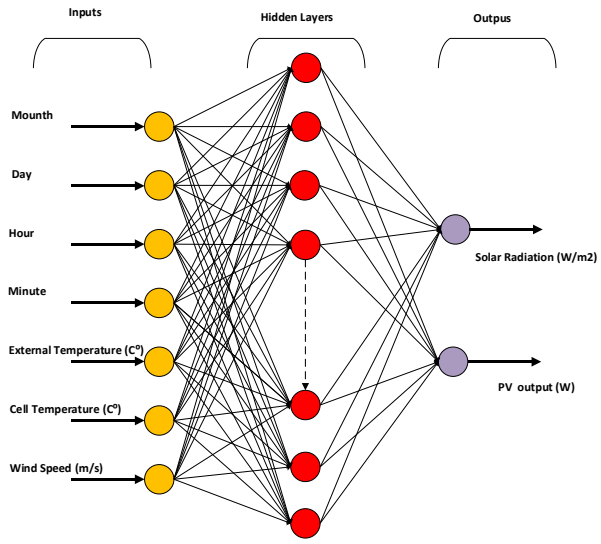


Figure 1. FFMLP ANN topology used in the prediction of solar irradiance and PV output power.

3. Results and Discussion

The winter dataset served as an open dataset by IEEE PES for 2.550 kWp Petrolina FV1/Enerq/USP system (Southeast of Brazil) for the period of 1-10 June 2014 was used as PV generation data in the analysis. The data for the period between 1-7 June 2014 were used as the training set, and data for 8-10 June 2014 were used as the test set [25].

The values of the Solar Irradiance and PV Output Power were predicted by performing ANN training according to the input parameters presented in Figure 1. The input data were trained using the Levenberg-Marquardt (LM) Back Propagation method. While 1730 samples were used as training data, 996 samples were used as test data.

The MSE performance obtained by prediction and the measured and predicted regression analysis for the best fit are given in Figure 2 and Figure 3, respectively. The charts of predicted and measured solar irradiance and PV output power are shown in Figure 4.

Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) values were calculated to statistically evaluate the predicted and measured values.

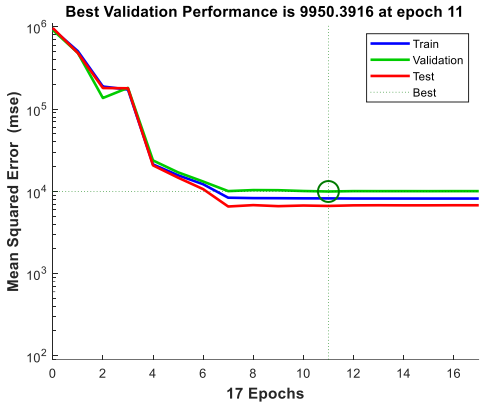


Figure 2. MSE performance of ANN Model in the prediction of Solar Irradiance and Power Output

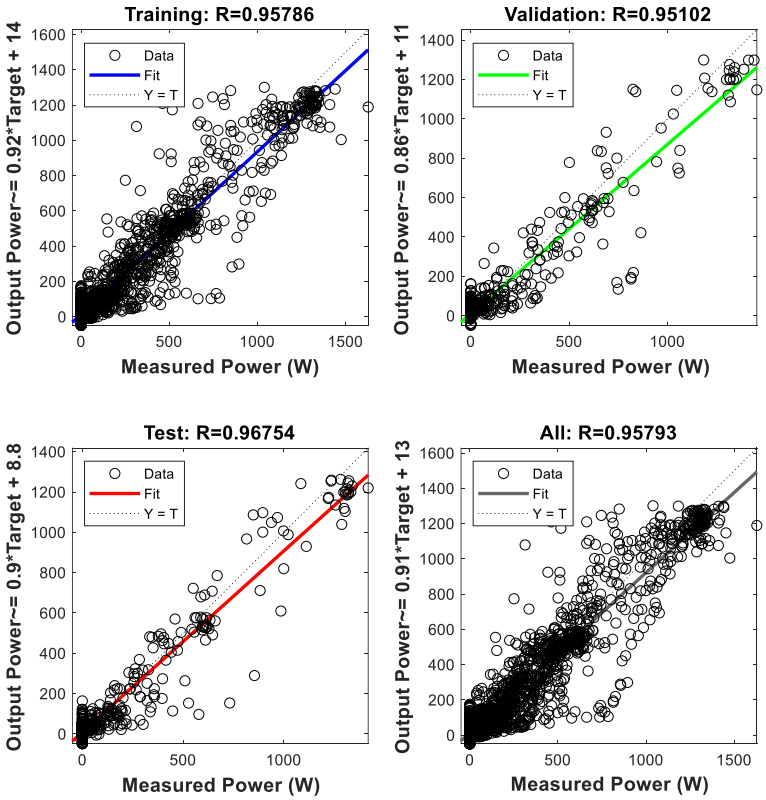


Figure 3. Comparison regression analysis between measured and predicted outputs for best fit.

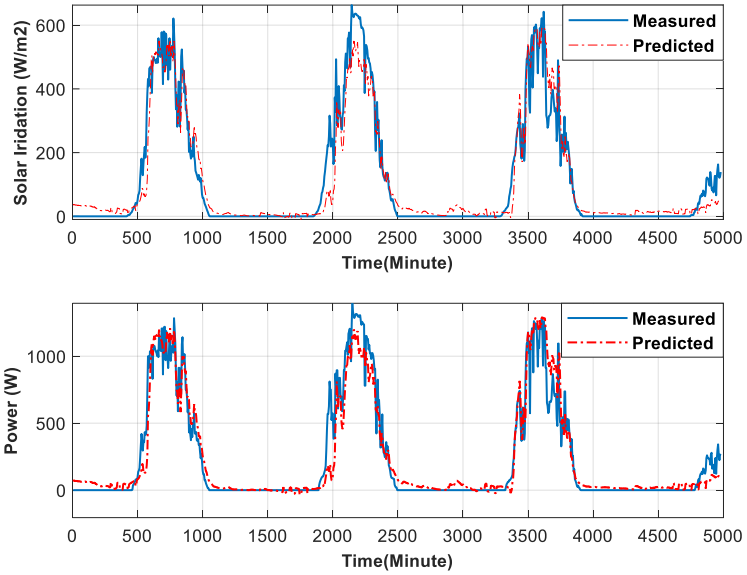


Figure 4. Comparison of the measured and predicted values of Solar Irradiance and Power Output.

$$MSE = \left[\frac{1}{N}.\sum(Measured\ Value - Predict\ Value)^2\right] \quad (3)$$

$$MAPE = \left[\frac{1}{N}\sum\frac{Measured\ Value - Predict\ Value}{Pradict\ Value}\right] \quad (4)$$

$$RMSE = \left[\sqrt{\frac{1}{N}\sum(Measured\ Value - Predict\ Value)^2}\right] \quad (5)$$

Calculated by equations (3-5), the statistical results of the measured true values and the predicted values obtained using the FFMLP ANN method are presented in Table 1.

Table 1. MSE, MAPE, and RMSE values for Solar Irradiance and PV Output Power

	MSE	MAPE	RMSE
Solar Irradiance	2.80	0.37	1.67
PV output	11.75	0.41	3.42

4. Conclusion

In this study, a prediction approach for predicting the PV output power and solar irradiance by using ANN was presented. The winter dataset served as an open dataset by IEEE PES for Petrolina FV1/Enerq/USP system (Southeast of Brazil) for the period of 1-10 June 2014 was used as PV generation data for the prediction analysis. The results revealed that the FFMLP ANN model, in which meteorological data and time series data were used together, could be used effectively in the prediction of solar irradiance and PV output power. The mean values of MSE, MAPE, and RMSE were determined to be 2.80, 0.37, and 1.67 for Solar Irradiance and 11.75, 0.41, and 3.42 for PV output power, respectively. The predictive accuracy of the presented approach can be improved by using different meteorological parameters such as humidity, pressure, latitude, longitude, and training set with more samples.

References

1. Ahmed A. Zaki Diab, Hegazy Rezk, "Global MPPT based on flower pollination and differential evolution algorithms to mitigate partial shading in bulding integrated PV system", Solar Energy, Vol. 157, pp. 171-186, 2017.
2. Jirada Gosumbonggot, Duy-Dinh Nguyen, Goro Fujita, "Partial Shading and Global Maximum Power Point Detections Enhancing MPPT for Photovoltaic System Operated in Shading Condition", 53rd International Universities Power Engineering Conference (UPEC), Glasgow-UK, 4-7 September, 2018.
3. Kashif Ishaque, Zainal Salam, Muhammad Amjad, and Saad Mekhilef, "An Improved Particle Swarm Optimization (PSO)-Based MPPT for PV with Reduced Steady-State Oscillation", IEEE Transaction on Power Electronics, Vol. 27, No. 8, August, 2012.
4. Rezk, H., "A compherensive sizing methodology for stand-alone battery-less photovoltaic water pumping system under the Egyptian climate", J. Cogent Eng. 3, 2016.
5. Rezk, H., Dousoky, G.M., "Technical and economical analysis of different configurations of stand-alone hybrid renewable power system-A case study", Reneawable Sustainable Energy Rev. Vol. 62, pp. 941-953, 2016.
6. Aygül, K., Cikan,m., Demirdelen, t., Tumay, M., Butterfly optimization algorithm based maximum power point tracking of photovoltaic systems under partial shading condition, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2019, DOI: 10.1080/15567036.2019.1677818.
7. Mekhtoup, F.B.S., PV cell temperature/PV output Relationships Homer Methodology Calculation, Ipco, Vol. 2, No. 1, pp. 1-12, 2014.
8. Dias, F.M., Antunes, A., Mota, A.M., Artificial Neural Network: A rewiev of commercial hardware, Engineering Application of Artificial Intelligence, Vol. 17, No. 8, pp. 945-952, 2004.
9. Bora. B., Gupta. K., Kumar, A., Renu, O.S., Dahiya, R., Artificial Neural Network based Modelling of PV Module to Predict the output, 4th International Conference on Advances in Energy Researchhe, held at IIT Bombay, December, 2013.
10. Ceylan, I., Erkaymaz, E., Gedik, E., Gurel, A.E., The prediction of photovoltaic module temperature with artificial neural network, Case Study Therm. Eng., Vol. 3, pp. 11-20, 2014.

11. Priya, S.S., Iqbal, M.H., Soalr irrradiance predicrtion using Artificial Neural Network, International Journal of Computer Application, Vol. 116, No.16, pp. 28-31, 2015.
12. Benghanem, A.M., Mellit, A., Alamri, S.N., ANN-based modelling and estimation of daily global solar radiation data: A case study, Energy Convers. Manag., Vol. 50, No. 7, pp.1644-1655, 2009.
13. Behrang, M.A., Assareh, E., Ghanbarzadeh, A., Noghrehabadi, A.R., The potetial of different artificial neural network (ANN) techniques in daily global solar radiation modelling, IOP Conf. Series: Journal of Physics: Conf. Series 1049-012088, 2018, doi:10.1088/1742-6596/1049/1/012088.
14. D., Elizondo, Hoogenboom, G., Mcclendon, R.W., Development of Neural network modl to predict daily solar radiation, Agric Forest Meteorol, Vol. 71, pp. 115-132, 1994.
15. Williams, B.D., Zazueta, F.S., Solar radiation estimation via neural network, In: ASAE, editör, Sixth International onference on computer in agriculture, Cancun, Mexico, 1994.
16. Mohandes, M., Rehman, S., Halawani, T.O., Estimation of global solar radiation using artificial neural networks, Renew Energy, Vol. 14. pp. 179-184, 1998.
17. Hontoria, L., Riesco, J., Zufuia, P., Aguilera, J., Application of neural networks in solar radiation fields, Obteinment of solar radiation maps, In 16th European photovoltaic for cemical engineers, Vol. 3, pp. 385-408, Amsterdam, 2000.
18. Tymvios, F.S., Jacovides, CP., Michaelides, S.C., Scouteli, C., Comporative study of Angstroms and artificial neural netwoks methodologies in estimating global solar radiaion, Solar Energy, Vol. 78, pp. 752-762, 2005.
19. Alam, S., Kaushik, S.C., Garg, S.N., Computation of beam solar radiation at normal incidence using artificial neural network, Renew Energy, Vol. 31, 2006.
20. Elminir, H.K., Azzam, Y.A., Younes, F.I., Prediction of hourly and daily diffuse fraction using neural network, as compared to linear regression models, Energy, Vol. 32(8), 1532-1523, 2007.
21. Mubiru, E.J., Banka, K.B., Estimation of monthly average daily global solar irradiation using arificial neural networks, Solar Energy, Vol. 82(2), pp. 181-187, 2008.
22. Basheer I.A., Hajmeer M., Artificial neural networks: fundamentals, computing, design, and Application,Journal of Microbiological Methods, 43: 3-31, 2000.
23. Graupe D., Princıples of artificial neural networks, (2nd Edition), advanced series on circuits and systems, 6, World Scientific Publishing Co. Pte. Ltd. 2007.
24. Cho V.A., Comparison of three different approaches to tourist arrival forecasting, Tourism Management, 24: 323-330, 2003.
25. Fernandes, F., Morais, H., Valdomiro, V., Luis Gomez, G, Vale, Z., Kagan, N., Dynamic loads and micro-generation method for a House Menegement System, Clemson University Power System Conference (PSC), Clemson, SC, USA, 2016. Doi:10.1109/PSC.2016.7462829. (https://site.ieee.org/pes-iss/data-sets/)

Vessel Segmentation using Shallow Water Equations

Fatih Nar[1][0000-0002-3003-8136] and Didem Gökçay[2][0000-0002-1101-0306]

[1] Ankara Yıldırım Beyazıt University, Ankara, Turkey
[2] Middle East Technical University, Ankara, Turkey
fnar@ybu.edu.tr

Abstract. In this study, we propose a new segmentation approach for vessel structures based on fluid flow. Vessels are medium for transportation of blood which makes segmentation using fluid flow physically plausible. So, streaming fluid regions can capture smooth and thin structures. It also provides robustness to radiofrequency field inhomogeneity and noise in images. More importantly, it can also segment disconnected areas as the fluid flow can pass through gaps. Thus, fluid flow can form a powerful vessel segmentation algorithm. In this study, we used linearized form of shallow water equations since they provide computational efficiency compared to original Navier-Stokes equations. We validated our method on a synthetic data set formed with BrainWeb database, and a clinical Magnetic Resonance Angiography (MRA) datasets. Experiments presented the accuracy of the vessel segmentation method we proposed.

Keywords: Vascular system, Shallow water equations, Vessel segmentation.

1 Introduction

In the literature, various vessel segmentation studies are proposed due to its clinical applications in human health. Vascular structures are required to determine target region in cancer treatments [1]. Artery segmentation is crucial for various clinical treatments since coronary artery disease is widest cause of death [2]. Hepatocellular carcinoma has third rank for death from liver cancer. It is treated by heating the target region using radiofrequency as nearby blood vessels act as coolers [3]. Also, accurate vessel structures allow characterization and assessment of vascular diseases by means of simulating blood flow on patient-specific 3D vessel models [4].

Noninvasive imaging modalities such as magnetic resonance angiography (MRA), magnetic resonance venography (MRV), computerized tomography angiography (CTA), and even ultrasound imaging can be used to acquire images of the vascular structures [5]. Vessels can be segmented using these medical images. Subsequently, segmented accurate vessel structures can be used to extract vessel characteristics such as shape, width, and branching. Accordingly, extracted anatomical vessel features may facilitate safe diagnosis and thereby efficient treatment of various diseases [6]. But, automatic segmentation of vessel structures in a precise manner is challenging due to various reasons such as radiofrequency field (RF) inhomogeneities, imaging noise, and high variability and complexity of vessel structures [7].

Sekiguchi et al suggested a vessel segmentation approach where multiple branches are segmented successively where local region growing schema is used for each branch [8]. Since success of region growing is limited, region growing is further improved by utilizing an atlas in [9]. In [5], active contour method is utilized as a deformable curve where image gradients are fitted to vessel structures as an external force in addition to common internal forces. Strzelecki et al investigated level set segmentation methods using simulated vessels with varying noise levels where they observed that results are satisfactory even for images with high noise levels [10]. Gazit et al employed level set formulation for surface evolution driven by a novel variational cost function with a minimal variance term in addition to edge-based term and geodesic active contour (GAC) regularization term [11]. Gooya et al modified the CURVES regularization such that level set method is employed for geometric regularization via GAC [12]. In [13], Hao and Shen used a watershed method to initialize the level set method for segmentation of vessel boundaries. Manniesing et al proposed a vessel enhancement scheme using anisotropic diffusion (AD) which applies more diffusion among the vessel direction inside the vessel like structures and applies less diffusion in other structures by means of a specially designed tensor [14]. Akram et al used a 2-D Gabor wavelet in conjunction with a sharpening filter to enhance the vessels in retinopathy images [15]. In [16], Staal et al employed ridge detection driven by characteristics of Hessian matrix to extract vessel centerlines in retinal images. In their approach, they define centerlines as piecewise lines and partitioned the pixels based on their distance to these lines. Ricci and Perfetti used a line detection approach where the filters for 12 different orientations are applied to find vessel-like structures [17]. In [1], Esneault and Lafon proposed a graph-cuts based approach for the segmentation of vessels in the liver. Researchers also applied model matching strategies for the segmentation of vascular structures. For example, superellipsoid models are used for vessel detection in [18]. Likewise, Rohr and Wörz investigated 3 different intensity models for tubular structures extraction which can be applied to vessel segmentation [19]. Li and Yezzi claimed that while methods utilizing tubular structures can produce paths aligned on vessel centerlines, global minimal-path methods cannot [20]. Comprehensive reviews of vessel segmentation methods are given by Kirbas and Quek [21], Lesage et al [7], Radaelli et al [5], and Moccia et al [22]. Limited to retinal images, [23] reviews and compares the retinal vessel segmentation methods.

Vessel segmentation literature also contains some fluid flow-based segmentation approaches. In [24], Chang and Valentino adapted charged-fluid method for vessel segmentation. They modeled repelling electric forces as a charged fluid that can flow through and flow around obstacles. Thus, Chang and Valentino claim that this model provides a motion like fluid flow. In [25], Yan and Kassim proposed a schema to minimize surface energy function that acts as a capillary active contour method. They used level set method for marching the boundary where variational calculus is used to define energy of wetting surfaces as a variational functional. In [26], Liu and Nixon suggested use of pressure, capillary actions, and surface tension are as flow attributes. They claim that these physical forces define edgewise and region-wise driving forces that also provides accurate and longer capture range. Finally, they form the fluid flow velocity by combining resistance and total driving forces.

The approach we propose is different from the above 3 approaches because we conceptualized the flowing fluid as shallow water acting inside the captivity of a 3D closed surface such as vessel. So far, shallow water equations have been utilized for modeling fluid flow through wide open environments, i.e. ocean [27]. In this study, we used fluid equations modeling shallow waters to define a deformable model with effective convergence criteria. The method we propose captures both thin and thick vessel structures in medical images with various noise levels and types [28][1].

For the remaining part of the paper, fluid equations and their brief description is given in section 2, algorithmic details of the proposed method is given in section 3, implementation issues are discussed in section 4, parameters are analyzed in section 5, and experiments are presented in section 6, and finally concluding in section 7.

2 Fluid Equations

2.1 Navier-Stokes Equations (NSE)

Irrespective from the strength of applied shear stress, fluids (also liquids and gases) are continually deforming materials. Navier-Stokes Equations (NSE) models the fluid flow which are partial differential equations (PDE). Nonlinearity of NSE makes it necessary to utilize state-of-the-art numerical solvers where analytical solution exists only for simple or special cases. There are different forms of NSE while the widely used one is incompressible fluid via Newtonian formulation which is defined as

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{u}) = 0 \tag{1}$$

$$\frac{\partial \vec{u}}{\partial t} = -(\vec{u} \cdot \nabla)\vec{u} + \nabla \cdot (\nu \nabla \vec{u}) + (-\frac{1}{\rho}\nabla p) + \vec{f} \tag{2}$$

$$\frac{\partial \rho}{\partial t} = -(\vec{u} \cdot \nabla)\rho + \kappa \nabla^2 \vec{u} + f_e \tag{3}$$

convection (advect) — diffusion (diffuse) — pressure — force sources

where

$\vec{u}\ (x, y, z)$ is 3D velocity vector field

ν : Kinematic viscosity, where $\nu = \mu / \rho$

μ : Viscosity

p : Pressure

$\vec{f}$: Body forces such as gravity

ρ : Fluid density

κ : Heat conduction coefficient

f_e : External force per unit mass

Equation (1) models the continuity of the fluid due to incompressibility assumption and preservation of the mass. Equation (2) models the velocity changes caused by internal and external forces with conservation of momentum assumption. Equation (3) models density change with conservation of energy assumption. Internal chemical reactions, electromagnetic fields or similar external sources may lead to heat flows that can cause density changes. Generally, Equation (3) is ignored.

[1] This paper is based on the PhD dissertation: Vessel segmentation using shallow water equations.

2.2 Shallow Water Equations (SWE)

The SWE models the flow of water and other incompressible fluids using a simplified water representation. Essentially, the fluid depth is assumed to be small compared to the wavelength of the fluid disturbance. With this assumption, fluid is represented as simple 2D height field instead of complex 3D description (Fig 1).

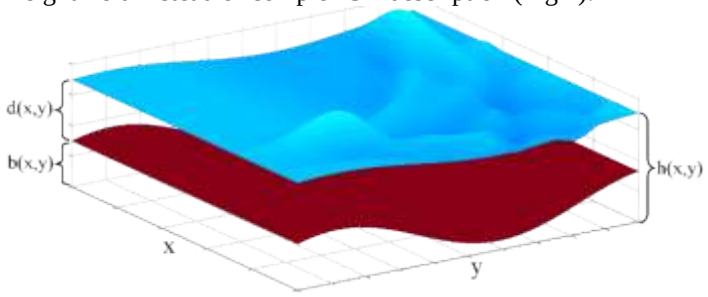


Fig 1. 2 dimensional (2D) height field representation: $h(x, y) = b(x, y) + d(x, y)$ is surface height where $b(x, y)$ is ground height and $d(x, y)$ is water height

Vertical velocity is small and can be ignored by the shallowness assumption (see Fig 2 for 1D shallow water representation). For an arbitrary (x, y) point in shallow water, this assumption holds since velocity has small variation in z-direction and pressure gradient is almost constant between the bottom and surface. Thus, a solution can be obtained with horizontal flow that is independent of height of water column. First, vertical velocity is taken as zero and no variations in z-direction is assumed in NSE. Then, NSE are depth integrated to obtain SWE. Further simplification is applied to obtain simpler Euler equations by assuming zero viscosity (inviscid) with constant density, i.e. for water. Pressure is proportional to the water height and gravity, and so the pressure term is modified accordingly [27,29].

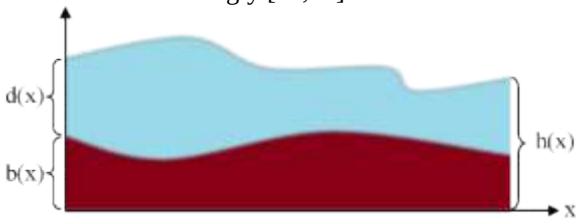


Fig 2. 1 dimensional (1D) height field representation: $h(x) = b(x) + d(x)$ is surface height where $b(x)$ is ground height and $d(x)$ is water height

Based on the above assumptions, 1D shallow water equations are defined in Equation (4) and Equation (5).

$$\frac{\partial d}{\partial t} + \frac{\partial}{\partial x}(ud) = 0 \tag{4}$$

$$\frac{\partial u}{\partial t} = \underbrace{-\left(u\frac{\partial u}{\partial x}\right)}_{\text{convection (advect)}} + \underbrace{0}_{\text{no diffusion (inviscid)}} + \underbrace{\left(-g\frac{\partial h}{\partial x}\right)}_{\text{pressure}} + \underbrace{0}_{\text{no force sources}} \tag{5}$$

2.3 Linearized Shallow Water Equations (LSWE)

After applying shallow water assumptions, the remaining non-linearity only pertains to Equation (5) due to the convection term. Note that, convection term is responsible for advection of fluid properties, i.e. velocity. Kass and Miller ignored the convection term and linearized the equation around *h* (a constant) to further simplify Equation (5) [27]. This simplification is only valid when fluid velocity is small, and the ground is smooth (slowly varying) [27].

$$\frac{\partial u}{\partial t} = \underbrace{0}_{\text{no convection}} + \underbrace{0}_{\text{no diffusion}} + \underbrace{(-g\frac{\partial h}{\partial x})}_{\text{pressure}} + \underbrace{0}_{\text{no force sources}} \tag{6}$$

By combining Equation (4) and Equation (6), equation (7) is obtained:

$$\frac{\partial^2 h}{\partial t^2} = gd\frac{\partial^2 h}{\partial x^2} \tag{7}$$

which is the 1D wave equation. Kass and Miller used finite-difference technique to discretize the Equation (7) [27]. Then, they preferred a first-order implicit integrator which leads to a linear system. Here, A is a tridiagonal matrix that is given as

$$Ah^{(n)} = 2h^{(n-1)} - h^{(n-2)} = y \tag{8}$$

where *n* is the iteration index. Matrix A is formed by *e* and *f* where they are given as

$$A = \begin{bmatrix} e_0 & f_0 & & & & \\ f_0 & e_1 & f_1 & & 0 & \\ & f_1 & \ddots & \ddots & & \\ & & \ddots & e_{w-3} & f_{w-3} & \\ & 0 & & f_{w-3} & e_{w-2} & f_{w-2} \\ & & & & f_{w-2} & e_{w-1} \end{bmatrix}$$

$$e_i = 1 + k\begin{cases} d_0 + d_1 & i = 0 \\ d_{i-1} + 2d_i + d_{i+1} & 0 < i < w-1 \\ d_{w-2} + d_{w-1} & i = w-1 \end{cases}$$

$$f_i = -k(d_i + d_{i+1})$$

$$k = g(\Delta t)^2 / (2(\Delta x)^2)$$

Here *w* is the number of water columns, d_i is water depth in i[th] column, *g* is gravity, Δt is simulation time step in seconds, Δx is water column width in meter, and variable *k* is a combination of *g*, Δt, and Δx parameters. Right hand side (RHS) of Equation (8) can be regarded as extrapolation of the fluid surface in last simulation iteration. Based on this assumption, Kass and Miller formed Equation (9) by adding damping in the RHS of Equation (8) to introduce artificial viscosity effect.

$$Ah^{(n)} = h^{(n-1)} + (1-\tau)(h^{(n-1)} - h^{(n-2)}) = y \tag{9}$$

where $0 \le \tau \le 1$ is the coefficient for damping [27]. Linear systems in Equation (8) or Equation (9) are solvable in linear, *O(n)*, time using tridiagonal matrix algorithm (TDMA) [27]. 2D/3D LSWE can be defined using Equation (10) and Equation (11).

$$\frac{\partial^2 h}{\partial t^2} = gd\left(\frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2}\right) = gd\nabla^2 h \tag{10}$$

$$\frac{\partial^2 h}{\partial t^2} = gd\left(\frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} + \frac{\partial^2 h}{\partial z^2}\right) = gd\nabla^2 h \tag{11}$$

Alternating Direction Implicit (ADI) method [30] is used to solve Equation (10) and (11). ADI splits the RHS of equation (10) into 2 terms for 2D and into 3 terms for 3D to make RHS independent from each other. So, linear systems remain tridiagonal and time complexity in each iteration remains proportional to volume size, i.e. the number of voxels in the volume.

1D: $O(w) = O(n)$

3D: $hwO(W) + wcO(H) + whO(C) = O(3whc) = O(n)$

where w is width of image, h is height of image, and c is slice count of image.

LSWE ignores fluid motions such as eddies, whirl, and splashing which are not necessary for our purposes. In Fig 3, height of the fluid surface is h*(x)* is used as *z*-direction (depth) of the fluid for 3D visualization of fluid on 2D height map.

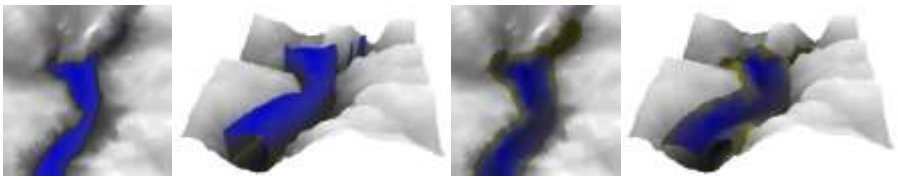


Fig 3. Fluid evolution example. a) Initial fluid placement, b) 3D visualization of *a*, c) Fluid surface at 500th iteration, and d) 3D visualization of *c*.

3 Proposed Method

Flowchart of the proposed LSWE based segmentation algorithm is shown in Fig 4.

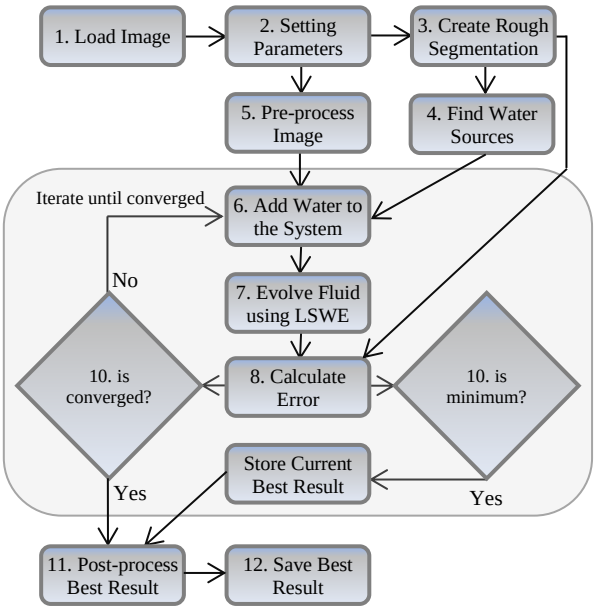


Fig 4. Flowchart of the proposed LSWE based vessel segmentation method.

Steps of the algorithm in

Fig **4** are explained as following:

Step 1. Loading image
In MRA images, vessels appear bright with high intensity values. So, voxel intensity values must be inverted to turn vessels into dark intensities. Once vessels become dark in the whole volume, fluid flow simulation can be used to fill these valleys.

Step 2. Setting parameters
Fluid simulation using LSWE assumes voxel to be in meters. Thus, medical images that are acquired in real world coordinates (i.e. millimeter resolution) are converted into meter resolution. Voxel intensity values act as ground depth. Thus, they are also normalized to satisfy shallowness assumption.

Step 3. Creating rough segmentation
Vessels are estimated using rough segmentation that is used for both water source determination and check for convergence. Otsu's thresholding method [31] is fast but not accurate. So, we developed a multiresolution local filtering approach for robust rough segmentation. Proposed filtering approach uses sliding window where voxel intensity at the window center is subtracted from the clamped maximum intensity in the window. We inspired from the study in [32] by Alanso and Vilari for our filtering approach although ours' is quite different. In our approach, maximum intensity value within a window is clamped to an average intensity value within a larger window.

$$R_{x,y,z} = \begin{cases} \hat{R}_{x,y,z} & if\ \hat{R}_{x,y,z} > T_{rs} \\ 0 & if\ \hat{R}_{x,y,z} \le T_{rs} \end{cases} \tag{12}$$

where $R_{x,y,z}$ represents the rough segmentation, T_{rs} represents the noise threshold, $\hat{I}_{x,y,z}$ represents the low-pass filter applied on pixel $I_{x,y,z}$, and r_v represents the estimation of maximum vessel radius for the given image,

$$\hat{R}_{x,y,z} = \max(\min(I^{(m)}_{x,y,z}, I^{(c)}_{x,y,z}) - \hat{I}_{x,y,z}, 0),$$

$$I^{(m)}_{x,y,z} = \max I_{x+\Delta x, y+\Delta y, z+\Delta z},\ -r_v \le \Delta x, \Delta y, \Delta z \le r_v \ : \text{maximum},$$

$$I^{(c)}_{x,y,z} = \frac{1}{(8r_v+1)^3} \sum_{\Delta x=-4r_v}^{4r_v} \sum_{\Delta y=-4r_v}^{4r_v} \sum_{\Delta z=-4r_v}^{4r_v} I_{x+\Delta x, y+\Delta y, z+\Delta z} \ : \text{clamped},$$

Rough segmentation is a rough view of vessels while LWSE segmentation result is the detailed view. As fluid flows, fluid heights become more similar to the rough segmentation. LSWE segmentation is obtained when they become most similar.

Step 4. Finding the water sources
First, $\tilde{R}$, a normalized rough segmentation is obtained from the rough segmentation, R, by using normalization via R / *max(R)*. Then using r_w as kernel radius and $2r_w+1$ as the width and height of a square sliding window, another image is generated to suppress the thin structures:

$$\tilde{R}^{(thinSupressed)}_{x,y,z} = [\prod_{\Delta x=-r_w}^{r_w} \prod_{\Delta y=-r_w}^{r_w} \prod_{\Delta z=-r_w}^{r_w} (\tilde{R}_{x+\Delta x, y+\Delta y, z+\Delta z} + 1)] - 1$$

Finally, thick vessels can be set as water sources using a fast and simple filtering mechanism, where T_{ws} is the threshold value such that values lower than T_{ws} in the suppressed image will be set to zero. Results of this processing are reflected in Fig 5.

$$W_{x,y,z}^{(source)} = \begin{cases} 0 & if\ \ \tilde{R}_{x,y,z}^{(thinSupressed)} \le T_{ws} \\ \tilde{R}_{x,y,z}^{(thinSupressed)} & otherwise \end{cases} \tag{13}$$

(a) MIP of original image *I* (b) MIP of rough segmentation (c) MIP of water source

Fig 5. Rough segmentation and water source detection in an MRA image.

Step 5. Preprocessing the image

One of the basic assumptions of LSWE is slow velocity in a moving fluid front on the smoothly varying ground. In each image, smoothly varying ground is equivalent to smoothly changing background, hence sharp intensity changes must be eliminated at this step. In our case, we are not doing any pre-processing for the MRA images.

Step 6. Pumping water from water sources

Below Equation (14) is used the add water into the simulation at each time step.

$$d_{x,y,z} = (d_{x,y,z} + W_{x,y,z}^{(source)})/2 \tag{14}$$

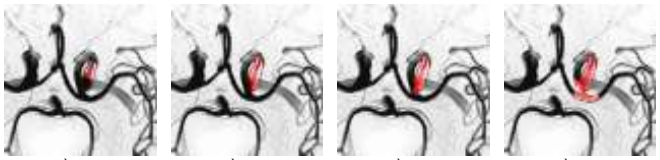
where $d_{x,y,z}$ represents the depth of water at voxel located at point *(x,y,z)*. Thus, slow velocity assumption is satisfied by pumping right amount of water into the vessel system. Water source is adaptively changed using Equation (15) to increase speed:

$$W_{x,y,z}^{(source)} = \max(W_{x,y,z}^{(source)}, d_{x,y,z})/2, \quad for\ \ d_{x,y,z} > W_{AVD} \tag{15}$$

where W_{AVD} is average vessel depth calculated from nonzero values.

Step 7. Fluid evolution using LSWE

The ADI method which is introduced in section 2 is used for integration of LSWE. Finite difference assumes that Δt goes to zero. Thus, time step (Δt) must be small enough where 0.001 and 0.01 are found sufficiently small and yield desired numerical accuracy with reasonable simulation execution times in our experiments.



a) 10th iteration b) 20th iteration c) 30th iteration d) 40th iteration

Fig 6. Fluid evolution on a slice of an MRA image.

As suggest by Kass and Miller to prevent volume loss, we set negative volumes as zero. Then we normalize the whole volume by recalculating the surface heights [27].

Minuscule amount of water pieces may form especially when low level of artificial viscosity is used. We developed a vaporization scheme to avoid this problem. Here, vaporization depth (V_d) is reduced by a vaporization factor (V_r) when water depth (d_i) is smaller than a constant value:

$$d_i = d_i - \begin{cases} d_i V_r & \text{if } d_i \le V_d \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

Note that, vaporization reduces to a simple thresholding when V_r is 1.

Step 8. Calculating the error

Water height is calculated as $H^{(n)}$ in step 7 for each iteration in time. Blended water height at nth iteration, $H_{(b)}^{(n)}$, is calculated using Equation (17).

$$H_{(b)}^{(n)} = (H_{(b)}^{(n-1)} + H^{(n)})/2 \tag{17}$$

where $H_{(b)}^{(0)} = 0$. Water fills vessels (dark intensities) almost simultaneously. As water fills deep (thick) vessels shallow (thin) vessels are also filled and similarity with rough segmentation increases. If water floods to non-vessel regions, then similarity starts to decrease. Error, inverse of similarity, is defined as mean absolute difference between $H_{(b)}^{(n)}$ (blended water heights) and R (rough segmentation).

$$E^{(n)} = \frac{1}{nE^{(0)}} \sum \left| H_{(b)\,x,y,z}^{(n)} - R_{x,y,z} \right| \tag{18}$$

where $E^{(0)}$ is maximum error, and error $E^{(n)}$ is divided by $E^{(0)}$ to normalize error to 1. As fluid flows, similarity with rough segmentation increases and error decreases. So best segmentation is obtained when minimum error is reached (Fig 7).

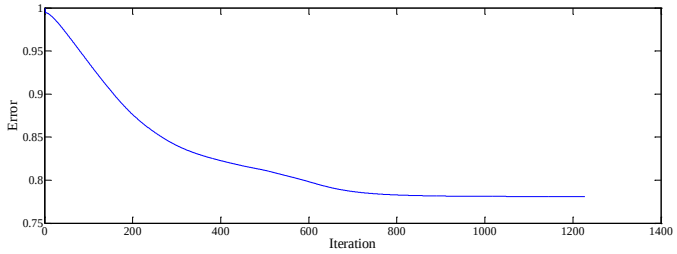


Fig 7. Error graph for an MRA image.

Step 9. Detecting possible minimums

When a minimum point in error is detected, current water heights must be stored as the current best. Yet, minimum can only be detected when error goes in an upward direction, which implies that actual minimum point has already been passed. At each detected minimum, water height needs to be copied and this may lead to large amount of memory copy operations if the error graph is oscillatory. This issue is addressed by doing copy with tolerance value check as shown in Equation (19).

$$(E^{(n)} - E_{\min}^{(m)})/(E^{(0)} - E_{\min}^{(m)}) > M_{tolerance} \tag{19}$$

where $M_{tolerance}$ is 0.01, $E^{(n)}$ is the error at n^{th} iteration, $E^{(0)}$ is the maximum error, and $E_{min}^{(m)}$ is the obtained minimum error so far for the m^{th} iteration.

Step 10. Testing for convergence

There can exist more than single local minimum in the error surface due to various reasons. Achieving the global optimum is not assured if algorithm terminates when it reaches the first minimum, i.e. a local minimum. To avoid being trapped in a local minimum, convergence is decided using a tolerance value as given in Equation (20).

$$(E^{(n)} - E_{\min}^{(m)}) / (E^{(0)} - E_{\min}^{(m)}) > C_{tolerance} \tag{20}$$

where $C_{tolerance}$ is 0.25 ($0 < C_{tolerance} \leq 1$), $E^{(n)}$ is the error at n^{th} iteration, $E^{(0)}$ is the maximum error, and $E_{min}^{(m)}$ is the minimum error found so far at the m^{th} iteration. As $C_{tolerance}$ increases number of iteration and chance of passing local minimums also increases. Error can reach a fixed value (see Fig 7), that also need to be tested.

Step 11. Post processing

Segmentation result can contain noise that is formed due to tiny fluid columns. Such situation can especially happen if input image is noisy and LSWE is executed with zero or very low viscosity parameter and vaporization schema is turned off.

As seen in this section, our algorithm depends on several parameters. In section 5, we will elaborate on how to determine these parameters in practice.

4 Implementation

The serial nature of TDMA would have prevented the parallelization. We used single-step Jacobi approximation on the linear system in Equation (8) or Equation (9) to enable parallel computation while solving the linear system. Thus, obtained numerical linear system solver for Equation (8) or Equation (9) becomes parallel in nature which allows utilization of OpenMP or CUDA to have an efficient implementation.

Jacobi method can be used for solving a linear system A$\boldsymbol{h}$=y if A matrix is strictly (or irreducibly) diagonally dominant. Matrix A defined in Equation (8) or Equation (9) is symmetric, positive definite, and diagonally dominant. Thus, we used Jacobi approximation, to define A matrix as a sum of D (diagonal entries) and R (remainder).

$$\begin{aligned} &Ah = (D+R)h = y \\ &\rightarrow\ h^{(n+1)} = D^{-1}(y - Rh^{(n)}) \end{aligned} \tag{21}$$

$$h^{(n+1)} = \begin{bmatrix} 1/e_0 & & & \\ & 1/e_1 & & \\ & & \ddots & \\ & & & 1/e_{w-1} \end{bmatrix} \left(\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{w-2} \\ y_{w-1} \end{bmatrix} - \begin{bmatrix} 0 & f_0 & & & \\ f_0 & 0 & f_1 & & \\ & f_1 & 0 & \ddots & \\ & & \ddots & \ddots & f_{w-1} \\ & & & f_{w-1} & 0 \end{bmatrix} \begin{bmatrix} h_0^{(n)} \\ h_1^{(n)} \\ \vdots \\ h_{w-2}^{(n)} \\ h_{w-1}^{(n)} \end{bmatrix} \right)$$

Then final 1D water height update equation becomes

$$h_i^{(n+1)} = \frac{y_i - f_{i-1} h_{i-1}^{(n)} - f_i h_{i+1}^{(n)}}{e_i} \tag{22}$$

This approximation is valid for LSWE integration since simulation time step is small in our settings. Thus, fluid heights change slowly which allows Jacobi iteration to converge mostly in one step.

5 Parameter Analysis

5.1 Vessel Shapes

Nine 3D synthetic vessel structures are created (Fig 8) to analyze the effect of vessel shapes on segmentation result.

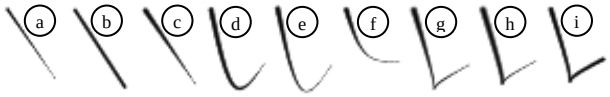


Fig 8. 3D synthetic vessel structures.

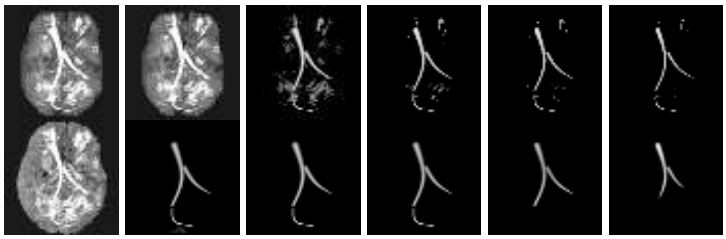
Table 1. Jaccard indices for different noise levels

	Original	PSNR=40	PSNR=30	PSNR=25	Avg	Min	Max
Vein a	100.0	98.9	99.3	93.4	97.9	93.4	100.0
Vein b	100.0	100.0	100.0	91.0	97.8	91.0	100.0
Vein c	100.0	99.7	99.7	90.1	97.4	90.1	100.0
Vein d	100.0	99.8	99.8	79.4	94.8	79.4	100.0
Vein e	99.9	99.7	99.2	74.8	93.4	74.8	99.9
Vein f	100.0	99.5	99.6	87.9	96.7	87.9	100.0
Vein g	91.1	90.6	83.6	55.1	80.1	55.1	91.1
Vein h	79.5	78.4	72.7	49.4	70.0	49.4	79.5
Vein i	100.0	100.0	100.0	85.7	96.4	85.7	100.0
Average	96.7	96.3	94.9	78.5			

Experiments expose the importance of vessel radius, noise level, and vessel curvature. Vessels in real-world data have low curvature and have slowly changing radius which are within well performing regime of LSWE. Also, LSWE is robust to moderate noise levels (PSNR: 30dB) that may exists in real world data.

5.2 Rough Segmentation

Both water source and convergence are determined using rough segmentation. Thus, parameter selection for rough segmentation is decisive to accurate vessel extraction. Estimated vessel radius parameter (V_r) and noise thresholding parameter ($0 \leq T_{rs} \leq 1$) determines the quality and noise level of vessel structures in the rough segmentation step. As shown in Fig 9, threshold value of 0.2 ± 0.1 leads to correct segmentation for the synthetic vessels. This threshold range covers the one-fifth of the whole threshold range which indicates that T_{rs} is not a fragile parameter.



(a) $T_{rs}=0$ (b) $T_{rs}=0.05$ (c) $T_{rs}=0.20$ (d) $T_{rs}=0.35$ (e) $T_{rs}=0.40$ (f) $T_{rs}=0.45$

Fig 9. Rough segmentation and its effect on final segmentation.

5.3 Water Shallowness

SWE assumes shallow water. However, water depth becomes negligible compared to the gravity for very shallow water. This indicates that Δh (maximum ground depth) must be chosen appropriately. Segmentation performance is shown in In Fig 10 for Δh values between 10^{-3} and 10^{3}. In this experiment, for each Δh value, average of the Jaccard indices of the first 5 patients in the BrainWeb database are taken. As seen in figure, Δh equals to 0.1 meter yields best segmentation result.

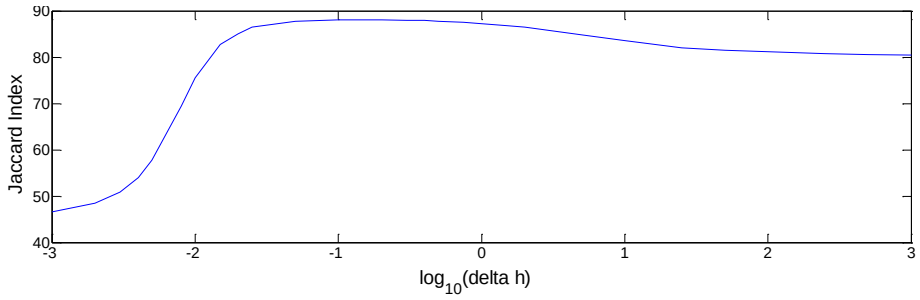


Fig 10. Segmentation performance with respect to Δh.

5.4 Simulation Time Step

Kass and Miller used a stable implicit numerical schema [27]. During the fluid flow simulation, small time step leads to longer simulation times while large time step leads to lower numerical accuracy. Segmentation performance is shown in Fig 11 for Δt between 0.01 and 10. For each Δt value, average of the Jaccard indices of the first 5 patients in the BrainWeb database are taken. As seen in figure, Δt equals to 1 second yields best segmentation result.

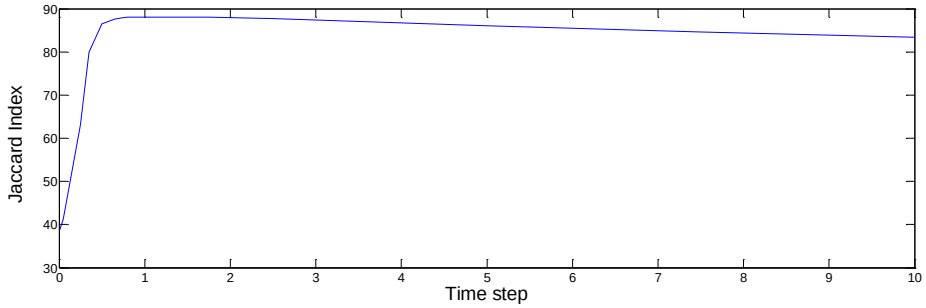


Fig 11. Segmentation performance with respect to Δt.

5.5 Viscosity

SWE hence LSWE both assume inviscid fluid therefore the mathematical model has no viscosity term. However, as explained before, Kass and Miller introduced artificial viscosity effect in Equation (9) by adding damping in the RHS of Equation (8) [27].

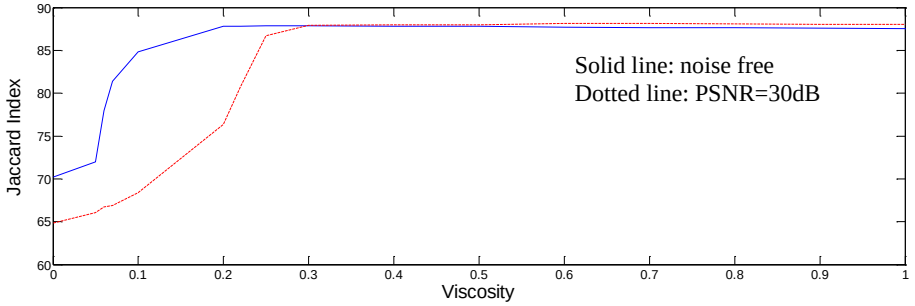


Fig 12. Segmentation performance with respect to viscosity.

In Fig 12, segmentation performance for the first 5 patients in BrainWeb database is shown for viscosity (τ) between 0 and 1. In here, solid line shows the segmentation performance for noise-free images where $\tau=0.3$ yields the best vessel segmentation result. Dotted line shows the vessel segmentation performance for the synthetically added noise (PSNR equals to 30dB) where best segmentation result is obtained for $\tau=0.6$. Fig 12 shows that higher viscosity is required to obtain smooth fluid boundary for noisy images. Thus, viscosity parameter apply regularization on fluid boundary. However, there are 2 disadvantages of setting higher viscosity values; fluid will evolve slower hence convergence will take more time, and thin structures may be under segmented since fluid boundary has to be smoother.

5.6 Vaporization

Segmentation performance with respect to vaporization depth and vaporization rate is shown in Fig 13. Vaporization depth of 0.015 and vaporization rate of 0.2 yields best vessel segmentation result. For the very noisy images, vaporization must be combined with viscosity since it may be insufficient itself as a regularization schema.

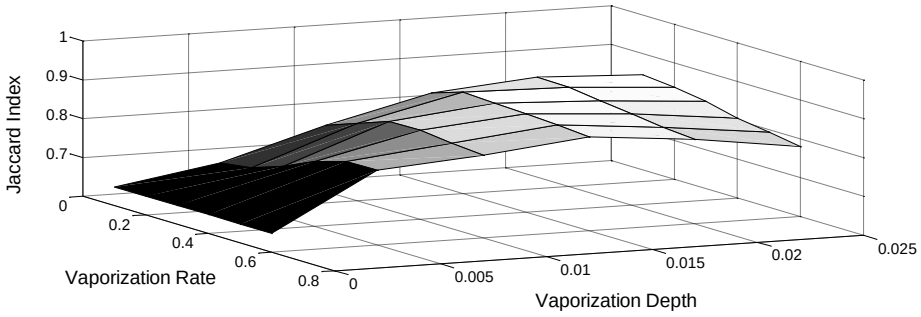


Fig 13. Segmentation performance for vaporization schema.

5.7 RF Inhomogeneity

RF inhomogeneities are inherent in MR imaging where image is corrupted with a smooth brightness field. RF inhomogeneity degrades the performance of the various segmentation algorithms, i.e. Chan-Vese level set segmentation [33].

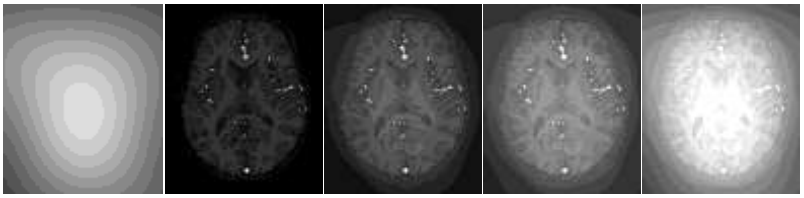


Fig 14. a) Synthetic RF field, b) example data set from BrainWeb database, c) 25% RF field is added, d) 50% RF field is added, e) 100% RF field is added.

In Fig 14, 4 different levels of RF inhomogeneity imposed on MRA image is shown. As seen in Fig 15, segmentation performance of the LSWE changes only slightly even for high RF inhomogeneity levels which shows that LSWE does not require a separate bias field correction mechanism.

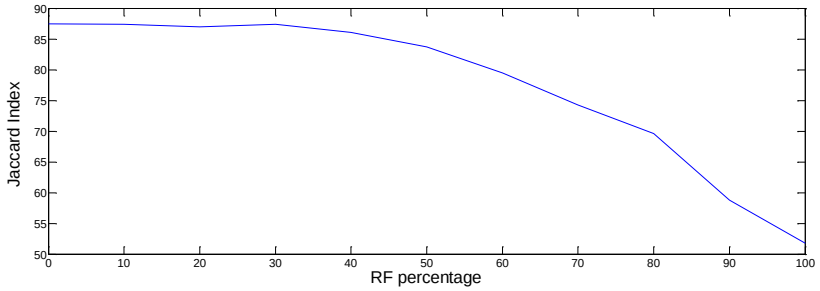


Fig 15. LSWE segmentation performance for low to high RF inhomogeneity levels.

6 Experiments

6.1 The BrainWeb Database

BrainWeb database contains 20 simulated brain MR images with a wide range of vessel structures. For each simulated image, its segmentation, and probabilities of 12 different tissues are given. We reconstructed realistic MRI volumes using these 12 class probabilities (Fig 16) for our comparison studies.

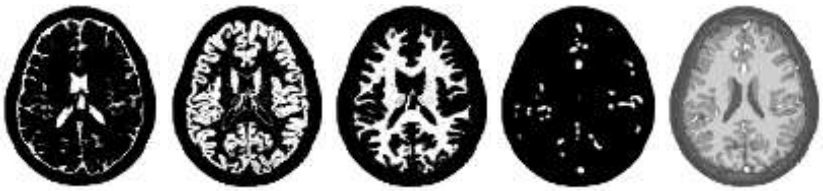


Fig 16. a) P(CSF), b) P(GM), c) P(WM), d) P(Vessel), e) reconstructed MRI

We used vessel class in BrainWeb database as our ground truth (Fig 16.d). Then, we compared our results with Chan-Vese level set results set using Jaccard index performance measure. In comparisons, different noise levels and RF inhomogeneity levels are used. Table 2 shows the superiority of the LSWE method compared to Chan-Vese method. LSWE tends to shrink the boundary slightly whereas Chan-Vese

method tends to enlarge and smooth the boundary. LSWE segments thin structures better and it is more robust to noise compared to Chan-Vese level set method.

Table 2. LSWE versus Chan-Vese level set comparison using Jaccard index

	Original image		PSNR=30dB		PSNR=25dB	
Patient	LSWE	Chan-Vese	LSWE	Chan-Vese	LSWE	Chan-Vese
04	86.4	82.4	82.8	74.8	73.1	65.6
05	88.6	83.5	84.1	75.9	75.2	66.9
06	88.4	83.9	84.5	76.6	76.1	67.8
18	87.8	83.8	84.7	76.6	76.0	67.9
20	87.9	83.5	84.0	76.0	75.1	67.4
38	86.8	82.3	80.7	75.0	72.4	66.0
41	86.8	82.9	82.1	75.3	74.0	66.2
42	88.4	83.3	84.4	76.0	75.1	67.2
43	84.8	81.2	80.2	73.6	71.4	64.4
44	88.2	83.2	79.7	75.8	73.2	67.3
45	87.3	83.1	82.5	75.5	73.7	66.5
46	89.1	83.8	85.3	76.7	77.1	68.4
47	86.2	82.8	80.7	75.7	73.3	66.5
48	85.0	81.7	74.9	73.9	68.6	65.0
49	85.0	81.2	79.9	73.2	71.3	64.3
50	83.9	82.4	79.1	74.8	70.7	65.4
51	89.0	83.3	84.4	76.1	76.3	67.5
52	88.3	84.1	79.8	76.6	64.1	67.4
53	85.1	80.8	82.1	74.1	73.6	65.6
54	87.9	82.6	83.9	75.0	74.5	65.9
Average	**87.0**	**82.8**	**82.0**	**75.4**	**73.2**	**66.5**

6.2 MRA Images

Clinical dataset contains 11 MRA volumes with size of 224x256x220 voxels in 0.7 mm rescaled isotropic resolution that are acquired using 1.5T Siemens scanner. Image acquisition parameters are: 0.35 mm x 0.35 mm x 0.70 mm voxel resolution, 12 bits, TR = 23 ms, TE = 7 ms, and FA = 25 degrees. Parameters of LSWE method and Chan-Vese method are optimized via trial and error for best performance (Table 3).

Table 3. Segmentation parameters for LSWE method

Parameter	Value	Unit
Δt	1	second
$\Delta x, \Delta y, \Delta z$	1	meter
Δh	0.5	meter
τ	0.1	
V_d	0.01	%
V_r	0.9	%

Parameter	Value	Unit
r_v	4	voxel
T_{rs}	0.15	%
r_w	2	voxel
T_{ws}	0.025	%
T_p	0.001	%
$C_{tolerance}$	0.25	%

A blindfolded validation study is carried out with the help of a radiologist. First, performance of LSWE and Chan-Vese methods for thick and thin vessels, number of false alarms, and overall quality are evaluated by the radiologist using 10 pt Likert scale. During expert evaluation, sagittal, coronal, and transverse views as well as 3D MIP rendering view are used in arbitrary angles and brightness levels for the original and segmented image slices. To guarantee the consistency of the blind observer, each segmentation result is displayed in a randomized order and twice (see Fig 17).

a) MIP view of MRA b) LSWE result c) Chan-Vese result

Fig 17. LSWE and Chan-Vese segmentation results for an MRA image.

Consistency of 2 observations for each vessel segmentation method is validated using paired t-test where no statistically significant difference is found ($p = 1.0$ for LSWE method and $p = 0.1$ for Chan-Vese method). Thus, 2 observations are combined to obtain a single observation set. Statistical analyses show that differences between LSWE method and Chan-Vese method are statistically significant, i.e. $p < 10^{-4}$ for thin structures, $p < 0.0015$ for thick structures, and $p < 10^{-4}$ for overall quality. As seen in Table 4, LSWE method performs better compared to Chan-Vese method.

Table 4. Gathered observer study statistics for 11 MRA volumes

		Thin structures	Thick structures	False alarms with a low rate	Overall quality	Weighted Average
LSWE	Average	7.7	8.8	9.9	8.6	8.7
	Min	6.0	7.0	8.0	7.0	7.9
	Max	9.0	9.0	10.0	9.0	9.1
	Std.Dev.	0.7	0.5	0.5	0.6	0.3
Chan-Vese	Average	5.9	7.7	9.7	6.8	7.5
	Min	4.0	5.0	5.0	5.0	5.5
	Max	7.0	9.0	10.0	8.0	8.3
	Std.Dev.	0.8	1.3	1.1	0.9	0.8

7 Conclusion

In this study, linearized SWE is parallelized to obtain a fast fluid-flow simulation which is used as a backbone of a novel fluid-flow driven vessel segmentation method. Rough vessel segmentation, detection of water source, and convergence criteria are proposed as novel building blocks where the facilitate fast and accurate segmentation via minimization of a cost function. This minimization schema, its ability to achieve global optimum, and method parameters are investigated in a comprehensive manner. Segmentation performance of the proposed LSWE method is shown on the BrainWeb database for different noise levels for 2 types of noise and proposed method found to be robust to noise. We also applied quantitative tests on BrainWeb and qualitative tests on clinical MRA images. Qualitative tests are carried out by a radiologist as a blind validation study. Both the quantitative tests and the qualitative tests showed that proposed method is superior to Chan-Vese level set method in all cases.

Acknowledgements

This paper is the outcome of my PhD dissertation. I would like to thank my PhD thesis advisor, Dr. Didem Gökçay, for her endless support and guidance in this long and fruitful academic progression of mine.

References

[1] S. Esneault, C. Lafon, and J.L. Dillenseger, Liver Vessels Segmentation Using a Hybrid Geometrical Moments/Graph Cuts Method, IEEE Transactions on Biomedical Engineering, 57(2), 276-283, (2010).

[2] World Health Organization. The top ten causes of death, (2008).

[3] B.T. Melanie, and X.Z. Andrew, Hepatocellular Carcinoma: The Need for Progress, Journal of Clinical Oncology. 23(13), 2892-2899, (2005).

[4] L. Antiga, B. Ene-Iordache, and A. Remuzzi, Computational geometry for patient-specific reconstruction and meshing of blood vessels from MR and CT angiography, IEEE Transactions on Medical Imaging, 22(5), 674-684, (2003).

[5] A. G. Radaelli, and J. Peiró, On the segmentation of vascular geometries from medical images, International Journal for Numerical Methods in Biomedical Engineering, 26(1), 3–34, (2010).

[6] A.C.S. Chung, Image segmentation methods for detecting blood vessels in angiography, Control, Automation, Robotics and Vision, ICARCV'06, 1-6, (2006).

[7] D. Lesage, E.D. Angelini, I. Bloch, and G. Funka-Lea, A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes, Medical Image Analysis, 13(6), 819-845, (2009).

[8] H. Sekiguchi, N. Sugimoto, S. Eiho, T. Hanakawa, and S. Urayama, Blood vessel segmentation for head MRA using branch-based region growing, System Computation Japan, 36(5), 80–88, (2005).

[9] N. Passat, C. Ronse, J. Baruthio, J. Armspach, C. Maillot, and C. Jahn, Region-growing segmentation of brain vessels: An atlas-based automatic approach, Journal of MRI, 1(6), 715-725, (2005).

[10] M. Strzelecki, P. Szczypinski, A. Materka, M. Kocinski, and A. Sankowski, Level-set segmentation of noisy 3D images of numerically simulated blood vessels and vascular trees, International Symposium on Image and Signal Processing and Analysis, 742-747, (2009).

[11] M. Holtzman-Gazit, R. Kimmel, N. Peled, and D. Goldsher, Segmentation of thin structures in volumetric medical images, IEEE Transaction on Image Processing, 15(2), 354-363, (2006).

[12] A. Gooya, K.H.L. Matsumiya, K. Masamune, Y. Masutani, and T. Dohi, A Variational Method for Geometric Regularization of Vascular Segmentation in Medical Images, IEEE Transaction on Image Processing, 17(8), 1295-1312, (2008).

[13] J. Hao, Y. Shen, and Y. Wang, Segmentation for MRA Image: An Improved Level Set Approach, IEEE Instrumentation and Measurement Technology Conference, 382-385, (2006).

[14] R. Manniesing, M.A. Viergever, and W.J. Niessen, Vessel enhancing diffusion: A scale space representation of vessel structures, Medical Image Analysis, 10(6), 815-825, (2006).

[15] M.U. Akram, A. Atzaz, S.F. Aneeque, and S.A. Khan, Blood Vessel Enhancement and Segmentation Using Wavelet Transform, International Conference on Digital Image Processing, 34-38, (2009).
[16] J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, and van B. Ginneken, Ridge-based vessel segmentation in color images of the retina, IEEE Transaction on Medical Imaging, 23(4), 501-509, (2004).
[17] E. Ricci, and R. Perfetti, Retinal Blood Vessel Segmentation Using Line Operators and Support Vector Classification, IEEE Transaction on Medical Imaging, 26(10), 1357-1365, (2007).
[18] J.A. Tyrrell, E. di Tomaso, D. Fuja, K.R.T. Kozak, R.K. Jain, and B. Roysam, Robust 3-D Modeling of Vasculature Imagery Using Superellipsoids, IEEE Transaction on Medical Imaging, 26(2), 223-237, (2007).
[19] K. Rohr, S. Worz, High-precision localization and quantification of 3D tubular structures, IEEE International Symposium on Biomedical Imaging: Nano to Macro, 1160-1163, (2006).
[20] H. Li, and A. Yezzi, Vessels as 4-D Curves: Global Minimal 4-D Paths to Extract 3-D Tubular Surfaces and Centerlines, IEEE Transaction on Medical Imaging, 26(9), 1213-1223, (2007).
[21] C. Kirbas, and F.K.H. Quek, A review of vessel extraction techniques and algorithms, ACM Computing Surveys, 36(2), 81-121, (2004).
[22] Sara Moccia, Elena De Momi, Sara El Hadji, Leonardo S. Mattos, Blood vessel segmentation algorithms - Review of methods, datasets and evaluation metrics, Computer Methods and Programs in Biomedicine, 158, 71-91, (2018).
[23] Azhar Imran, Jianqiang Li, Yan Pei, Ji-Jiang Yang, Qing Wang, Comparative Analysis of Vessel Segmentation Techniques in Retinal Images, IEEE Access, 114862–114887, (2019).
[24] H. Chang, D.J. Valentino, G.R. Duckwiler, and A.W. Toga, Segmentation of Brain MR Images Using a Charged Fluid Model, IEEE Transaction on Biomedical Engineering, 54(10), 1798-1813, (2006).
[25] P. Yan, and A.A. Kassim, Segmentation of volumetric MRA images by using capillary active contour, Medical Image Analysis, 10(3), 317-329, (2006).
[26] X.U. Liu, and M.S. Nixon, Image, and volume segmentation by water flow, LNCS, 4842, 62-74, (2007).
[27] M. Kass, and G. Miller, Rapid, stable fluid dynamics for computer graphics, ACM SIGGRAPH Computer Graphics, 24(4), 49-57, (1990).
[28] F. Nar, Vessel segmentation using shallow water equations, PhD Thesis, Middle East Technical University, Ankara, Turkey, (2011).
[29] M. Shinbrot, The shallow water equations, Journal of Engineering Mathematics, 4(4), 293-304, (1970).
[30] J. Douglas, R.B. Kellog, and R.S. Varga, Alternating direction iteration methods for n space variables, Mathematics of Computation, 17(83), 279-279, (1963).
[31] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transaction on Systems, Man and Cybernetics, 9(1), 62-66, (1979).
[32] C. Alonso-Montes, D.L. Vilari, P. Dudek, and M.G. Penedo, Fast retinal vessel tree extraction: A pixel parallel approach, International Journal of Circuit Theory and Applications, 36(5-6), 641–651, (2008).
[33] T.F. Chan, and L.A. Vese, Active contours without edges, IEEE Transactions on Image Processing, 10(2), 266-277, (2001).

Internet of Things Based Intelligent Facial Expression Monitoring using EMG Signals

Ilhan Aydin [1] and Masood A. Othman [2]

[1] Firat University, Elazig, Turkey, iaydin@firat.edu.tr

[2] Firat University, Elazig, Turkey, masood.sindy@hotmail.com

Abstract:-Facial expressions are important aspects of human behaviors used to develop automated tools to assess anger, smile, fear, or surprise. These tools are effective to replace self-reporting methods, suitable for patients who are incapable to report their health status, especially patients in the intensive care unit (ICU) and minors**.** In the present study, we propose a wearable device with a biosensing mask to monitor the intensity of the human status of a patient using facial surface electromyography (sEMG). Wearable devices are the wireless sensor nodes, integrated into the Internet of Things (IoT) systems for remote monitoring of patients. At 1000 Hz, approximately nine sEMG channels are sampled for transmission to the cloud server and cover the entire frequency range through the gateway. We also consider low energy consumption wearing devices in the design process to carry out long-term monitoring. We develop a mobile web application to transmit large amounts of real-time sEMG data. We also perform digital signal processing, visualization, and interpretation to explain real-time sick situations such as angry, fear surprise, or smile data to caregivers. The cloud platform is a bridge between the web browser and sensor nodes. It helps to manage the wireless intercommunication between the web application and server. This study proposes a real-time scalable IoT system biopotential monitoring, a wearable solution for automatic sick situation assessment using classification and facial expressions to assess fear, surprise, anger, or smile.

Keywords: Biopotentials, SEMG, IoT, Facial Expression Recognition, Internet of Things

1. INTRODUCTION

The facial expression monitoring system (FEMS) is critical for optimizing the health process [1], and this facilitates the workflows and reduces medical costs. The FEMS is a pivotal tool for remote sensing, patients' diagnosis, and for relieving patients' discomfort. Previous studies have used many tools such as telemedicine, survey methods, and focus groups to collect data on patients' health and suitable medication [1-2]. Their outcomes reveal that angry, surprise, or smile monitoring systems and feedback systems are the pivotal tools to monitor daily habits. However, many constraints have been recorded to meet this self-reporting approach. For example, people face challenges to enter and monitor daily information manually. Moreover, this method is unsuitable for people with limited cognitive and expression skills, such as newborn babies, sedated patients, children with limited intelligence, and older adults' patients subjected to intensive care units (ICU). Lack of effective real monitoring system for patient continuous is the third shortcoming, leading to persistence and severe delay in patient's long-time intensive care.

Recent studies are explored to evaluate the facial expressions in the video [2] or physiological signal fusion [1], regarding pain intensity monitoring that involves surprise, fear angry, or smile. Nevertheless, limited studies are carried out on an integrated remote health monitoring system for pain assessment. Embedded surface sensors integrated into the wearable devices are the effective systems to monitor the activities of facial muscles and remotely monitored through the surveillance system.

To identify facial expressions, extracted features must be categorized into unique different classes. Different classification algorithms using machine learning (ML) can be used for the facial expression's classifications. Fright, annoyance, happiness, unhappiness, anxiety, and profound disapproval are remarked daily through specific facial expressions. However, studies that explore the context of facial expression in pain are scanty, essentially for the academic, psychological, medical research. Pain is a pivotal symptom experience by chronic-ill patients. However, measurement of pain in an unbiased manner is still challenging [2].

In clinical settings, common bioelectrical signals to measure mental and physical activities, such as heart, eyes, brain, and muscle are "Electrocardiography (ECG), Electrooculography (EOG), Electroencephalography (EEG), and Electromyography (EMG)" signals, respectively. These applications perform the pivotal roles in monitoring bio-signal activities that include analysis of muscle activation and diagnosis of heart disease.

This study aims to design an automatic system to distinguish and read facial expressions (i.e., happy, surprise, anger, sadness, fear, and disgust).

The rest of the paper is organized as follows:

The second section provides the related work where we discuss the remote monitoring sensing that reveals the IoT technology. The experimental method is discussed in the subsequent section. The acquisition, pattern training, and processing are also discussed. In the third section, we provide the results. We conclude the paper in the final section.

2. RELATED WORK

Numerous studies have discussed different functions of facial expression monitoring systems. The face is an object of major importance as it identifies people. Facial expressions are evaluated through pain in facial muscles because they communicate emotions [3]. Facial expressions are observed through conversation and interactions with other persons, neighborhood, and environment. Facial expressions play an active role in human feeling and enhancing verbal communication [4]. In Darwin's theory, all humans have similar facial expressions; the difference is only observed in cultures [5]. According to Ekman, emotions such as fear, sadness, surprise, anger, and happiness are types of facial expressions. Non-verbal communication is another example of facial expression as this displays nervousness [6]. Eyes are also significant features of facial expressions; eyes as organs of vision can be used to evaluate an individual nervousness. Eyes contact is also an important aspect of interpersonal communication. In the next section, we discuss the important methods to analyze facial expressions.

The “facial action coding system (FACS)” (developed by Ekman and Friesen in 1978 [7]) is the most used technique to code facial expressions in the behavioral sciences [15], and it is a research tool to measure facial expression [9]. In this method, we describe 46 components movements corresponding to the movement of individual facial muscle. As shown in Figure 1 [8], the method provides a comprehensive technique to investigate the elementary component of expressions, assisting in decomposing a complete speech into phonemes, proven to discover facial movements that reveal emotional and cognitive states [13]. Manual coding is the major shortcoming of FACS as this is time inefficient. Moreover, it is cost-ineffective as it consumes more than 100 hours of training to master and proficient FACS tasks. The manual coding of video takes also nearly two hours. FACS can assist in describing the facial expressions and Action units (AU). This represents a small fraction of the method to build facial expression. In essence, FACS AU depends on the action of facial muscles [10]. AU is a small fraction that cannot be interpreted as originated from the action [11].

EMG (Electromyography) is a medical diagnostic technique to assess nerve cells and muscles as well as recording and evaluating the electrical activity from skeletal muscle dysfunction and nerve dysfunction [16]. The EMG application is carried out with an electromyogram to record electrodes and translate an electrical signal into a graph. The EMG can detect the potential signal when myocytes are activated by electricity or nerves. Signals help to detect medical abnormalities. Moreover, the activation level help to evaluate the “biomechanics of human or animal movement” [12].

Facial EMG offers a more precise method to record and measure facial expression than visual observation. EMG is effective in measuring muscle activity, which helps to amplify and detect small electrical pulses emitted through the contraction of muscle fibers. Through the electrical activity, EMG can be used to measure small changes in the facial muscles, thereby reflecting small muscle movements. In this, EMG can measure both facial activities and produce weak emotional stimuli. Moreover, EMG plays a leading role in biomedical applications and clinical diagnosis [17].

3. System Design and Architecture

The IoT architecture consists of a variety of technologies, such as cloud services, sensors, communication layers to understand how different technologies are linked and set-up through IoT arrangement [9]. All layers of IoT are depicted below:

Application layer: The application provides a user experience.

Management service layer: Certain information processes are completed through security control, analysis, management of the equipment, and process modeling.

Gateway layer: This layer applies powerful and high-performance wired or wireless network infrastructure such as vehicle media.

Sensor connection and network: This layer can interconnect the digital and physical world to collect and process real-time data.

4. Experimental design

The purpose of the experiment was to collect facial muscle data. . (see Fig. 1) the process and hardware or environment for recording facial EMG signals. Physiological EMG recordings were performed in university laboratories.

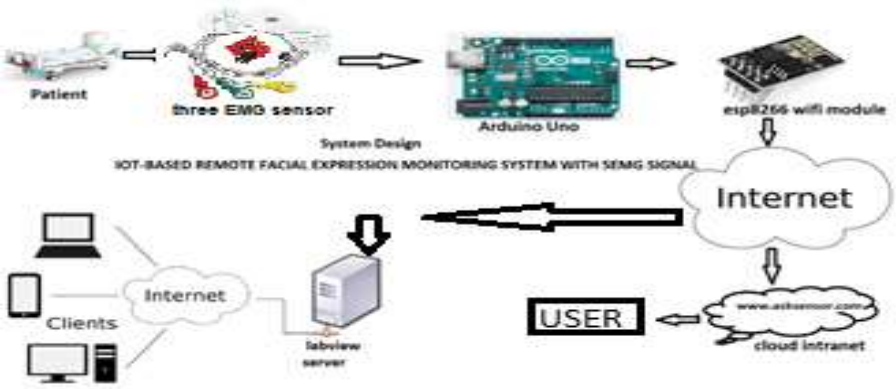


Fig. 2. System design.

5. Classification

In classification, we extract features in facial expressions . (see Fig.3). Different factors lead to misclassification. These factors are sweat, electrodeposition, and muscle fatigue, this cause variation in the EMG pattern and misclassification. A classifier can address these flaws. However, the classifier must be proficient to address real-time restraints and classify the new patterns in online training. We apply different classification algorithms based on pattern recognition. These are applied to classify facial expressions.

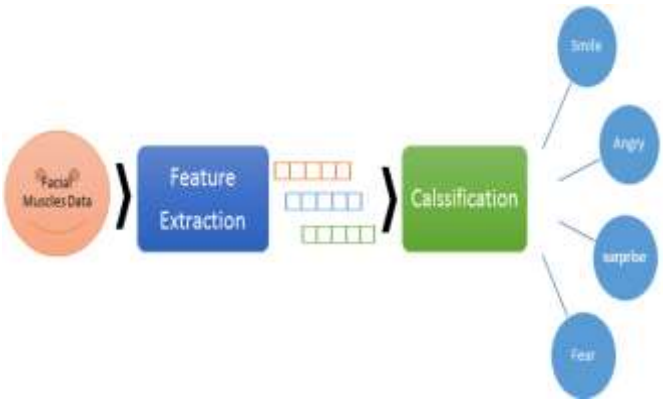


Fig. 3. Facial Expression Analysis and Classification.

Figure 9 shows the steps to process sEMG data to train a classifier. From three EMG sensors, the sEMG signal is sampled at 1000 SPS. After sampling, we apply a 50Hz notch Butterworth filter and 20Hz high-pass Butterworth filter to each signal to reduce the effects of artifacts coupled to the wires and power line interference. Before the RMS feature extraction and construction of the feature matrix, the data is segmented into 200-millisecond slices. We extract the RMS features using the following formula:

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$$

Training expression classification is carried out on multivariate naive Bayes classifiers. For classification tasks, the probabilistic ML model is the Naive Bayes classifier. The Bayes' theorem is based on the classifier.

Bayes' theorem:

$$P(A\backslash B) = \frac{P(B\backslash A)P(A)}{P(B)}$$

Gaussian Naive Bayes:

With predictors continuous and not in discrete values, we assume that these values are sampled from a Gaussian distribution. . (see Fig. 4) the training of multiple classifiers for expression classification. From the training data,

the feature matrix estimates the Gaussian distribution parameters of each expression. In the test data, we calculate the posterior probability of the category c for pattern recognition [23s].
The univariate Gaussian Bayes' theorem is the equation below, where the continuous random variable x of the probability density function for a category c is Gaussian with mean π_c and variance σ_c^2.

$$P(c\backslash x) \propto \frac{1}{2\pi\sigma_c}\exp\left(\frac{-(x-\mu)^2}{2\sigma_c^2}\right)P(c)$$

$\mu = Mean$
$\sigma = Standard\ Deviation$
$\pi \approx 3.14159 \ldots$
$e \approx 2.71828 \ldots$

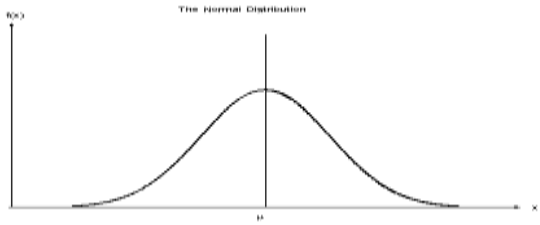


Fig. 4. Gaussian Distribution (Normal Distribution).

6. RESULT AND ANALYSIS

The results of facial expressions of Person 1, Person, 2, and Person 3 are provided in Mean, Standard Deviation, Skew, and time series. The results show that Person 3 records the highest signal in Smile, Fear, Angry, and Surprise when compared its Mean to that of Person 1 and Person 2 as revealed in Figure 6 and Table 1. Moreover, Person 3 records the highest standard deviation and skewness compared to the other two. The time series of entire Person 1, Person, 2, and Person 3 is provided in Figures 6, 7, and 8. Figure 5 provides the average of Person 1, 2, and 3

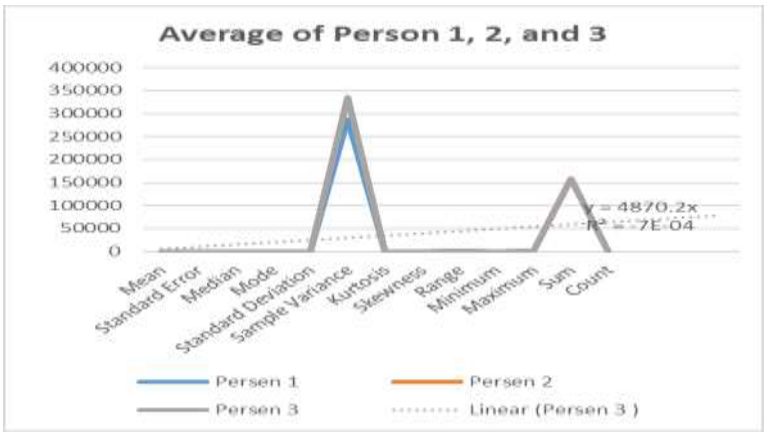


Fig. 5. Average of Person 1, 2, and 3.

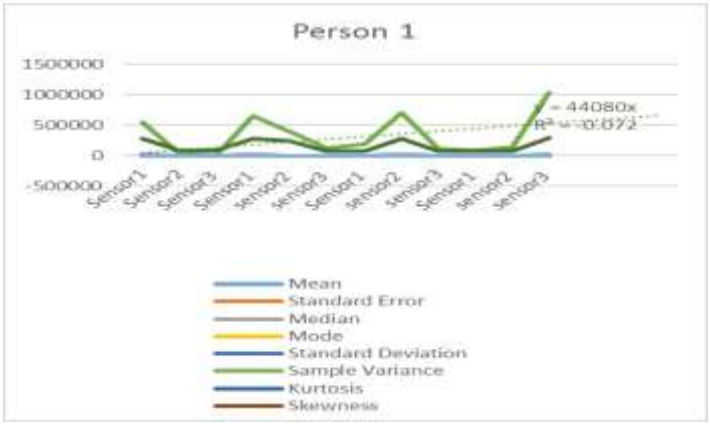


Fig. 6. Time Series of Person 1.

Fig. 7. Time Series of Person 2.

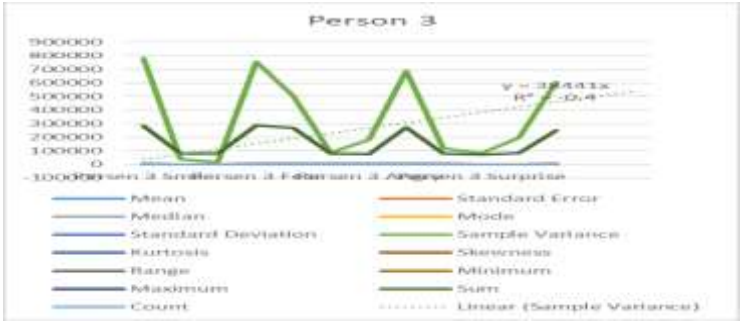


Fig. 8. Time Series of Person 3.

Table 1. Average of Person 1, 2, and 3.

Average	Person 1	Person 2	Person 3
Mean	782.4909	789.5845	790.7773
Standard Error	33.93343	35.99604	36.28239
Median	728.5	719.4167	701.25
Mode	700.9167	743.5	564.3333
Standard Deviation	481.0895	510.332	514.3917
Sample Variance	288547.2	335172.8	337049.2
Kurtosis	4.641563	3.576313	3.305591
Skewness	1.394661	1.165878	1.257305
Range	2360.667	2366.333	2402.667
Minimum	22.5	30.58333	26.25
Maximum	2383.167	2396.917	2428.917
Sum	157280.7	158706.5	158946.3
Count	201	201	201

Using supervised machine learning, a remote multi-channel biopotential monitoring system is implemented based on IoT. Biopotential measurement equipment carries wireless data streaming. We measure Three-channel biopotential according to application requirements. However, the current system supports full eight channels of wireless data transmission and online processing, where bytes are transmitted at a 24KB / s data rate. Moreover, the system in other multi-channel biopotential applications requires remote monitoring and online data . (see Fig. 9).

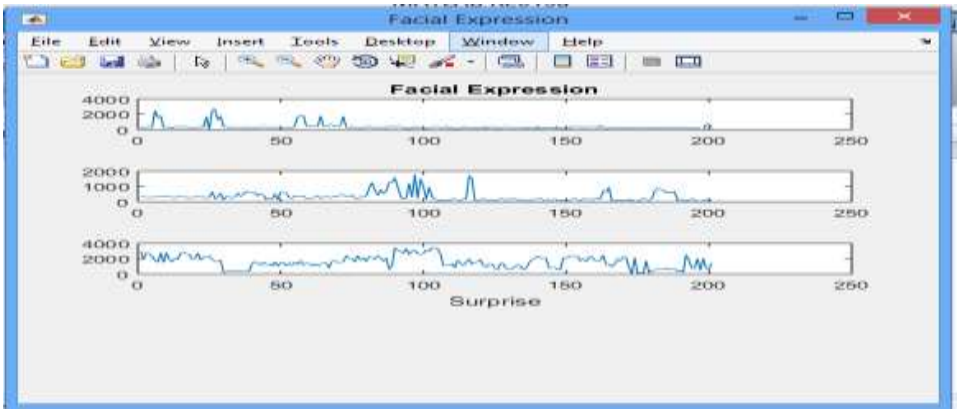


Fig. 9.. EMG data stream processing and classification in Matlab.

7. Conclusion

EMG measurement technology is a critical tool for clinical purposes to detect facial expressions because when someone has a facial expression, their facial muscles change, and this method can be used for detection and analysis. However, FACS is a time-consuming technology that requires expert knowledge or training. Our current work requires further improvement.

8. Future work

This project can be applied to the patient. The signals show emotional excitement (fear, surprise, anger, or smile) / injury or false. Our main goal is to analyze the emotional facial expressions/injuries of the intensive care unit (ICU) and newborns. To distinguishing newborn pain from crying is a challenge. Another important issue is that newborns grow very fast; thus, we must change the size according to their age. It is also important to have a system to automatically detect pain to increase the efficiency and overhead associated with monitoring patient progress.

REFERENCES

1. Jiang, M., Rahmani, A. M., Westerlund, T., Liljeberg, P. & Tenhunen, H. Facial expression recognition with sEMG method. Proc. - 15th IEEE Int. Conf. Comput. Inf. Technol. CIT 2015, 14th IEEE Int. Conf. Ubiquitous Comput. Commun. IUCC 2015, 13th IEEE Int. Conf. Dependable, Auton. Se 981–988 (2015) DOI:10.1109/CIT/IUCC/DASC/PICOM.2015.148.
2. Yang, G. et al. IoT-Based Remote Pain Monitoring System: From Device to Cloud Platform. IEEE J. Biomed. Heal. Informatics 22, 1711–1719 (2018).
3. Yang, G. et al. An IoT-Enabled Stroke Rehabilitation System Based on Smart Wearable Armband and Machine Learning. IEEE J. Transl. Eng. Heal. Med. 6, 1–10 (2018).
4. Hamedi, M., Salleh, S. H., Ting, C. M., Astaraki, M. & Noor, A. M. Robust Facial Expression Recognition for MuCI: A Comprehensive Neuromuscular Signal Analysis. IEEE Trans. Affect. Comput. 9, 102–115 (2018).
5. Instrumentation, E. & Singh, Y. ANALYSIS AND CLASSIFICATION OF EMG SIGNAL USING LabVIEW WITH DIFFERENT WEIGHTS Master of Engineering in Electronics Instrumentation and Control Submitted by Yadvendra Singh Under the Guidance of : Department of Electrical and Instrumentation Engineering T. (2018).
6. Boot, L. & Muhl, C. Facial expressions in EEG/EMG recordings. Dep. Hum. Media Interact. Master of, 112 (2009).
7. Chiaberge Sebastián Aced López, M. Design and Construction of an Emg Multichannel Acquisition System Prototype. (2012).
8. Yang, G. et al. IoT-Based Remote Pain Monitoring System: From Device to Cloud Platform. IEEE J. Biomed. Heal. Informatics 22, 1711–1719 (2018).
9. Science, C. & Publications, S. Surface Electromyography-Based Facial Expression Recognition in Bi-Polar Configuration Mahyar Hamedi , Sh-Hussain Salleh , Tan Tian Swee , and Kamarulafizam Department of Biomedical Instrumentation and Signal Processing , Faculty of Biomedical and Health. 7, 1407–1415 (2011).
10. Muhammad, G., Alsulaiman, M., Amin, S. U., Ghoneim, A. & Alhamid, M. F. A Facial-Expression Monitoring System for Improved Healthcare in Smart Cities. IEEE Access 5, 10871–10881 (2017).
11. Verma, L. T. O. Emg Analysis of Facial Activity. (2015).
12. Perdiz, J., Pires, G. & Nunes, U. J. Emotional state detection based on EMG and EOG biosignals: A short survey. ENBENG 2017 - 5th Port. Meet. Bioeng. Proc. (2017) doi:10.1109/ENBENG.2017.7889451.
13. Boccignone, G., Cuculo, V., Grossi, G., Lanzarotti, R. & Migliaccio, R. Virtual EMG via facial video analysis. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 10484 LNCS, 197–207 (2017).
14. Yang, G. et al. An IoT-Enabled Stroke Rehabilitation System Based on Smart Wearable Armband and Machine Learning. IEEE J. Transl. Eng. Heal. Med. 6, (2018).
15. Chen, Z., Ansari, R. & Wilkie, D. Learning Pain from Action Unit Combinations: A Weakly Supervised Approach via Multiple Instance Learning. IEEE Trans. Affect. Comput. PP, 1–1 (2019).
16. Boxtel, A. Van. Facial EMG as a tool for inferring affective states. Proc. Meas. Behav. 2010, 104–108 (2010).
17. Djelouat, H., Amira, A. & Bensaali, F. Compressive sensing-based IoT applications: A review. J. Sens. Actuator Networks 7, (2018).

Extreme Gradient Boosting for Multivariate Wind Speed Prediction

Ayşe Tuğba Dosdoğru and Aslı Boru İpek

Department of Industrial Engineering, Adana Alparslan Türkeş Science and Technology University, 01250, Adana, Turkey
adosdogru@atu.edu.tr

Abstract. Nowadays, one of the most important renewable energy sources is wind energy that is remarkable importance among the low-carbon energy technologies. The primary aim of wind energy production is to reduce dependence on fossil fuels that affect environment adversely. Therefore, wind energy is analyzed to develop new energy resources. The main issue related to evaluation of the wind energy potential is wind speed prediction. Due to the high volatility and irregularity, wind speed prediction is difficult. To cope with complex data structure, this study presents the development of extreme gradient boosting (XGBoost) within Grid Search parameter optimization for daily average wind speed prediction. Also, results are compared with artificial neural network (ANN). The results showed that XGBoost exhibits superior CPU time performance than that of ANN.

Keywords: Extreme Gradient Boosting, Artificial Neural Network, Wind Speed Prediction.

1 Introduction

In the early 1970s, the first worldwide oil crisis was occurred. On account of the crisis, energy supply has become tighter and tighter. In order to reduce fossil fuels usage and to increase renewable energy usage, many countries have enacted laws. In the twenty-first century, the crude oil price has been dramatically increased in the international oil market. Furthermore, the oil resources have been decreasing. Hence, renewable energy has attracted more and more attention from many countries [1]. Especially, wind energy assures the successful operation of multi-megawatt sized wind turbines. It is also competitive with traditional fossil fuel power plants [2]. Countries' economic growth is influenced by wind energy that generates additional jobs and technology development. For this reason, wind industry has been developing very quickly in the past decade. In addition, its ratio in the total energy production increases every year [3].

To effectively analyze wind energy potential, the statistical properties of wind and the determination of appropriate wind turbines are important [4]. Timely updates and

innovative design of turbine help improving the turbine efficiency. In this manner, the energy produced by wind can be maximized. However, we should analyze the benefits or drawbacks between wind and conventional power systems. For example, economic analyses such as the turbine system cost are important to evaluate the new technical solutions [3].Wind energy is proportional to the cube of wind speed. Therefore, wind speed prediction is important to wind energy evaluation and wind farm design [5].

Efficient and effective prediction of wind speed can provide enough information for decision-makers to evaluate power generation of wind turbine. Using predicted values, they can make a detailed planning to provide the maximum yield of wind energy. For example, if the values of wind speed are smaller (or larger) than wind turbine capacity, the turbines should be shut down to reduce the operating costs [6]. In recent past, significant amount of research has been focused on prediction of the wind speed. Single linear statistical model cannot describe wind speed time series comprehensively. The climate conditions and the geomorphic characteristics directly affect the wind speed time series. For this reason, different regions can have different characteristics. In literature, no one best prediction model is available for wind speed prediction [7]. Hence, it is strongly recommended that different models should be employed to provide the best result for considered region [8]. At this point, ANN is gaining popularity for wind energy evaluation in recent years. Especially, it has been used to predict wind speed and then, predicted wind speed has been employed to estimate the energy potential. In addition, ANN can be directly used to estimate the energy potential [9]. In this paper, XGBoost and ANN are employed to predict wind speed. Furthermore, the performances of the proposed methods are compared to evaluate the prediction results.

2 Literature Review

The utilization of fossil fuels negatively influence environment. On the other hand, renewable energy sources have a much lower environmental impact. Therefore, increased attention has been paid to increase the efficiency of renewable energy sources. Especially, wind speed prediction is of increasing importance since wind power generation, power grid operation scheduling, and wind farm planning directly influenced by wind speed [10]. In literature, a variety of approaches and methods has been created to predict wind speed. Wu and Hong [11] presented literature review related to wind prediction including persistence models, numeric weather prediction, statistical methods, ANN methods, and hybrid methods. Okumus and Dinler [12] presented an extensive review related to wind speed and power prediction. Hybridization of different methods is favorable to increase the accuracy of prediction methods. The used method for wind speed prediction can change according to the available information, considered regions, timescale of the prediction. We summarized some studies related to wind speed prediction in chronological order as follows.

More and Deo [13] developed neural networks to forecast monthly, weekly and daily wind prediction. In the study, back-propagation and cascade correlation were used in training processes of neural networks. Torres et al. [14] evaluated the applicability of

the autoregressive moving average (ARMA) to predict hourly average wind speed. The transformation and standardization of dataset were firstly applied. Then, ARMA model was created and its performance was compared with persistence model. Barbounis et al. [15] presented the long-term wind speed and power prediction. Three types of recurrent neural networks were constructed to evaluate seventy two–hours-ahead wind forecasts. They also introduced two novel learning algorithms that have considerably smaller computational and storage requirements. Results demonstrated that proposed method produce robust multi-step ahead predictions.

Similarly, Barbounis and Theocharis [16] presented the long-term wind speed and power prediction. A global recursive prediction error model that combines the atmospheric and the time-series method was developed. In addition, three simplified versions were created to decrease the complexity of proposed method. Hocaoglu et al. [17] developed hidden Markov models to model the wind speed. The paper aim was to create the time transition properties of the wind speed data. Proposed method was tested on hourly and daily data. The results demonstrated that the daily data based model was better than the hourly data based model. The reason of this situation was that daily averages exhibited a smoother variation. Guo et al. [1] proposed Generalized Autoregressive Conditional Heteroscedasticity (GARCH) based methods and ARMA based methods to predict long-term wind speed. Mohandes et al. [2] used neuro-fuzzy method to predict wind speed at high altitudes based on measurements at lower heights.

Cassola and Burlando [18] proposed numerical weather prediction based on the kalman filtering method to predict the daily-averaged wind speed and daily wind power production. Different kinds of Kalman filter were implemented to test the proposed method. They also showed how the performance of the filter varies according to the prediction period. Philippopoulos and Deligiorgi [19] developed two feed forward neural networks model to estimate mean hourly wind speed values. In the paper, different spatial interpolation methodologies were used to compare the results of proposed methods. Tatinati and Veluvolu [20] presented the hybrid method including three stages. Empirical mode of decomposition was firstly employed to decomposed signal into meaningful local time scales. The decomposed components were predicted. Finally, the final prediction result was obtained using the aggregation of predictions. In the study, the two data sets were used to perform single-step prediction and six-step ahead prediction. Jung and Kwon [9] presented two new ANNs that employ weighted error functions to enhance the behavior of the long-term wind speed prediction. In the paper, the energy potential was computed with the predicted wind speed. Shamshirband et al. [21] used three adaptive neuro-fuzzy inference system (ANFIS) networks for wind speed estimation. Three ANFIS networks were created using different input variables but their output was same, instantaneous wind speed.

Azad et al. [22] predicted mean hourly wind speed values using a combination of different approaches. Firstly, average monthly wind speed for the past four years were used to create neural network and hourly data from the previous year were used to create statistical analysis. Then, yearly trend, monthly trend, and daily trend were employed to construct neural network that also included data modification. Lydia et al. [23] employed the autoregressive (AR) model, AR model with exogenous variable

and non-linear ARX model as wind speed prediction models. Linear, wavenet, neuralnet, sigmoidnet, and tree partition were applied as modeling techniques. They used power curve models and wind speed to estimate the wind resource. Wang et al. [24] used Cuckoo Search and Chaotic Particle Swarm Optimization to determine the parameters of distributional functions. In the paper, extreme wind speeds were predicted using the analysis of the monthly and quarterly maximum wind speeds.

Noorollahi et al. [8] developed the ANN methods to predict the wind speed in three wind observation stations. Three variable selection methods were employed to determine the variables. The trial and error approach was utilized to determine the learning parameters of ANN. Doucoure et al. [25] proposed multi-resolution analysis and adaptive wavelet neural network based method with Hurst predictability analysis to analyze the wind speed profiles. They used hourly wind speed data to test the approach. Wang et al. [26] proposed wind speed prediction approach that includes wavelet transform, deep belief network and spine quantile regression. Firstly, wavelet transform was employed to decompose the wind speed data into four frequencies. Then, deep belief network was designed for each wind speed frequency. Finally, the probabilistic model was statistically created based on the spine quantile regression.

Peng et al. [27] proposed hybrid methodology including four steps. In first step, the two-stage decomposition algorithm was used to denoise the original wind speed time series. AdaBoostELM models were created and trained in second step. To select the input variables, the partial autocorrection function values were used. Thirdly, proposed models were implemented to the test datasets. In final step, the final prediction was obtained.

The contribution of this study can be summarized as follows. Firstly, XGBoost with Grid Search is used to predict the wind speed. Secondly, ANN with Grid Search is employed for wind speed prediction. Thirdly, the performance of proposed methods is compared to evaluate the average wind speed prediction.

3 Materials and Methods

3.1 Dataset

Table 1. Descriptive statistics of features used in the study.

	Average Temperature	Average Dew Point	Average Wind Speed
count	3181	3181	3181
mean	3.312584	-1.588557	18.576705
std	14.32844	13.157638	6.910508
min	-35.2	-40.1	4.5
25%	-7.75	-10.6	13.5
50%	5.19	-0.2	18
75%	16.14	9.8	23
max	28.3	21.4	47.5

In this paper, dataset covers the period from 01.01.2011 up to 16.09.2019. Average dew point and average temperature are used as features. Also, average wind speed is selected as target in proposed methods. Descriptive statistics of features are given in Table 1 that includes count, mean, standard deviation (std), minimum (min), first quartile (25%), second quartile (50%), third quartile (75%), maximum (max).

3.2 Proposed Methodology

In this paper, XGBoost is used to predict daily average wind speed. Due to the difficulty of determining parameters of XGBoost algorithm, Grid Search with 3-fold cross validation is used. In XGBoost, colsample_bytree (the subsample ratio of columns when constructing each tree), maximum depth of a tree, the number of trees in the forest, reg_alpha (L2 regularization term on weights), reg_lambda (L1 regularization term on weights), and subsample (the subsample ratio of the training instances) are determined. Also, results are compared with ANN. Likewise Gird Search is used to optimize the parameters of ANN that includes the epoch number, the number of neuron in hidden layer, and the activation function. In addition, MinMaxScaler is used to transform features to a given range by scaling each feature. Details about the Grid Search, XGBoost, and ANN are given in following subsections.

3.2.1 Grid Search

Any possible combination of hyper-parameters is implemented in Grid Search using the trial and error procedure and the combination. This process results in a relatively high run-time. However, this process outperforms the manual approach. A parameter tuning based on Grid Search is used to evaluate each combination of parameters to figure out which combination provides the true minimum. Additionally, Grid search is completely automatic until the entire grid is completed [28].
A simple way of setting up a Grid Search is to define a lower bound vector and an upper bound vector for parameters. Grid Search requires taking n points equally spaced at each interval of the form between the lower bound and the upper bound. This generates a total possible grid points to control. Details about the Grid Search can be found in Vinod, and Rao [29].

3.2.2 Extreme Gradient Boosting (XGBoost)

XGBoost can be described as a scalable end-to-end tree boosting system. It is used as an open-source package. The key factor behind XGBoost's success is its scalability across all scenarios. XGBoost's scalability is attributed to many significant structures and algorithmic optimizations. On a single machine, the system operates more than ten times faster than existing popular solutions. Parallel and distributed computing makes the learning process faster that allows for faster exploration of the model. XGBoost takes advantage of out-of-core computation, allowing data scientists to process hundreds of millions of examples on a desktop. Finally, integrating these methods to create an end-to-end system is much more exciting [30]. Details about XGBoost can be found in Chen and Guestrin [30].

In this paper, Grid Search determines the parameters of XGBoost. Note that Grid Search optimizes the parameters according to mean squared error (MSE). It is determined that the value of colsample_bytree (the subsample ratio of columns when constructing each tree) is 0.7. Maximum depth of a tree is determined as 15. The number of trees in the forest is 400. In addition, reg_alpha (L2 regularization term on weights), reg_lambda (L1 regularization term on weights), and subsample (the subsample ratio of the training instances) are specified as 1.3, 1.1, and 0.9, respectively. MinMaxScaler is used to transform features by scaling each feature between 0.1 and 0.9. Note that dataset is divided into two separate parts as training (80%) and testing (20%).

3.2.3 Artificial Neural Network (ANN)

In simple ANN, an input layer, hidden layer(s), and an output layer are used to create the network. The ANN accepts the independent variables in the input layer. In the output layer, predictions are generated. ANN is capable of modeling a complex non-linear relationship and extracting the input-output dependency. ANN can learn from past data. It also uses these to predict future values. It is easy to create the model and requires only short development times. However, the performance of the ANN methods directly depends on the parameters of the ANN. Adam is used as optimization algorithm, and other parameters are determined by Grid Search. Note that Grid Search is optimized parameters according to MSE. The epoch number is determined as 1000. Single layer is used in ANN and the number of neuron in hidden layer is specified as 64. In addition, rectified linear unit is determined by the Grid Search as the activation function. MinMaxScaler is used to transform features by scaling each feature between 0.1 and 0.9. Note that dataset is divided into two separate parts as training (80%) and testing (20%).

4 Results and Discussion

To develop sustainable energy and protect environment, wind energy has become increasingly significant as a renewable and clean energy source. At this point, wind speed plays a crucial role in wind power generation and wind farm planning, but also economic and social benefits.

In this paper, Grid Search is used to optimize parameters of XGBoost and ANN. The results of cross-validation on train dataset are given in Table 2. The results indicated that Grid Search with XGBoost required significantly less CPU time than Grid Search with ANN even though less number of parameters are optimized for ANN. Also, MSE values of two algorithms are relatively close.

Table 2. The results of the Grid Search optimization.

Methods	MSE	CPU time
XGBoost	0.017045	3.7 minutes
ANN	0.016232	10.9 minutes

The results for test data including Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are given in Table 3. According to the results of the performance measures, the ANN performance is slightly better than XGBoost. Therefore, XGBoost and ANN can be efficiently utilized to predict the average wind speed. The algorithms are run on a computer with i7 processor technology, 2.60 GHz processor speed and 16 GB RAM capacity.

Table 3. Other performance measures for test data.

Performance Measures	XGBoost	ANN
MSE	0.01933	0.01852
RMSE	0.13905	0.13610
MAE	0.11025	0.10839

5 Conclusion

Wind speed has a high autocorrelation and inherent volatility. Therefore, many researchers have been tried to improve the performance of wind speed prediction. In this case, Grid Search, XGboost, and ANN can be efficiently used to cope with real world problems. In this study, the results showed that XGboost's outperforms ANN in terms of CPU time performance. This property is significantly important especially when parameter optimization is required. Besides this, XGBoost achieved almost same results with ANN on multivariate wind speed prediction.
In future work, other optimization methods, such as metaheuristics, can be used when optimizing XGBoost parameters in order to get better performance on wind speed prediction. Also, other features on dataset can be used or a feature selection method can be applied.

References

1. Guo, Z., Dong, Y., Wang, J., Lu, H.: The forecasting procedure for long-term wind speed in the Zhangye area. Mathematical Problems in Engineering, Article ID 684742 (2010).
2. Mohandes, M., Rehman, S., Rahman, S. M: Estimation of wind speed profile using adaptive neuro-fuzzy inference system (ANFIS). Applied Energy, 88(11), 4024-4032 (2011).
3. Wais, P.: A review of Weibull functions in wind sector. Renewable and Sustainable Energy Reviews, 70, 1099-1107 (2017).
4. Wang, J., Hu, J., Ma, K.: Wind speed probability distribution estimation and wind energy assessment. Renewable and Sustainable Energy Reviews, 60, 881-899 (2016).
5. Sun, S., Qiao, H., Wei, Y., Wang, S.: A new dynamic integrated approach for wind speed forecasting. Applied Energy, 197, 151-162 (2017).
6. Wang, J., Du, P., Niu, T., Yang, W.: A novel hybrid system based on a new proposed algorithm—Multi-objective whale optimization algorithm for wind speed forecasting. Applied Energy, 208, 344-360 (2017).
7. Guo, Z., Zhao, J., Zhang, W., Wang, J.: A corrected hybrid approach for wind speed prediction in Hexi Corridor of China. Energy, 36(3), 1668-1679 (2011).

8. Noorollahi, Y., Jokar, M. A., Kalhor, A.: Using artificial neural networks for temporal and spatial wind speed forecasting in Iran. Energy Conversion and Management, 115, 17-25 (2016).
9. Jung, S., Kwon, S.-D.: Weighted error functions in artificial neural networks for improved wind energy potential estimation. Applied energy, 111, 778-790 (2013).
10. Wang, J., Qin, S., Zhou, Q., Jiang, H.: Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China. Renewable Energy, 76, 91-101 (2015).
11. Wu, Y. K., Hong, J. S.: A literature review of wind forecasting technology in the world. In 2007 IEEE Lausanne Power Tech, pp. 504-509, Lausanne (2007).
12. Okumus, I., Dinler, A.: Current status of wind energy forecasting and a hybrid method for hourly predictions. Energy Conversion and Management, 123, 362-371 (2016).
13. More, A., Deo, M. C.: Forecasting wind with neural networks. Marine Structures, 16(1), 35-49 (2003).
14. Torres, J. L., Garcia, A., De Blas, M., De Francisco, A.: Forecast of hourly average wind speed with ARMA models in Navarre (Spain). Solar Energy, 79(1), 65-77 (2005).
15. Barbounis, T. G., Theocharis, J. B., Alexiadis, M. C., Dokopoulos, P. S.: Long-term wind speed and power forecasting using local recurrent neural network models. IEEE Transactions on Energy Conversion, 21(1), 273-284 (2006).
16. Barbounis, T. G., Theocharis, J. B.: Locally recurrent neural networks for long-term wind speed and power prediction. Neurocomputing, 69(4-6), 466-496 (2006).
17. Hocaoğlu, F. O., Gerek, Ö. N., Kurban, M.: A novel wind speed modeling approach using atmospheric pressure observations and hidden Markov models. Journal of Wind Engineering and Industrial Aerodynamics, 98(8-9), 472-481 (2010).
18. Cassola, F., Burlando, M.: Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output. Applied Energy, 99, 154-166 (2012).
19. Philippopoulos, K., Deligiorgi, D.: Application of artificial neural networks for the spatial estimation of wind speed in a coastal region with complex topography. Renewable Energy, 38(1), 75-82 (2012).
20. Tatinati, S., Veluvolu, K. C.: A hybrid approach for short-term forecasting of wind speed. The Scientific World Journal, Article ID 548370 (2013).
21. Shamshirband, S., Petković, D., Anuar, N. B., Kiah, M. L. M., Akib, S., Gani, A., Ćojbašić, Z., Nikolić, V.: Sensorless estimation of wind speed by adaptive neuro-fuzzy methodology. International Journal of Electrical Power & Energy Systems, 62, 490-495 (2014).
22. Azad, H. B., Mekhilef, S., Ganapathy, V. G.: Long-term wind speed forecasting and general pattern recognition using neural networks. IEEE Transactions on Sustainable Energy, 5(2), 546-553 (2014).
23. Lydia, M., Kumar, S. S., Selvakumar, A. I., Kumar, G. E. P.: Wind resource estimation using wind speed and power curve models. Renewable Energy, 83, 425-434 (2015).
24. Wang, J., Qin, S., Jin, S., Wu, J.: Estimation methods review and analysis of offshore extreme wind speeds and wind energy resources. Renewable and Sustainable Energy Reviews, 42, 26-42 (2015).
25. Doucoure, B., Agbossou, K., Cardenas, A.: Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data. Renewable Energy, 92, 202-211 (2016).
26. Wang, H. Z., Wang, G. B., Li, G. Q., Peng, J. C., Liu, Y. T.: Deep belief network based deterministic and probabilistic wind speed forecasting approach. Applied Energy, 182, 80-93 (2016).

27. Peng, T., Zhou, J., Zhang, C., Zheng, Y.: Multi-step ahead wind speed forecasting using a hybrid model based on two-stage decomposition technique and AdaBoost-extreme learning machine. Energy Conversion and Management, 153, 589-602 (2017).
28. Basnayake, W. M. N. D., Attygalle, D., Hansen, L. L., Nandala, K. D. W.: Grid search based parameter tuning of dynamic neural network to forecast daily reservoir inflow. In Proceedings of the International Conference on Computational Modeling and Simulation, pp. 291-295, Colombo, Sri Lanka (2017).
29. Vinod, H.D., Rao, C.R.: Handbook of Statistics Volume 41, Conceptual Econometrics Using R, (2019).
30. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 785-794, San Francisco, CA, USA (2016).

A Proposed Speaker Recognition Model Using Optimized Feed Forward Neural Network And Hybrid Time-Mel Speech Feature

S. Eltanashi[1] and F.ATASOY[2]

[1]Karabük University, Department of Computer Engineering, 78050, Karabük, Turkey
[2]Karabük University, Department of Computer Engineering, 78050, Karabük, Turkey

Abstract. Speaker Recognition Process is susceptible to a several challenges which are critical to speaker modeling. The main obstacle in speaker identification is the nature of voice signals. Such signals are termed by their randomness nature which can be caused by the time-varying nature of speech electric properties. The spectral information of speech varies overtime; therefore, it is difficult to rely only on the spectral domain in order to model a system for speakers recognition. In text-independent speaker recognition; frequency component analysis used to recognize the speakers. As voice signals are time-variant signal, the frequency spectrum information is changing by time. In this paper, Mel Frequency Cepstral Coefficient (MFCC) alongside Fundamental Frequency are implemented to propose text-independent speaker recognition system approximation. In addition, Feed Forward Neural Network (FFNN) used for speaker prediction. To further improve the performance, particle Swarm optimization algorithm was integrated by Freezing-FFNN. The simulation has shown that newly proposed technique, namely PSO-FFNN has achieved accuracy by 83.4 and reduced the Mean Square Error significantly.

Keywords: Prediction, FFNN, Neural, Features, Speaker.

1 Introduction

Speaker recognition is a vital stage in various personal authentication and security systems. It constructs the logic of individual verification using more precise bio-metric tools known as voice characteristics. Generally the speaker recognition system involves in two major stages known as feature extraction and speaker classification [1]. These processes might be initiated by voice pre-processing which involves in setting the voice signals [2]. Endpoint Detection (EPD) is one of the most popular algorithms which performs pre-processing of the voice signal. This algorithm finds the starting point and end point in speech signal. EPD algorithm determines the zero-crossing number in the spoken speech signal and hence it detects the noise level of the speech [3]. Speech features include time domain and frequency domain processing, both are integral part of speech processing and they can be used to form a final recognition system [4]. Speech signal pre-processing consists of signal enhancement by reducing the noise level, removing unnecessary components and clipping the distortion [5]. It involves silence removal which clips the low power signals that represents the silence in the uttered sentence, i.e.; breaks while speaking [6]. These processes are very important to enhance the signal quality which makes the signal more readable by the further stages. Pre-processing has remarkable effect in saving the computational resources, thus the performance of the entire system is improved. The several approaches have been proposed to perform features extraction for speech signal such as: Linear Predictive Cepstral Coefficients (LPCC), Real Cepstral Coefficient (RCC), Fundamental Frequency and Mel Frequency Cepstrum Coefficients (MFCC) [6–8]. Recent approaches empowered by MFCC outperformed all other methods such as LPCC and RCC [9]. Furthermore, artificial neural networks and deep learning approaches are widely adapted for mapping the features to a particular speaker [10]. The ultimate stage of processing is set to be system performance examination which monitors the system performance to recognize the speakers. The system is examined during the training and testing in order to evaluate the accuracy of speaker recognition and to highlight the error during the model training [11]. This paper proposes an improved speaker identification system which is based on Fundamental Frequency and Mel Frequency Cepstrum Coefficients (MFCC). The prediction model is treated under several stages for optimization purpose to further improve the performance by using Feed Forward Neural Network (FFNN), Freezing Feed Forward Neural Network (FFFNN) and particle swarm optimization (PSO). Fig.1 illustrates the overall speaker recognition system.

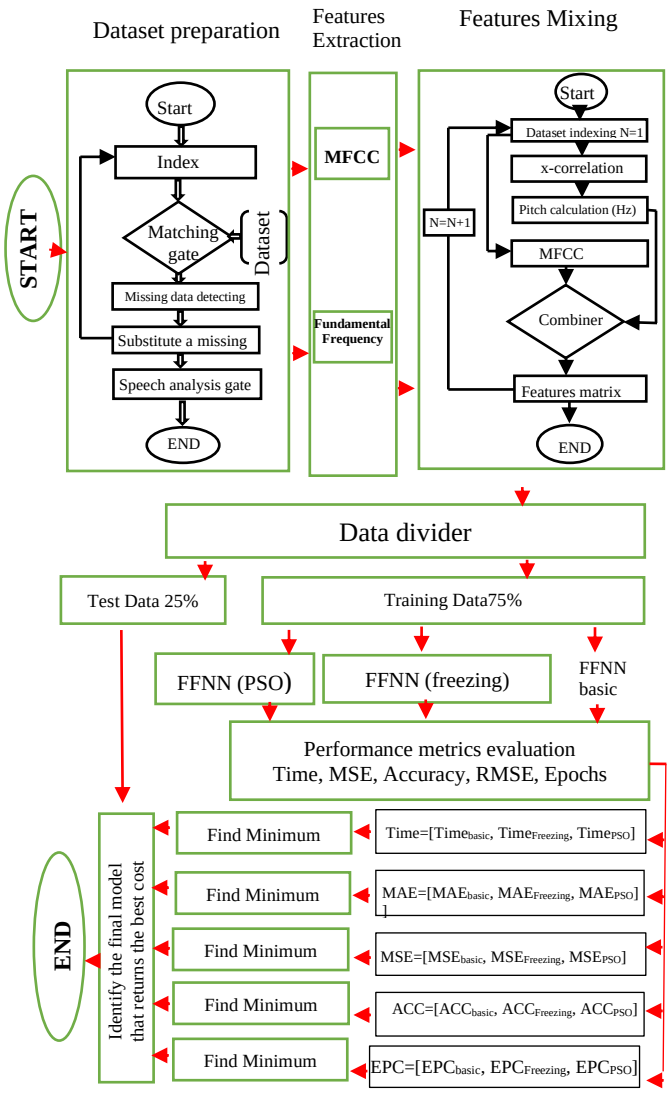


Fig.1: Overall process diagram of the speaker recognition system.

2 Voice Pre-Processing and Data Set

Pre-processing of the voice signal is the stage that applies some filtering to denoising and enhancing the signal before it is passed to the analyzers. To process the signal by digital means, the continuous-time signal must be filtered into a discrete-time (digital) valued signal. In pre-processing stage, the signal is sampled by converting it to a set of samples for efficient analysis [12]. Data set in the study contains 12 Males and 13 females, each speaker reads ten different sentences, and each sentence duration varies from 5 to 15 seconds. Speed sampling rate is 6 KHz. Dataset is obtained from Open Speech and Language Resources [13]. The dataset contains 250 voice clips recorded from 25 speakers and the same clips are sorted and named in a numerical or alphabetical form in order to feed them into the processing stage smoothly.

3 Feature Extraction

Features of the speech signal are generated from both MFCC method and the fundamental frequency methods i.e. pitch frequency. The speaker model was formulated by means of one feature from fundamental frequency and eight

features from MFCC. In addition, the dataset includes 25 records for different speakers and read sentence for each speaker.

3.1 Mel Frequency Spectrum Coefficients

MFCC is well-known algorithm in speech context domain; it simulates the range of human ear sensation level to a speech signal. Mel frequency is different from the local frequency of a signal. Mel spectrum coefficients are a formulation of the Mel set which represents the sensitivity of the human ear to a particular voice signal [14]. Consequently, each voice signal has different effects on the ear. Mel frequency spectrum coefficients represent the ear response in the form of a vector of eight values. Each element in the presented array represents a coefficient in the MFCC model. Mel frequency spectrum coefficient vectors are presented in the following expression.

$M_f = [m_1, m_2, m_3, m_4, m_5, m_6, m_7, m_8]$

Firstly, the voice signal is passed through a pre-emphasis filter which amplifies the low power samples as shown in Fig.2.

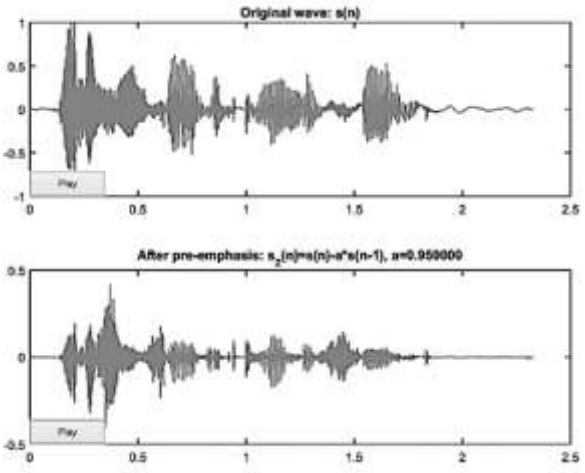


Fig. 2: Pre-emphasis filter input and output signals [15].

Speech signal remains stationary in a very short time frame approximately within 25 milliseconds. For this purpose, and in order to convert speech into a time-invariant signal, framing the signal is must. A 25 milliseconds frame window with 10 seconds overlapping frames for signal windowing are applied over the speech signal.

The last step in Mel frequency spectrum algorithm is to simulate the human ear perception with the voice signal. Hence, filter banks are used to mimic the non-linear human ear perception of sound. Filter bank with the transfer functions given is implemented to produce the human ear voice perception as shown in Fig.3. The flow diagram shown in Fig.4, demonstrates the process of the MFCC algorithm.

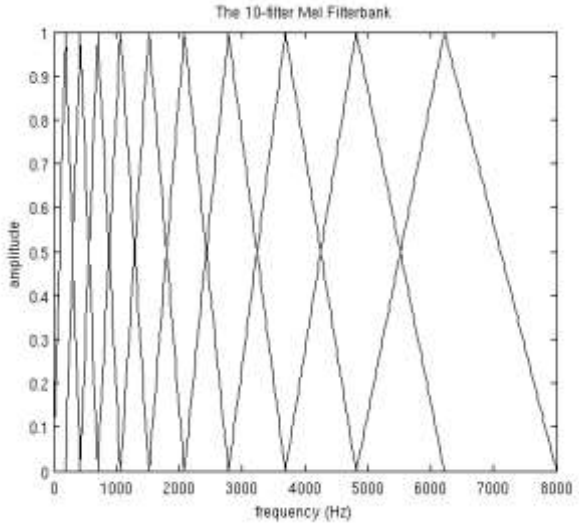


Fig.3 Filter bank response on Mel frequency scale [16].

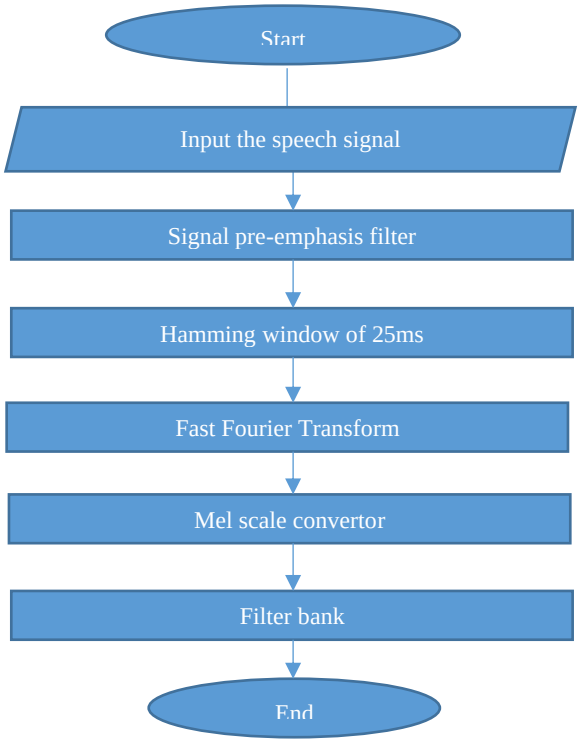


Fig.4: Mel frequency spectrum coefficients flow diagram.

3.2 Fundamental Frequency Coefficient

The fundamental frequency is one of the interesting features in speech signal domain; it is produced in time domain analysis using the cross-correlation approach. The aim of this feature is to identify the fundamental frequency in the speech signal [17].

The fundamental frequency is also called pitch frequency and calculated using the pitch period. This period is placed on the cross-correlation signal and represents the time between the minimum local maxima and maximum local maxima on the signal corps [2].

Assuming sampled speech signal is represented by S[n] and let the S'[n]=S(n-1) is the time-shifted copy of the same signal, cross-correlation is given in Equation (1) [18].

$$C[n] = \sum_{n=1}^{N} S[n].S'[n] \tag{1}$$

4 Classification

4.1 Feed Forward Neural Network

The neural network was implemented in many fields of sciences and engineering due to its numerous advantages [19]. Neural networks technology is a subset of artificial intelligence. This technology is rapid, accurate, and efficient. Neural as indicated by its name; it is mimicking the way that neural cells of human brain operate. Biologically, the neural cell initiates the order and passes it to high order cell through neurons. The neurons propagate the information from first layer to last layers. Artificial neural network (ANN) is the most popular terms in the machine learning area. ANN takes the same concept of biological brain cells to classify, predict, and cluster the data. The most popular type of artificial neural networks is the Feed-Forward neural networks (FFNN). This type of machine learning is mostly used for predicting future incidents or problems by learning about the incident/problem [20]. Fig.5 presents the topology of feed forward neural network; it consists of three layers which are input, hidden and output layer. Layers are built up with several nodes; the neurons are used to connect these layers.

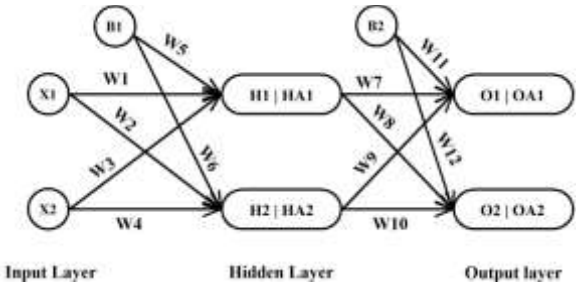


Fig.5: Feed Forward Neural Network structure [20].

In Fig 5 the input layer has two nodes $X1$ and $X2$, $W1$ to $W12$ is connection weight of the network, hidden layer has two nodes $H1$ and $H2$, output layer has two nodes $O1$ and $O2$, HA1 and $HA2$ are known as hidden node outputs and $OA1$ and $OA2$ are known as output node outputs. The hidden node values are expressed as:

$$\begin{aligned} H1 &= X1 * W1 + X2 * W3 + B1 * W5 \\ H2 &= X1 * W2 + X2 * W4 + B1 * W6 \end{aligned} \tag{2}$$

Let activation function in the hidden layer is sigmoid function; then node output of the hidden layer expressed as:

$$\begin{aligned} HA1 &= \frac{1}{1 + e^{-H1}} \\ HA2 &= \frac{1}{1 + e^{-H2}} \end{aligned} \tag{3}$$

The output node values are expressed as:

$$\begin{aligned} O1 &= HA1 * W7 + HA2 * W9 + B2 * W11 \\ O2 &= HA1 * W8 + HA2 * W10 + B2 * W12 \end{aligned} \tag{4}$$

Let activation function in the output layer is linear; then output of the output layer expressed as:

$$OA1 = O1$$
$$OA2 = O2 \quad (5)$$

For total error calculation, consider OAi is the output value of the output node i and yi is the output target then:

$$E = \frac{1}{2}\sum_{i=1}^{2}(yi - OAi)^2 \quad (6)$$

By employing propagation algorithm, weights of the connection ($W1$ to $W12$) are adjusted to minimize error between actual output value and target value [19].

Accordingly, features from Mel frequency spectrum coefficients methods are obtained and used for recognition process. For 250 speech signals and nine features for each signal, a total of 2250 features are generated from the speaker model. Table 1 illustrates the FFNN parameters customization to learn the features and hence to predict the speaker.

Table 1: FFNN First Experiment Parameters

Term	Values
Total layers	Three
Nodes distribution	Thirty, twenty and one
Learning algorithm	Built-in Levenberg–Marquardt (LM)
MSE goal	1e-200
Iteration	50

The first experiment conducted based on the parameters given in Table 1. During the training stage of this experiment, results were varying at every time model is restarted since the LM algorithm is distributing the weight values arbitrarily and it behaves the same whenever the same input output data is used. In order to monitor the model performance and to tackle this stochastic nature in the output, the experiment is repeated 100 times, results are recorded, and the performance of the model was evaluated by considering the mean of the findings.

4.2 Feed Forward Neural Network (Freezing)

The second experiment was performed according to the results obtained from the first one, the performance of a neural network is obtained for all 100 epoch and hence the weight of every epoch is recorded. However, weight freezing technology involves adjusting of the weight values of the FFNN model to return the best cost. The technology of freezing doesn't require any further training algorithms since readymade weights can be fed into the model with pre-determined performance. The selection of proper weight values depends totally on the previous experiment which involves the record of weights and their cost values.

4.3 Particle Swarm Optimization algorithm

The third experiment is made to further enhance the performance of the prediction. Therefore, a new algorithm is used as a training algorithm. Particle Swarm Optimization algorithm (PSO) is widely implemented in optimizing the feed-forward neural network due to its proven efficiency and remarkable performance [19]. PSO optimizes the performance as follows:

- Process is started by generating the set of weights (swarms). PSO will determine the best weight.
- Identifying the best weight is done by using the cost function; cost functions are the mean square error of the results obtained after training.

- PSO will search the weight that minimizes the cost function by setting the parameters such as velocity and inertia coefficients.
- The solution proposed by PSO will be proposed to the Feed Forward Neural Network as the best weight, which in turn applies the proposed weight to predict the speakers.

The flowchart for PSO optimized feed forward neural network is shown in Fig.6.

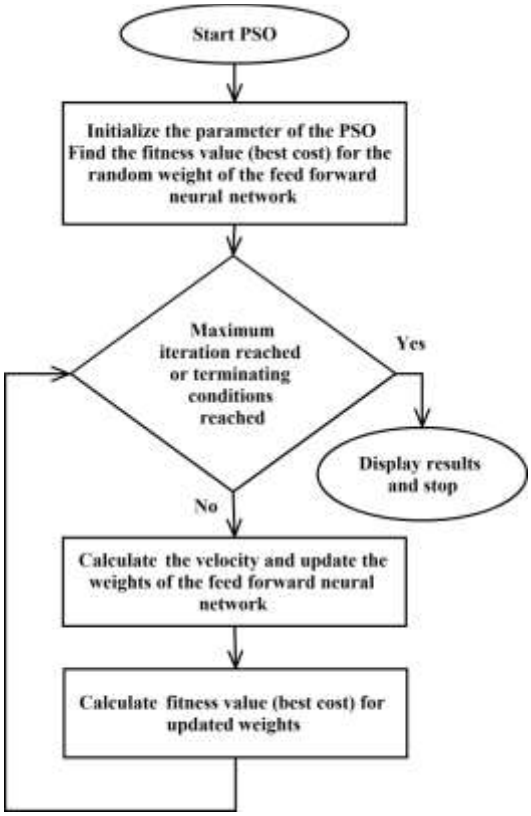


Fig.6: PSO optimized Feed forward neural network.

5 Result and Discussion

As discussed previously, FFNN is evaluated using several performance metrics in order to identify the best model which predicts the speaker's identity. Three models are simulated, namely: Plain Feed Forward Neural Network, Weight-Freezing based Feed Forward Neural Network and finally the Particle Swarm optimization based on Feed Forward Neural Network. The results of those models' performances are presented in the following sections. The performance of the Feed Forward Neural Network is evaluated periodically using a set of performance metrics which are decisive criteria for speaker recognition. The simulated performance metrics are briefly explained as follows:

5.1 Time and Accuracy Measurement

The Time in seconds required by the neural network to find the optimal solution and accuracy represents the total number of correct decisions (predictions) with respect to all other inputs. The accuracy and time are calculated in all three models and the results are presented in Table 2. The findings have shown that the best accuracy and time of speaker prediction is achieved by the PSO-FFNN model. Fig.7 shows the time and accuracy calculation for FFNN, Freezing-FFNN and PSO-FFNN.

Table 2: Time and Accuracy calculations for proposed run models.

Algorithm	Accuracy	Time (sec)
FFNN	59.68	3.983
Freezing-FFNN	63.04	3.27
PSO.FFNN	83.4	3.2639

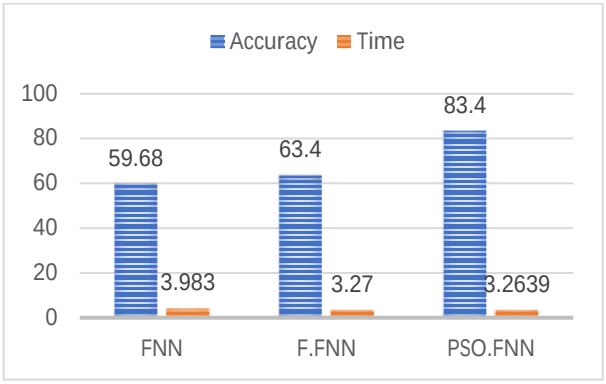


Fig.7: Time and Accuracy for the proposed models.

5.2 Mean Square Error

This term refers to the error range in the results after the training process, if the target vector which expected to be produced from the neural network is $Z_{tar}[n] = [1, 1, 2, 4]$ and the actual output resulted from the neural network after the training is $Z_{act}[n] = [1, 0, 2, 5]$. Then, the error vector is calculated using the following expression:

$$E[n] = Z_{tar}[n] - Z_{act}[n] \tag{7}$$

Mean square error [20] derived using the expression (4):

$$MSE = \sum_{n=1}^{N} \frac{E_n^2}{N} \tag{8}$$

Table 3: MSE for the three proposed models.

Algorithm	MAE
FFNN	34.755
Freezing-FFNN	29.73
PSO.FFNN	17.1889

Table 3 and Fig.8 represent the MSE values for all three models. The results show that PSO-FFNN achieves the minimum MSE.

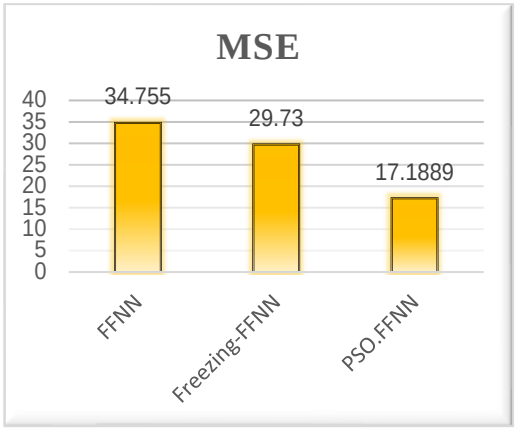


Fig. 8: MSE for proposed models.

5.3 Number of Epochs Values

Epochs represents the number of trails that neural network doses till reaching the fair performance. This is represented by a positive integer number [21]. As shown in Table 4 and presented in Fig.9; the minimum number of epochs and the minimum time are achieved by the PSO-FFNN model.

Table 4: The Epochs calculations for proposed run models.

Algorithm	Epochs
FFNN	3
Freezing-FFNN	2
PSO.FFNN	2

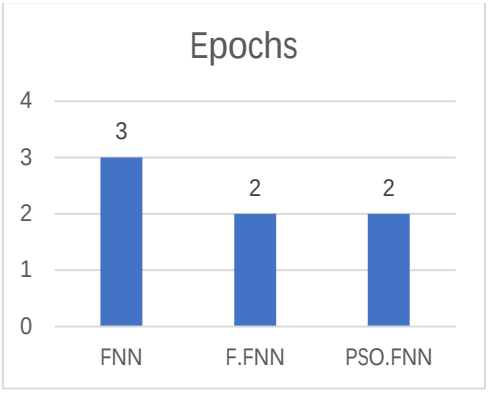


Fig.9: Epochs calculations for proposed run models.

5.4 Mean Absolute Error

This metric represents the absolute value of the error vector and the mean of them as shown in the following expression [18].

$$MAE = \sum_{n=1}^{N} \frac{|E[n]|}{N} \tag{9}$$

Table 5: MAE performance for the simulated models.

Algorithm	MAE
FFNN	2.482
Freezing-FFNN	1.41
PSO.FFNN	0.45570368

Similarly, root mean absolute error is demonstrated in Table 5 and presented in Fig.10. Results show that PSO-FFNN achieves less error rate.

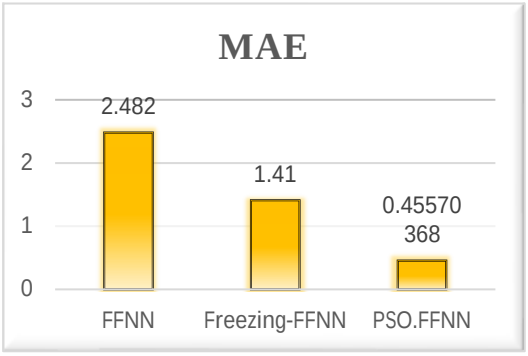


Fig.10: MAE calculations for proposed run models.

6 Conclusion

In this study, a new speaker model in text-independent voice identification system is proposed. The signal feature extraction is performed by using fundamental frequency and Mel Frequency Cepstrum Coefficients (MFCC). In addition, a smart voice recognition system is built-in using neural network algorithm which predicts the speakers by training and implementing data recognition voice features. This stage is carried out by seizing the advantages of Feed Forward Neural Network (FFNN). Furthermore; optimization is required to further improve the performance; this is accomplished by integrating the Freezing-FFNN with Practical Swarm Optimization algorithm (PSO), which significantly improves the recognition accuracy. The findings have shown that PSO-FFNN method outperforms FFNN and Freezing-FFNN in all decisive objective functions tackled in this study.

References

1. E. B. Tazi and N. El Makhfi, "An hybrid front-end for robust speaker identification under noisy conditions," in *2017 Intelligent Systems Conference (IntelliSys)*, 2017, pp. 764–768.
2. A. Akula, V. R. Apsingekar, and P. L. De Leon, "Speaker identification in room reverberation using GMM-UBM," in *2009 IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*, 2009, pp. 37–41.
3. A. Maazouzi, N. Aqili, A. Aamoud, M. Raji, and A. Hammouch, "MFCC and similarity measurements for speaker identification systems," in *2017 International Conference on Electrical and Information Technologies (ICEIT)*, 2017, pp. 1–4.
4. B. Wang, J. Zhao, X. Peng, and B. Li, "A Novel Speaker Clustering Algorithm in Speaker Recognition System," in *2006 International Conference on Machine Learning and Cybernetics*, 2006, pp. 3298–3302.
5. N. P. Jawarkar, R. S. Holambe, and T. K. Basu, "Speaker identification using whispered speech," in *2013 International Conference on Communication Systems and Network Technologies*, 2013, pp. 778–781.
6. M. F. R. Chowdhury, S.-A. Selouani, and D. O'Shaughnessy, "Distributed automatic text-independent speaker identification using GMM-UBM speaker models," in *2009 Canadian Conference on Electrical and Computer Engineering*, 2009, pp. 372–375.

7. X. Fan and J. H. L. Hansen, "Speaker identification with whispered speech based on modified LFCC parameters and feature mapping," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 4553–4556.
8. R. Martsyshyn, M. Medykovskyy, L. Sikora, Y. Miyushkovych, N. Lysa, and B. Yakymchuk, "Technology of speaker recognition of multimodal interfaces automated systems under stress," in *2013 12th International Conference on the Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*, 2013, pp. 447–448.
9. B. G. Nagaraja and H. S. Jayanna, "Efficient window for monolingual and crosslingual speaker identification using MFCC," in *2013 International Conference on Advanced Computing and Communication Systems*, 2013, pp. 1–4.
10. S. Dagtas, M. Sarimollaoglu, and K. Iqbal, "A multi-modal virtual environment with text-independent real-time speaker identification," in *IEEE Sixth International Symposium on Multimedia Software Engineering*, 2004, pp. 557–560.
11. I. N. Abu-Isbeih, K. Dagrouq, and W. Ali-Sawalmeh, "Speaker identification wavelet transform based method," in *IEEE 5th International Multi-Conference on Systems, Signals and Devices*, 2008.
12. A.-E. Maazouzi, N. Aqili, M. Raji, and A. Hammouch, "A speaker recognition system using power spectrum density and similarity measurements," in *2015 Third World Conference on Complex Systems (WCCS)*, 2015, pp. 1–5.
13. A. Meyer-Baese and V. J. Schmid, *Pattern recognition and signal analysis in medical imaging*. Elsevier, 2014.
14. Y.-H. Chao, "Speaker identification using pairwise log-likelihood ratio measures," in *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*, 2012, pp. 1248–1251.
15. Y. Shan and Q. Zhu, "Speaker identification under the changed sound environment," in *2014 International Conference on Audio, Language and Image Processing*, 2014, pp. 362–366.
16. C. Kumar, F. ur Rehman, S. Kumar, A. Mehmood, and G. Shabir, "Analysis of MFCC and BFCC in a speaker identification system," in *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 2018, pp. 1–5.
17. K. Daqrouq, W. Al-Sawalmeh, A.-R. Al-Qawasmi, and I. N. Abu-Isbeih, "Speaker identification wavelet transform based method," in *2008 5th International Multi-Conference on Systems, Signals and Devices*, 2008, pp. 1–5.
18. A. Shahab and D. Lestari, "An investigation of Indonesian speaker identification for channel dependent modeling using I-vector," in *2016 Conference of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Techniques (O-COCOSDA)*, 2016, pp. 151–155.
19. F. E. F. Junior and G. G. Yen, "Particle swarm optimization of deep neural networks architectures for image classification," *Swarm Evol. Comput.*, vol. 49, pp. 62–74, 2019.
20. M. Bazyar and R. Sudirman, "A new speaker change detection method in a speaker identification system for two-speakers segmentation," in *2014 IEEE Symposium on Computer Applications and Industrial Electronics (ISCAIE)*, 2014, pp. 141–145.
21 S. C. KOTHAR and H. Oh, "Neural Networks for Pattern," *Adv. Comput.*, p. 119, 1993.

Effects of Image Sentiment, Image Content and Image Characteristics on Social Media Post Interaction Rates

Mehmetcan Gayberi[1] and Mustafa Erşahin[2]

[1] Commencis Technology, İstanbul, Turkey, mehmetcan.gayberi@commencis.com
[2] Commencis Technology, İzmir, Turkey. mustafa.ersahin@commencis.com

Abstract. A number of variables affect the interaction rate of social media posts. This paper focuses on the effects of image-related variables. These features are image sentiment scores, object types that images contain and image characteristics such as brightness, pixel count and standard deviation of pixels. A large Instagram post dataset was used in the study, which contains 125.784 images. These images were crawled specifically for this study by using crawling tools. 17 variables were composed, and state-of-the-art methods and tools were used to form these variables. As interaction rate, the proportion of like count to follower count was used. Regression methods were used to calculate the correlation between variables and the post interaction rate. In conclusion, a 0,717 rank-order correlation was obtained using Spearman's Rank Correlation between the variables and the post interaction rate. Moreover, performing prediction of posts interaction rates by using image-related variables resulted in with 0,039 Mean Absolute Error and 0,072 RMSE (Root Mean Square Error) where interaction rate values start from 0 and mostly goes up to 1 (it can exceed 1 as a small number of posts can be interacted more than the number of followers).

Keywords: Social Media, Post Interaction, Post Popularity, Image Sentiment, Image Characteristics.

1 Introduction

Social media usage increases dramatically over the years. People share images, videos, blogs and many more on social media platforms. There are billions of contents shared every day.

These huge numbers push many academic and commercial studies. In the literature, there are a number of studies focusing on social media platforms and big data they contain. There exist a huge number of studies focusing on interactions, sentiment analysis, popularity.

This paper combines various methods about image and presents the effect of these outputs to the interaction rate. In short, image sentiment, image contents and basic image characteristics were utilized in the study and their effect on interaction rate was investigated.

For the content of this study, Instagram was selected as the social media platform due to its popularity and only image-based posts were focused. Various state-of-the-art

methods and algorithms were used to obtain the relation between variables and interaction rates. All the experiments were performed on a quite big dataset containing 125.784 images from publicly available Instagram posts.

2 Related Works

Lin et al. [2] presented a new dataset for Image Object Detection in 2014 in their study "Microsoft coco: Common objects in context". They created a model that can detect 91 different objects in images. Moreover, Caesar et al. in 2018 [3] presented the object and category mappings in their study. They created 12 different categories for objects and these 12 categories were used in our study as variables.

Stokowiec et al. aimed to predict the popularity of online content using only titles in their study in 2017 [6]. They used a dataset of 40.000 videos and news. The interesting point was using only the title variable. In our study, we followed a similar "simple" approach and used only image-related variables by using some part of Gayberi and Oguducu's dataset from their popularity prediction study [1].

Vadicamo et al. [8], Yuan et al. [7], You et al. [5] and Campos et al. [4] all studied image sentiment in their studies. Yuan et al. [7] focused on mid-level image attributes rather than low-level image attributes and they resulted in accuracy up to 82% in 2013. You et al. [5] in 2015 studied image sentiment by using Convolutional Neural Networks. They proposed that well trained neural networks perform better than low-level or mid-level attribute sentiment classifiers. Also using CNN's helped them to transfer knowledge. Vadicamo et al. [8] proposed a method for image sentiment analysis. They used a database of 3 million tweets from Twitter. These tweets contain both images and text. They tried to train their image sentiment classifiers using textual contents. In 2018, Campos et al. [4] struggled to get better results and they performed fine-tuning of CNNs. They used the dataset from You et al.'s [5] study and they could improve the performance. In our study, Campos et al.'s sentiment prediction model was used.

Gayberi and Oguducu [1], Khosla et al. [9], Mazloom et al. [10] and many other researchers studied popularity prediction. They all used large datasets and a number of features to predict popularity. These features were user-based, post based, textual based, statistical, image-related, image characteristics, low-level vision features and image content.

Fontanini et al. [12] studied on popularity prediction of videos and they used sentiment and content visual features. They outperformed the state-of-the-art video popularity prediction methods. They used the number of views as the popularity metric while we used the proportion of likes to the number of followers in our study as the interaction rate.

In another study, Gelli et al. [11] performed popularity prediction by using user features, object features, sentiment and context features. Their sentiment features were based on visual emotion and they used Visual Sentiment Ontology and DeepSentiBank. They result in with 0.72 to 0.74 correlation by using and combining all the features not only image-related features.

3 Dataset

3.1 Base Dataset

For building the dataset, Gayberi and Oguducu's dataset [1] was taken as the base. The base dataset contains 210.630 Instagram posts which were collected from 22.359 different Instagram accounts.

These Instagram accounts were selected randomly based on some pre-defined seed accounts [1]. This dataset contains various features about the user, post, image, as well as some statistical features. Only some image-related features and the number of followers were used in this study and all other features are omitted. The details are given in the next sections.

3.2 Image Collection

First, Instagram post shortcodes were taken from the base dataset and images related to these posts were collected. The base dataset was created in March and April 2019. For this study, images were collected in April 2020. Some of the posts were not accessible or not collected for this study in April 2020 due to some reasons:

- Account does not exist anymore
- Account became private
- Related post does not exist anymore
- Videos were not used in this study

125.784 images were still valid and accessible and collected specifically for this study.

3.3 Image Content Features

By using methods that were presented by Lin et al. [2] in 2014, in their study "Microsoft COCO: Common Objects in Context", Image Object Detection task performed on collected images. Caesar et al. [3] in 2018 shared objects and category mappings for the COCO dataset. The shared 12 categories were used, and image content-related features were created. These features are based on existence in the image of categories. They are given in Table 1 with some statistical information.

Table 1. Image Content Features.

Category	Number of Images Containing	Mean	Standard Deviation
Human	62.559	0,49	0,49
Animal	8.534	0,06	0,25
Vehicle	14.217	0,11	0,31
Indoor Things	13.246	0,10	0,30
Outdoor Things	7.689	0,06	0,23

Accessories	13.504	0,10	0,30
Sports	8.456	0,06	0,25
Kitchen	14.336	0,11	0,31
Food	10.696	0,08	0,27
Furniture	13.499	0,10	0,30
Electronics	8.605	0,06	0,25
Appliances	1.638	0,01	0,11

3.4 Image Sentiment Scores

Campos, Jou and Giro-i-Nieto presented their study on Image Sentiment Analysis in 2017 [4]. They trained machine learning models by using Convolutional Neural Networks (CNNs). They built their deep network over Caffe. They used Twitter images to train their models. The Twitter dataset was published by You et al. within their study "Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks". It contains 1.269 images from Twitter, and they are publicly available [5].

We used the model that Campos et al. built in their study "From Pixels to Sentiment: Fine-tuning CNNs for Visual Sentiment Prediction". Their source code is publicly available on GitHub [4]. The model was used, and sentiment scores were calculated for all 125.784 images in our dataset.

Sentiment score was calculated as a scalar value between 0 and 100, where 100 is completely positive and 0 is the most negative. Some statistical information on sentiment scores are given in Table 2 and the distribution is given in Figure 1.

Table 2. Image Sentiment Calculations.

Average	Standard Deviation	25%	50%	75%	100%
82,35	27,61	77	97	100	100

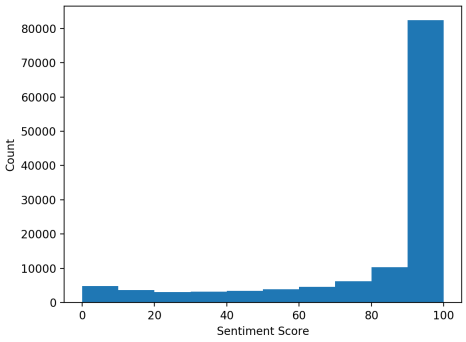


Fig. 1. Distribution: Image sentiment score.

3.5 Image Characteristics

By using an open-source image analysis tool, we obtained some basic characteristics of images. These variables are brightness, pixel count and standard deviation of all bands in the image. Table 3 gives the statistical explanations of these variables.

Table 3. Image Characteristics Statistics.

	Average	Std Dev.	Min	Max	25%	50%	75%
Brightness	140,36	38,37	1	255	117	139	163
Pixel Count	1.037.714	338.076,7	53.440	1.466.640	834.840	1.166.400	1.166.400
Standard Deviation	66,50	16,00	0	126,09	57,05	67,31	76,97

Figure 2, 3 and 4 show the distribution of brightness, pixel count and standard deviation respectively.

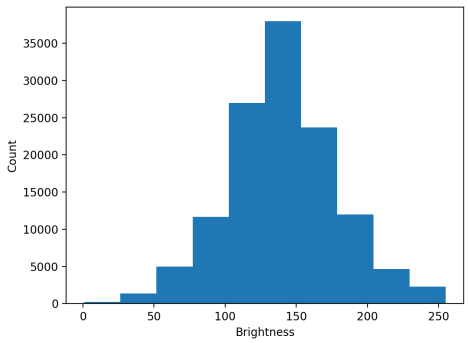


Fig. 2. Distribution: Brightness of images.

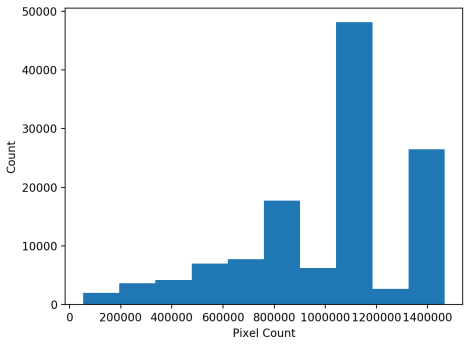


Fig. 3. Distribution: Pixel count in images.

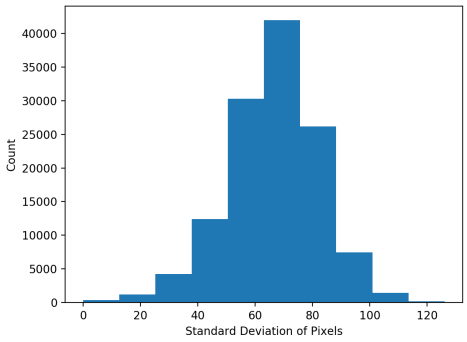


Fig. 4. Distribution: Standard deviation of pixels in images.

4 Experiments and Results

4.1 Image Interaction Rate

In various studies, the proportion of likes (views, clicks, retweets etc.) to the number of followers is used as an interaction rate or popularity metric. In this study, we also used it as an interaction rate which shows the interaction of users to related posts.

Figure 5 shows the interaction rate distribution over post instances. As there are some posts that have more likes than the number of followers, there are some outliers. The second chart in Figure 5 depicts the interaction rates after filtering by up to 1.

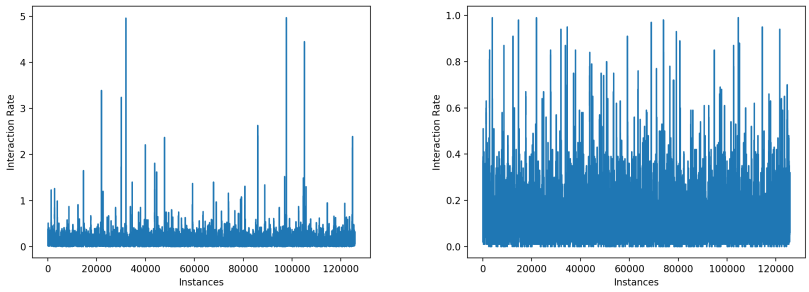


Fig. 5. Distribution: Interaction rates unfiltered (left) and filtered up to 1 (right).

4.2 Results

After preparing data and variables, we had 12 image content related variables, sentiment score variable and 3 image characteristics variables (brightness, pixel count and standard deviation of pixels) for 125.784 images. We had calculated the interaction rate as the proportion of likes to the number of followers.

We used variables and interaction rates and performed Random Forest Regression to calculate correlation, predictability and errors based on predictability. Table 3 shows the results.

As shown in Table 4, variables present a high rank-order correlation of 0,717. Also, a low MAE was obtained which means the interaction rate for an image post can be predicted with a low error rate.

Table 4. Correlation Results.

Metric	Value
MAE (Mean Absolute Error)	0,039
MSE (Mean Squared Error)	0,005
RMSE (Root Mean Square Error)	0,072
Spearman's rank-order correlation	0,717

Importance values related to the variables are given in Table 5. In general, image characteristics and image sentiment scores are far more effective on the interaction rate. The number of followers has the highest importance which tells the interaction rate is highly correlated with the number of people following the account.

Table 5. Variable Importance Values.

Metric	Value
followers	0,395
stddevpixel	0,169
brightness	0,128
pixelcount	0,119
sentimentscore	0,083
image_has_human	0,018
image_has_accessory	0,010
image_has_indoor	0,010
image_has_vehicle	0,009
image_has_kitchen	0,009
image_has_electronic	0,008
image_has_furniture	0,007
image_has_food	0,007
image_has_sports	0,006
image_has_outdoor	0,006
image_has_animal	0,006
image_has_appliances	0,001

5 Conclusion and Future Works

In this paper, we proposed to obtain the relationship between image-related variables and the interaction rate of this image on social media. Image Object Detection methods were used to understand the content of images and variables were derived according to the outputs of detection. Image sentiment scores were calculated by using state-of-the-art image sentiment models and basic image characteristics were extracted. By using regression methods, the effects of these variables were investigated and a high rank-order correlation of 0,717 was obtained. Moreover, it was revealed that the interaction rate could be predicted with an MAE value of 0,039. Also, another outcome showed that image characteristics and image sentiment are more important variables than image content variables.

For future works, image characteristic variables can be enriched, and more advanced variables can be added. Some other indicators may be added for calculation of interaction rates, such as the number of comments or various other outcomes for different social media platforms. Obviously, a bigger dataset can give better and sharper results.

References

1. Gayberi, M., & Oguducu, S. G. (2019, November). Popularity Prediction of Posts in Social Networks Based on User, Post and Image Features. In Proceedings of the 11th International Conference on Management of Digital EcoSystems (pp. 9-15).
2. Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham.
3. Caesar, H., Uijlings, J., & Ferrari, V. (2018). Coco-stuff: Thing and stuff classes in context. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1209-1218).
4. Campos, V., Jou, B., & Giro-i-Nieto, X. (2017). From pixels to sentiment: Fine-tuning CNNs for visual sentiment prediction. Image and Vision Computing, 65, 15-22.
5. You, Q., Luo, J., Jin, H., & Yang, J. (2015, February). Robust image sentiment analysis using progressively trained and domain transferred deep networks. In Twenty-ninth AAAI conference on artificial intelligence.
6. Stokowiec, W., Trzciński, T., Wołk, K., Marasek, K., & Rokita, P. (2017, June). Shallow reading with deep learning: Predicting popularity of online content using only its title. In International Symposium on Methodologies for Intelligent Systems (pp. 136-145). Springer, Cham.
7. Yuan, J., Mcdonough, S., You, Q., & Luo, J. (2013, August). Sentribute: image sentiment analysis from a mid-level perspective. In Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining (pp. 1-8).
8. Vadicamo, L., Carrara, F., Cimino, A., Cresci, S., Dell'Orletta, F., Falchi, F., & Tesconi, M. (2017). Cross-media learning for image sentiment analysis in the wild. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 308-317).
9. Khosla, A., Das Sarma, A., & Hamid, R. (2014, April). What makes an image popular?. In Proceedings of the 23rd international conference on World wide web (pp. 867-876).
10. Mazloom, M., Rietveld, R., Rudinac, S., Worring, M., & Van Dolen, W. (2016, October). Multimodal popularity prediction of brand-related social media posts. In Proceedings of the 24th ACM international conference on Multimedia (pp. 197-201).
11. Gelli, F., Uricchio, T., Bertini, M., Del Bimbo, A., & Chang, S. F. (2015, October). Image popularity prediction in social media using sentiment and context features. In Proceedings of the 23rd ACM international conference on Multimedia (pp. 907-910).
12. Fontanini, G., Bertini, M., & Del Bimbo, A. (2016, June). Web video popularity prediction using sentiment and content visual features. In Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval (pp. 289-292).

Software Reliability Assessment Using Polynomial Regression Approach

Shayma Mustafa Mohi-Aldeen[1], Makera Moayad Aziz[2], and Dena Rafaa Ahmed[3]

[1,2] Computer Science Department College of Computer Science and Mathematics,
Mosul University, Mosul, Iraq
[3] Software Department College of Computer Science and Mathematics,
Mosul University, Mosul, Iraq

Abstract. Software Reliability is a method of measuring the quality of software that predicts the number of faults in the software. Different models were proposed from time to time to estimate the number of software failures. This paper proposed a new model depend on polynomial regression to predict the number of software faults for a particular duration of time in a specific environment. The mean square error (MSE) was used for measuring the performance of the proposed model and the results show that the proposed model outperforms the other models in predicting the faults of the systems when compared with previous work, the results also show that the model has the best performance when calculating the MSE.

Keywords: Software Reliability, Polynomial Regression, Mean Square Error (MSE), Least Square Method(LSE).

1 Introduction

Nowadays, software being used dramatically in different areas of life and has been increasingly important and a critical part of applications. Due to the high importance of software quality assurance, thus the software requirements like reliability, security, performance, and accuracy became a significant issue. However, software reliability is the most important quality attribute of software [1, 2]. Software Reliability quantifies the faults and failure of the software, that lead to dire effects on safety-critical systems and normal systems as well [3]. So, predicting software reliability and estimating it became widely required in the projects to accomplish high-reliability systems. It is evolving as the main part of the study [4]. Software Reliability is described as the likelihood of software free of failure in a specific duration of time under particularized environmental [2, 5]. Different models of software reliability were suggested to measure software reliability. Since 1972 when Jelinski and Moranda [6] proposed the first software reliability growth model (SRGM) and many reliability models have been increasingly developed until now and are a great amount of researchers' interest and utilized in different fields. The models of reliability depend on detection the faults through a particular testing duration time [7]. There are two types of SRGMs; the parametric and non-parametric models. The traditional reliability growth models which have been used are based on various parameters, and these parameters representing the properties of each model. In general, these parameters have a great effect on the models' accuracy [8] and

are predicted using Mathematical approaches such as Maximum Likelihood Estimation (MLE) or Least Square Estimation (LSE), etc. However, these approaches involved some limitations in calculating or estimating [9, 10] and the major restrictions of these approaches are the parameters' number. The limitations of the traditional methods of parameter estimation motivate the researchers to use metaheuristic methods such as Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), etc. [5]. This paper proposed a new technique for assessing software reliability, which depends on a statistic approach called the polynomial regression model, this model will explain later. The rest of the paper is ordered as follows: In Section 2, the literature review was explained, which includes the software reliability and the polynomial regression model followed by the proposed method in Section 3, while Section 4 describes the results and discussion and the conclusion in Section 5.

2 Literature Review

2.1 Software Reliability Models

Software reliability is a quantitative scale that is used to measure the failure chance of the software through the duration of time and a particular environment [1, 9]. Due to the importance of software reliability, many of the software reliability models have been developed. These SRGMs are used to represent the detection of the error process as a discrete or continuous process at the rate of error detection that depends on the time [8]. The detected faults of the traditional SRGMs will be removed immediately, which is a general assumption. Though, it is rare to immediately correct the discovered faults which imply that this assumption is not very realistic [11]. The goal of SRGMs is to specify the random process which represents the behavior of the software respecting to failures of software. This process can be utilized for the reliability estimation, present status measurement, and the prediction of future status [2, 6]. In 1972, Jelinski and Moranda developed the first SRGM which has been received great consideration, and then various models have been suggested to measure the software reliability [3]. These models could be categorized as the time between failure models and failure calculation models [4]. Various mathematical methods have been used in predicting software reliability such as MLE and LSE, these models have different limitations as explained earlier [5, 9]. The increase in parameters in any mathematical model will increase the faults of the data. In the beginning, the interest of researchers was focused on mathematical models. In 1992, Knafl proposed models for software reliability depend on the maximum probability method of two factors, and in 1996, Knafl et. al suggested an MLE to SRGMs. However, the constraints of these models lead the researchers to use other models for solving the problem of parameter estimation [5, 7].

2.2 Polynomial Regression Model

Regression analysis is considered as one of the predictive modeling techniques that are used to investigate the relationship between a dependent and independent variable. The relationship between the variables can be used to make predictions. There are many

types of regression analysis, and the polynomial regression is one of them, which is defined as a regression analysis form where the relation between the independent variables (x) and the dependent variable (y) is considered as an nth degree polynomial in x [12]. Polynomial regression is a statistical model that gives the most accurate estimation of the relationship between the dependent and independent variables and it's suitable for a broad range of functions [13]. Different researches have been used polynomial regression, especially in statistical estimation problems in engineering and manufacturing applications [14]. In this paper, it has been used in predicting the reliability of software as will be explained in the next section.

3 Proposed Method

In reliability, the faults' number was increased during the time and there is a positive relationship between the number of faults and the time [3]. The polynomial regression is described as the relationship between the independent variable x and the dependent variable y. The number of faults grows during the time that means the faults are dependent on time. The number of faults will be the dependent variable and the time will be the independent variable. Polynomial regression is suitable for a nonlinear relationship between the value of x and the identical conditional mean of y [12]. The polynomial regression is represented by the following model [13]:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_n x^n + \varepsilon \quad (1)$$

Where,
y: represents the estimated faults number,
x: represents the time at which the number of faults will be estimated,
β: represents the coefficient of variable x, and
ε: represents the intercept (constant)

The following steps represent the proposed algorithm:

Begin
Read data_set
Calculate the cumulative faults
Var1 Independent_variable ⟵ time
Var2 dependent_variable ⟵ No_faults
Calculate the polynomial regression model depends on the equation (1)
End

The dataset (DS1, DS2, and DS3) which have been used in this paper are collected from previous studies as shown in Table 1, these data included different working time per weeks for different systems with different numbers of faults. Table 1 included the information and details of the dataset. These datasets were used in the previous studies from time to time for testing the reliability models of the software.

Table 1. Description of the datasets.

Dataset number	Working Time/weeks	Number of faults	Resource
DS1	19	328	[15]
DS2	20	100	[16]
DS3	34	181	[17]

The Mean Square Error (MSE) also called the variance of estimation and the square root of it is known as the standard error of estimation. It is used to measure the proposed model performance and the small value of MSE indicates better performance [13, 18]. It calculates the total deviation between the real value of the faults for each system and the estimated value by using the following equation [13]:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(R_i - E_i)^2 \quad (2)$$

Where,
N is the number of total observations
Ri is the real value of the faults
Ei is the estimated value of faults

3.1 Estimate and solve the coefficient of the regression model

The Least Square Method (LSM) is one of the common methods used to estimate the coefficient of the polynomial regression. The goal of this method is to minimize the deviation between the real values and the estimated value of the dataset [18]. The following system of linear equations can be used to determine the coefficient of the polynomial regression model [13].

$$\begin{bmatrix} N & \sum_{i=1}^{N} x_i & \cdots & \sum_{i=1}^{N} x_i^k \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 & \cdots & \sum_{i=1}^{N} x_i^{k+1} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{N} x_i^k & \sum_{i=1}^{N} x_i^{k+1} & \cdots & \sum_{i=1}^{N} x_i^{2k} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i y_i \\ \vdots \\ \sum_{i=1}^{N} x_i^k y_i \end{bmatrix}$$

Where,
a: polynomial coefficient
k: The polynomial degree
N: The points' number to be the degree

Many methods can be used to solve the coefficient of the regression one of them is Cramer's rule and determinant which is used in this study. Equation (3) represents the

determinant (Matlab 2015b is used to find the determinant values for each matrix) which can be used to determine each coefficient [13].

$$a_k \frac{\det(Mi)}{\det(M)} \tag{3}$$

Where, Mi represents the matrix M with the ith column substituted by vector b column (the result of the least square method). For example, M_0 could be calculated as follows [13]:

$$M_0 = \begin{bmatrix} \sum_{i=1}^{N} y_i & \sum_{i=1}^{N} x_i & \cdots & \sum_{i=1}^{N} x_i^k \\ \sum_{i=1}^{N} x_i y_i & \sum_{i=1}^{N} x_i^2 & \cdots & \sum_{i=1}^{N} x_i^{k+1} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{N} x_i^k y_i & \sum_{i=1}^{N} x_i^{k+1} & \cdots & \sum_{i=1}^{N} x_i^{2k} \end{bmatrix}$$

4 Results and Discussion

In this paper, the polynomial regression of degree three (cubic regression) has been used to estimate the faults of the system. The results of the proposed model have been compared with the results of a paper selected from literature studies [19]; the paper used two reliability models: The G_O model and the INFS model which has been included in Table 2.

Table 2. The reliability growth models U.

Model	Mean Value Function	No. of parameters	Resources
G_O	$\mu(t) = a(1 - e^{-bt})$	2	[20]
INFS (Inflection S-Shaped model)	$\mu(t) = \frac{a(1 - e^{-bt})}{1 + ce^{-bt}}$	3	[21]

Where t represents the time at which the fault will be estimated.

In [19] The researchers tried to develop the model by estimating the parameters of each model to get estimated values that close to the real values. The Algorithm used in this research gave a better result compared with the previous studies. The proposed model gave better results than results that gave by the algorithm suggested in [19] as

shown in Table 3. For each of the three datasets, a polynomial regression model that has been designed and executed in this paper gives the minimum MSE when compared with G_O and INFS models as presented in Table 3.

Table 3. The results of comparison with other models.

Da-tasets	Proposed Model (Polynomial Model)	MSE of each model		
		G_O Model [19]	***INFS Model [19]***	***Proposed Model***
DS1	$-0.054x^3+1.17x^2+13.98x+10.3$	139.815	82.704	63.79
DS2	$-0.015x^3+0.026x^2+5.2x+10.56$	11.617	8.98	1.79
DS3	$-0.00315x^3+0.141x^2+4.2x+0.25$	22.863	5.82	5.69

The results in Table 3 show that the polynomial regression model has the least value of MSE. For DS1, the MSE of the proposed model is equal to 63.79 which is less than the MSE of the G_O model and INFS model that is equal to 139.815 and 82.704 respectively. While in DS2, the MSE of the proposed method is equal to 1.79 whereas in the G_O model and INFS model equal to 11.617 and 8.98 respectively. The MSE of DS3 of the proposed method is equal to 5.69 and in the G_O model and INFS model equal to 22.863 and 5.82 respectively. This means that the proposed model has better performance of prediction when compared with previous models for all the datasets and all models. The results also show that the number of faults per week predicted by the proposed method is too close to the real number of faults and for all systems as shown in Fig. 1.

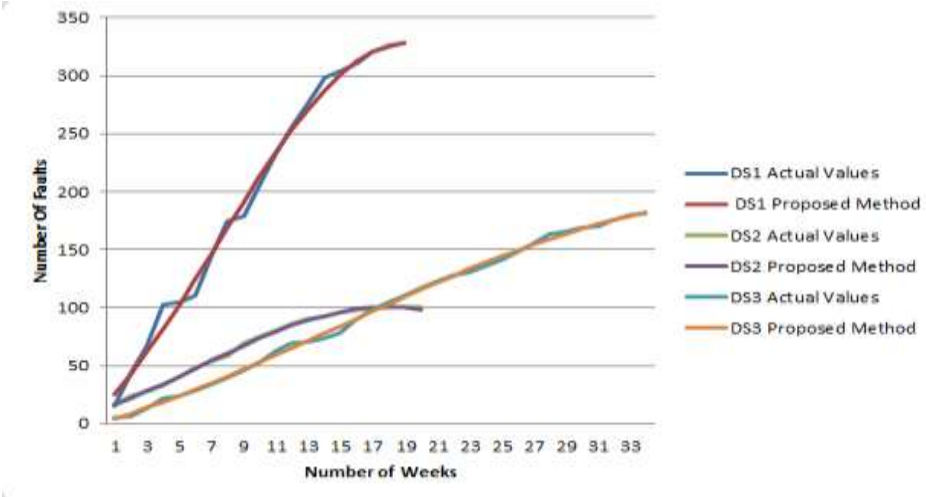


Fig. 1 Comparison between the actual values of the number of faults with the proposed model.

Figure 1 shows that the number of faults predicted by the proposed model is very near to the real values of many faults that presented in Table 1 and for all the datasets. This means that the proposed model has the best performance in predicting the number of faults of the system.

5 Conclusion

Reliability is one of the software quality attributes, and it is the most significant factor because it calculates the faults and failures of the software, which has significant effects on the systems. The SRGMs have been used widely in estimation and prediction software reliability. Nevertheless, these models have many limitations. Many researchers were proposed different solutions to outperform these limitations using different approaches. This paper proposed a new predictive approach depend on the polynomial regression model, the performance of the proposed approaches was evaluated by using MSE, and the results compared with the literature studies. The results of a comparison present that the proposed approach outperformed the other models when calculating the MSE for all the datasets and all models, for example, in DS1 the MSE is equal to 139.815 and 82.704 in G_O and INFS models respectively, while the proposed method reduced the MSE to 63.79. The proposed method also reduced the MSE in DS2 and DS3.

References

1. M. Palviainen, M., A. Evesti, E. Ovaska, “The Reliability Estimation Prediction and Measuring of Component-based Software,” Journal of Systems and Software, Vol. 84 , No. 6, 2011, pp. 1054–1070.

2. M. R. Lyu, “Handbook of Software Reliability Engineering”, McGraw-Hill, New York, 1996.

3. M. R. Lyu, “Software Reliability Engineering: a Roadmap”, Future of Software Engineering (FOSE’07), 2007, pp. 153–170.

4. A. Amin, L. Grunske, A. Colman, “An approach to software reliability prediction based on time series modeling”, The Journal of Systems and Software, Vol. 86, No. 2013, 2007, pp. 1923– 1932.

5. A. Kumar, R. P. Tripathi, P. Saraswat, P. Gupta, “Parameter Estimation of Software Reliability Growth Models Using Hybrid Genetic Algorithm”, Fourth International Conference on Image Information Processing (ICIIP), 2017.

6. Z. Jelinski, P. Moranda, “Software reliability research. In: Statistical Computer Performance Evaluation”. Academic Press Inc, New York, USA, 1972, pp. 465–484.

7. A. Kumar, "Software Reliability Growth Models, Tools and Data Sets-A Review", Proceedings of the 9th India Software Engineering Conference. ACM, 2016.

8. P. J. Ganesh, "A survey of software reliability models", Pham, Hoang, "Springer Series in Reliability Engineering." New Jersey, USA, 2005.

9. N. Raj Kiran, V. Ravi, “Software reliability prediction by soft computing techniques”, Journal of Systems and Software, Vol. 81, No. 4, 2008, pp. 576–583.

10. M. Moura, E. Zio, I. Didier Lins, E. Droguett, “Failure and reliability prediction by support vector machines regression of time series data”, Reliability Engineering and System Safety, Vol 96, No. 11, 2011, pp. 1527–1534.

11. J. H. Lo, C. Y. Huang, “An integration of fault detection and correction processes in software reliability analysis”, The Journal of Systems and Software, Vol. 79, No. 2006, 2011, pp. 1312–1323.

12. C. Douglas Montgomery, A. Peck Elizabeth, G. Geoffrey Vining, “Introduction to Linear Regression Analysis” fifth ed., 2012.

13. D. Jonathon Brown, “Linear Models in Matrix Form: A Hands-On Approach for the Behavioral Sciences”, Springer International Publishing, 2014.

14. I. O. Ajao, A. Abdullahi, and I Raji, “Polynomial Regression Model of Making Cost Prediction In Mixed Cost Analysis”, Mathematical Theory and Modeling, Vol.2, No.2, 2012, pp. 14-24.

15. M. Ohba, "software reliability analysis models", IBM J. RES. DEVELOP, Vol. 28 No. 4, 1984, pp.228-443.

16. A. Wood, "Predicting Software Reliability," IEEE Computer, Vol. 29, No. 11, 1996, pp. 69-77.

17. D. R. Jeske, X. Zhang, L. Pham, " Adjusting Software Failure Rates That Are Estimated From Test Data", IEEE TRANSACTIONS ON RELIABILITY, Vol. 54, No. 1, 2005, pp.107–114.

18. A. Dean, D. Voss, D.Draguljić, “Design and Analysis of Experiments”, 2nd Edition, Springer, 2017.

19. J.S. Alneamy, M.M.A.Dabdoob, “The Use of Original and Hybrid Flower Pollination Algorithm In Estimating The Parameters of Software Reliability Growth Models”, Educational and Science Journal, Vol. 28, No. 2, 2019, pp. 196-218.

20. P. H. Meyfroyt, "Parameter Estimation for Software Reliability Models", M.Sc. thesis, Eindhoven: Technische Universiteit Eindhoven, 2012.

21. R. S. Prasad, K. P. Rao, G. K. Mohan, "Software Reliability using SPRT: Inflection S-shaped Model", International Journal of Application or Innovation in Engineering & Management, Vol. 2, No. 6, 2013, pp. 349-355.

Realization of Electronically Tunable Square-Root-Domain Trans-admittance Filter

M. Serhat Keserlioglu[1] and Ali Kircay[2]

[1] Electrical and Electronics Engineering Department, Pamukkale University, Denizli, Turkey
[2] Electrical and Electronics Engineering Department, Harran University, Sanliurfa, Tukey
mskeserlioglu@pau.edu.tr, kircay@harran.edu.tr

Abstract. Square Root Domain (SRD) electronically adjustable trans-admittance lowpass filter proposes in this paper. The filter circuit input is voltage and output responses of filter is current. The design is based on the state-space synthesis method. State-space synthesis method is one of the methods used to systematically design filters. State variables of the filter are obtained by state space synthesis method. The circuit of the filter is realized from the obtained equations. One of the important types of filters using the state-space synthesis method are the square-root and log-domain filters where the sign is compressed and expanded. The transfer admittance parameter g_0, the quality factor Q and the natural frequency f_0 of the transadmittance filter can be electronically controlled by external DC currents. Using PSPICE simulations program are verified the theory and to show the performance of filter circuits' responses. The proposed trans-admittance filter is simulated by using TSMC 0.35 μm Level 3 CMOS process parameters.

Keywords: Square-root-domain Filters, State-Space Synthesis, Companding systems, Electronically Tunable.

1 Introduction

The translinear principle is based on the exponential (current-voltage) I-V characteristics of bipolar junction transistors (BJTs) and metal-oxide-semiconductors (MOS) transistors in weak inversion region [1, 2]. MOS quadratic law is the linear trans-conductor that derived by Bult [3]. The MOS translinear (MTL) principle is proposed by Seevinck [4] using the bipolar translinear (BTL) approaches [1]. MOS quadratic law in strong inversion and saturation region and the voltage translinear principle were studied in [5-10].

The square-root-domain filter is important subclass of companding filters. This filter class were studied by different researchers, because the companding filters have many advantages, for example: electronic tunability, high frequency applications, and large dynamic range under low voltage/power supply [11, 12]. The principle of this filters; the signals are compressed at the first (input) stages, then processed and then expanded at the last (output) stages.

A number of square-root-domain circuits were proposed by the authors in the literature: first-order filters [13-16], second-order voltage-mode (VM) [17, 18] and current-mode (CM) filters [7, 13, 18, 19], third-order VM filters [20] and trans-admittance circuits [10, 14, 21] Additionally, there are some papers in the literature about SRD and the squarer/divider structures [7, 10, 12, 22]. These sub-circuit blocks are used to SRD circuits [9, 15, 22].

The input signal of a trans-admittance type filter is voltage and output signal is current. Therefore, a trans-admittance type filter described as an interface connecting a voltage-mode circuit to a current-mode circuit [25, 26, 27]. As a result, both filtering and voltage-current transforming processes can be achieved by trans-admittance type filter. Trans-admittance and trans-impedance type filter circuits have various application areas [28, 29, 30]. For example, trans-admittance type filters are used in the base-band structure of modern radio systems [25, 26]. However, few works have been proposed in the area of square-root-domain trans-admittance filter design by using steady-space method [23, 24].

In this work, square-root-domain, second-order trans-admittance lowpass filter is proposed by using steady-space synthesis method. The trans-admittance lowpass filter circuit contains of square-root and squarer/divider blocks. In addition these circuit blocks, current mirrors, current sources, power supply and two grounded capacitors are included in the filter circuit. The trans-admittance g_0, cut-off frequency f_0, quality factor Q of the proposed filter can be tuned by changing current values.
does not have an indent, either.

2 The Design of SRD Second Order Trans-Admittance Type LP Filter

Second order trans-admittance type low-pass filter transfer function can be written as follows,

$$Y_{21}(s) = \frac{I_{out}}{V_{in}} = \frac{g_0 \omega_0^2}{s^2 + \frac{\omega_0}{Q} s + \omega_0^2} \tag{1}$$

where V_{in} is the input voltage, I_{out} is the output current, g_0 is the trans-admittance parameter, ω_0 is the cut-off frequency and Q is the quality factor of the trans-admittance circuit. Transfer function of trans-admittance filter was turned into the following state-space equations:

$$\dot{I}_1 = -\frac{\omega_0}{Q} I_1 + \omega_0 I_2 \tag{2}$$

$$\dot{I}_2 = -\omega_0 I_1 + g_0 \omega_0 U \tag{3}$$

where, U is represented by V_{in} and I_1 and I_2 currents are the state variables. The output variable is defined by (4) as given below:

$$I_{out} = I_1 \tag{4}$$

The I_1 and I_2 currents that were represented by state-variable in (2) and (3) are drain currents of MOS transistors in saturation mode and these currents can be given as follows [18, 19].

$$I_1 = \frac{\beta}{2}(V_1 - V_{th})^2 \tag{5}$$

$$I_2 = \frac{\beta}{2}(V_2 - V_{th})^2 \tag{6}$$

The input voltage U that was represented by V_{in} in (3) is the gate to source voltage of a MOS and this voltage can be defined as given (7),

$$I_U = \frac{\beta}{2}(U - V_{th})^2 \tag{7}$$

where, $\beta = \mu_0 C_{ox}(W/L)$ is the device trans-conductance parameter, and V_{th} is the threshold voltage respectively. The input voltage U and the derivatives of I_1 and I_2 are written as:

$$U = \sqrt{2I_U/\beta} + V_{th} \tag{8}$$

$$\dot{I}_1 = \dot{V}_1\sqrt{2\beta I_1} \tag{9}$$

$$\dot{I}_2 = \dot{V}_2\sqrt{2\beta I_2} \tag{10}$$

(8) and (9) Eq.s are applied to (2) and (3) then given by

$$\dot{V}_1\sqrt{2\beta I_1} = -\frac{\omega_0}{Q}I_1 + \omega_0 I_2 \tag{11}$$

$$\dot{V}_2\sqrt{2\beta I_2} = -\omega_0 I_1 + g_0\omega_0\left[\sqrt{2I_U/\beta} + V_{th}\right] \tag{12}$$

The both side of (11) and (12) can be multiplied by coefficient C_1 and C_2 and these equations can be rearranged as follows:

$$C_1\dot{V}_1 = -\frac{C_1\omega_0 I_1}{Q\sqrt{2}\sqrt{\beta}\sqrt{I_1}} + \frac{C_1\omega_0 I_2}{\sqrt{2}\sqrt{\beta}\sqrt{I_1}} \tag{13}$$

$$C_2\dot{V}_2 = -\frac{C_2\omega_0 I_1}{\sqrt{2}\sqrt{\beta}\sqrt{I_2}} + \frac{C_2\omega_0}{\sqrt{2}\sqrt{\beta}\sqrt{I_2}}\left[\frac{g_0}{\sqrt{\beta}}\sqrt{2I_U} + g_0 V_{th}\right] \tag{14}$$

I_{g0}, I_{T0} and I_k currents can be written as [26].

$$\sqrt{I_{g0}} = g_0/\sqrt{\beta} \tag{15}$$

$$I_{T0} = g_0 V_{th} \tag{16}$$

$$I_k = 2\sqrt{\frac{I_U I_{g0}}{2}} + I_{T0} \tag{17}$$

These current equations are used in (13) and (14) and they can be revised to form the following nodal equations:

$$C_1\dot{V}_1 = -\frac{1}{Q}\sqrt{\frac{I_{01}I_1}{2}} + \sqrt{\frac{I_{01}I_2^2}{2I_1}} \tag{18}$$

$$C_2\dot{V}_2 = -\sqrt{\frac{I_{02}I_1^2}{2I_2}} + \sqrt{\frac{I_{02}I_k^2}{2I_2}} \tag{19}$$

where I_{01} and I_{02} currents can be defined as given (20) and (21).

$$\sqrt{I_{01}} = \frac{C_1\omega_0}{\sqrt{\beta}} \tag{20}$$

$$\sqrt{I_{02}} = \frac{C_2\omega_0}{\sqrt{\beta}} \tag{21}$$

The proposed square-root-domain trans-admittance filter circuit with lowpass output can be achieved via (18), (19), and (17) as shown in Fig. 1.

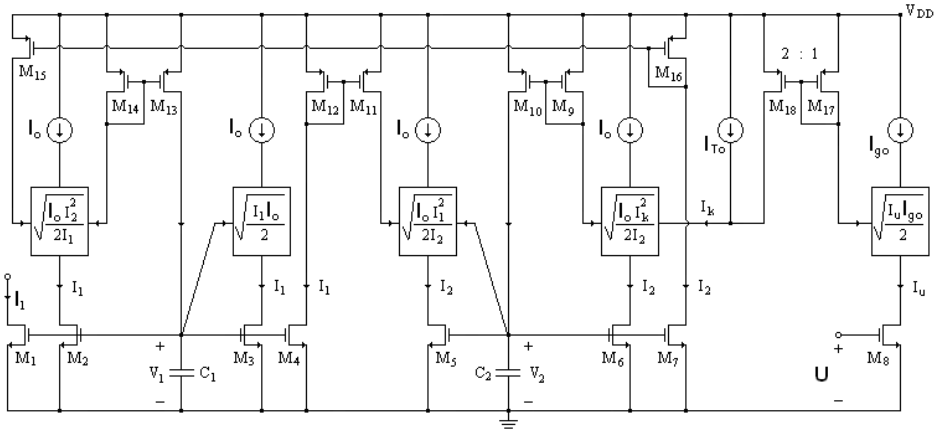


Fig.1. SRD trans-admittance type LP filter.

The trans-admittance parameter of the filter can be written as given (22) by using (15) and (16) [23].

$$g_0 = \beta V_{th}\left(\frac{I_{g0}}{I_{T0}}\right) \tag{22}$$

The cut-off frequency of the filter is given by (23) as follow [9, 18]:

$$\omega_0 = \frac{\sqrt{\beta I_0}}{C} \tag{23}$$

where $I_{01} = I_{02} = I_0$ and $C_1 = C_2 = C$.

3 Simulation Results

The proposed second order square-root-domain trans-admittance lowpass filter was simulated by TSMC 0.35um Level 3 CMOS process parameters [29]. The transistor dimensions were chosen as $W/L = 10\mu m/10\mu m$ for $M_1 \sim M_8$, $W/L = 220\mu m/2\mu m$ for $M_9 \sim M_{17}$ and $W/L = 440\mu m/2\mu m$ for M_{18}. The parameters of circuit were selected as, $V_{DD} = 3V$ and $C = 25pF$. The gain response of the filter changes from about $135kHz$ to $1008kHz$, when I_0 dc external currents are changed from $3.1\mu A$ to $232\mu A$.

Thus the cut-off frequency of the filters can be adjusted in a frequency range of about $870kHz$.

The cut-off frequencies of filter are $135kHz$, $380kHz$ and $1008kHz$ for dc control currents $3.1\mu A$, $29\mu A$ and $232\mu A$ respectively. The trans-admittance parameter value is 23.4μS for $I_{g0} = 18\mu A$ and $I_{T0} = 32.88\mu A$. The gain response obtained for the different values of the dc current sources of the filter circuit have been shown in Fig. 2.

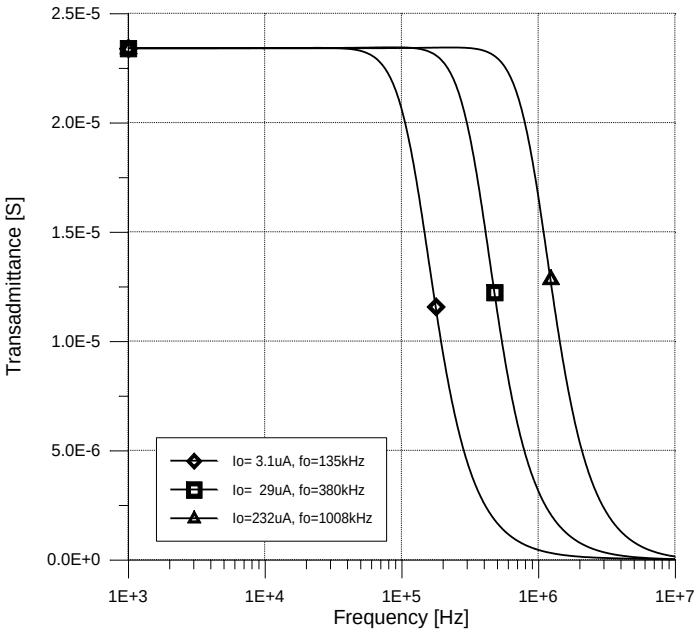


Fig.2. Gain responses of proposed filter at different values of $\boldsymbol{I_0}$.

The phase response obtained for the different values of dc current sources of filter circuit have been shown in Fig. 3.

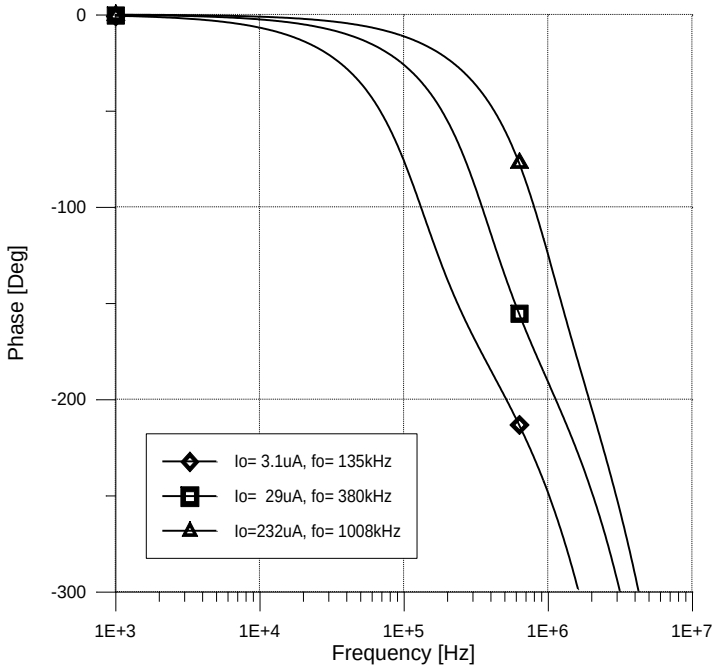


Fig.3. Phase responses of the filter at different values of $\boldsymbol{I_0}$.

The trans-admittance parameter g_0 of trans-admittance type filter can be tunable via dc control current I_{g0}. The trans-admittance parameters of filter changes from about $11.8\mu S$ to $82.7\mu S$ or I_{g0} dc control current is changed from $4\mu A$ to $240\mu A$. Thus the trans-admittance parameter of the filter can be adjusted in a gain range of about $70\mu S$. The trans-admittance parameters of trans-admittance type filter are $11.8\mu S$, $43.4\mu S$ and $82.7\mu S$ for I_{g0} dc control current $4\mu A$, $62\mu A$ and $240\mu A$ respectively. The natural frequency value is $330kHz$ for $I_0 = 29\mu A$. For these situations the gain responses of filter are shown in Fig. 4.

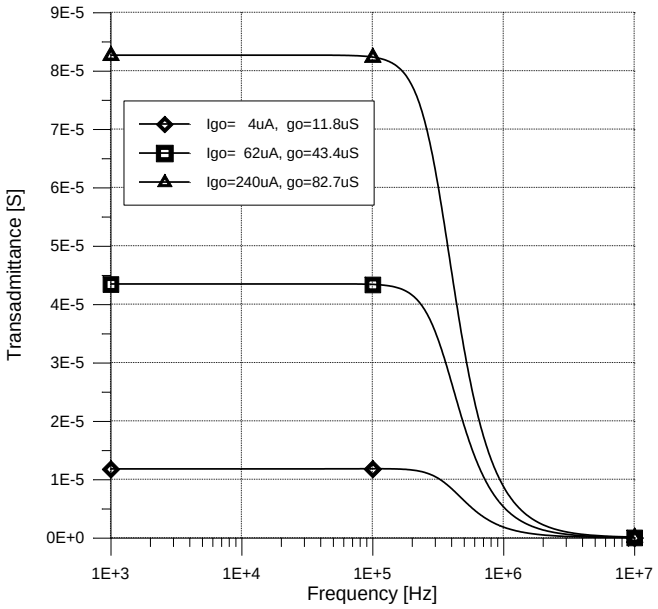


Fig.4. Gain responses at different values of $\boldsymbol{g_0}$

When the signal that has $1.5V$ dc and $200mV$ peak value at $1MHz$ frequency is applied to the input of the filter, only ac components of time domain responses of LP output for variety trans-admittance parameter values that are given in Fig. 4 are shown in Fig.5.

The quality factor of trans-admittance type filter are 0.71, 1 and 1.41 for I_q dc control current $67.6\mu A$, $37.5\mu A$ and $20.86\mu A$ reswpectively. The natural frequency value is $330kHz$ for $I_0 = 29\mu A$. For these situations the gain responses of filter are shown in Fig. 6.

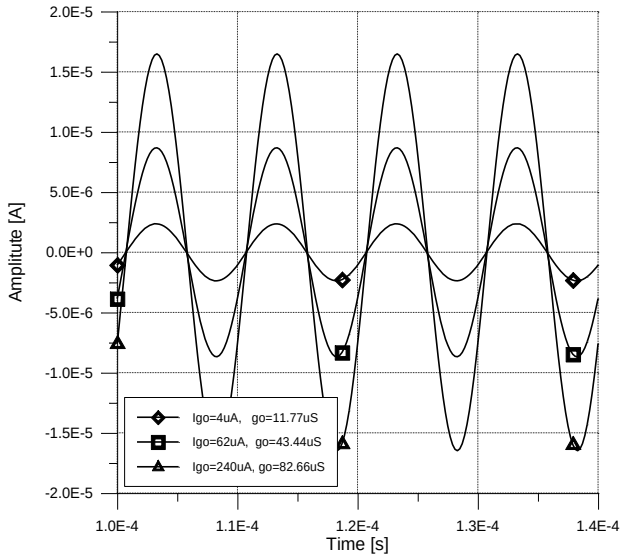


Fig. 5. Time domain responses at different values of g_0

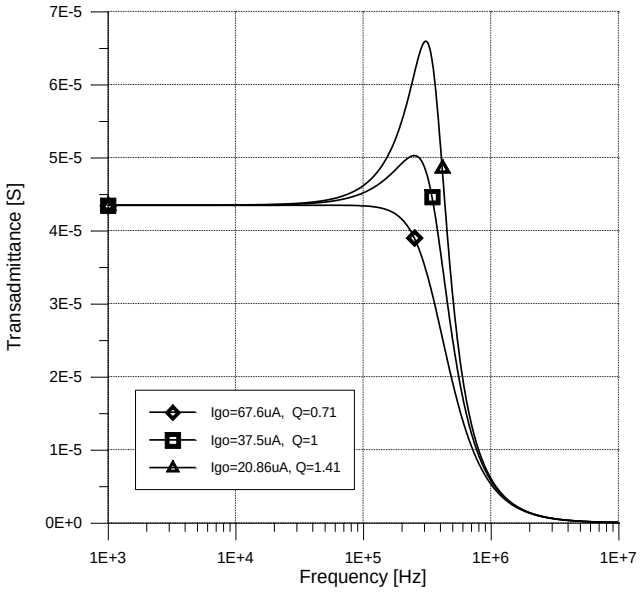


Fig.6. Gain responses at different values of Q

When peak values of sinusoidal signal input voltages at $1MHz$ frequency are varied from $70mV$ to $850mV$, Total Harmonic Distortion (THD) of output currents less than 2% for $g_0 = 23.4\mu S$.

4 Conclusion

A novel square-root-domain trans-admittance second order lowpass filter using steady-space synthesis method proposes in this study. Square-root and squarer/divider building blocks, current mirrors, dc current sources, grounded capacitors and dc voltage source are used in the filter circuit. The cut-off frequency f_0 and trans-admittance parameter g_0 of trans-admittance filter can be tuned by changing values of external dc current sources. PSPICE simulations results confirm the theoretical analysis.

References

1. B. Gilbert, Translinear circuits: A proposed classification, *Electron. Lett., 11(1),*, 1974, 14-16.
2. J. Ngarmnil and C. Toumazou, Micropower log-domain active inductor, *Electron. Lett., 32(11)*, 1996, 953-955.
3. K. Bult and H. Wallinga, A class of analog CMOS circuits based on the square-law characteristic of an MOS transistor in saturation, *IEEE J. Solid-State Circuits, 22(3),* 1987, 357-365.
4. E. Seevinck and R. J. Wiegerink, Generalized translinear circuit principle, *IEEE J. Solid-State Circuits, 26(8),* 1991, 1098-1102.
5. A. Payne and C. Toumazou, Linear transfer function synthesis using non-linear IC components, *Proc. ISCAS'96, Atlanta, USA*, 1996, 53-56.
6. M. H. Eskiyerli, A. J. Payne and C. Toumazou, State-space synthesis of integrators based on the MOSFET square law, *Electronics Letters, 32(6),* 1996, 505-506.
7. M. H. Eskiyerli and A. J. Payne, Square root domain filter design and performance, *Analog Integrated Circuits Signal Process, 22( ),* 2000, 231-243.
8. J. Mulder, *Static and dynamic translinear circuits,* (Netherlands, Delft University Press, 1998).
9. J. Mulder, W. A. Serdjn, A. C. Van Der Woerd and A. H. M. Van Roermund, A 3.3 volt current-controlled √-domain oscillator, *Analog Integrated Circuits and Signal Processing, 16(1),* 1998, 17-28.
10. A. J. Lopez-Martin and A. Carlosena, Systematic design of companding system by component substitution, *Analog Integrated Circuits Signal Process, 28(1),* 2001, 91-106.
11. G. J. Yu, B. D. Liu, Y. C. Hsu and C. Y. Huang, Design of square-root domain filters, *Analog Integrated Circuits and Signal Processing, 43(1),* 2005, 49-59.
12. S. Menekay R. C. Tarcan and H. Kuntman, The second-order low-pass filter design with a novel higher precision square-root circuit, *Istanbul Univ. J. Electr. Electron., 7(1),* 2007, 323-329.
13. J. V. Kumar and K. R. Rao, A low-voltage low power CMOS companding filter, *Proceedings of the 16th International Conf. on VLSI design (VLSI'03)*, 2003.

14. A. J. Lopez-Martin and A. Carlosena, Very low voltage CMOS companding filters based on the MOS translinear principle, *Mixed-Signal Design SSMSD Southwest Symposium on*, Austin, Texas, 2001, 105-109.
15. A. J. Lopez-Martin and A. Carlosena, 1.5V CMOS companding filter, *Electronics Lett., 38(22),* 2002, 1346-1348.
16. M. S. Keserlioglu and A. Kircay, The design of current-mode electronically tunable first-order square-root domain filters using state-space synthesis method, *Int. Review on Modelling and Simulation, 2(2),* 2009, 124-128.
17. M. H. Eskiyerli, A. J. Payne and C. Toumazou, State-space synthesis of biquads based on the MOSFET square law, *Circuits and Systems ISCAS'96*, Atlanta, GA, 1996, 321-324.
18. A. Kircay and M. S. Keserlioglu, Novel current-mode second order square-root-domain highpass and allpass filter, *6th Int. Conf. on Electrical and Electronics Eng.*, Bursa, Turkey, 2009.
19. A. Kircay, M. S. Keserlioglu and U. Cam, A new current-mode square-root-domain general notch filter, *Journal of Circuits, Systems and Computers, 22(1),* 2013, 1-10.
20. A. Kircay, M. S. Keserlioglu and F. Z. Sagi, Design of third order square-root-domain filters using state-space synthesis method, *9th Int. Conf. on Electrical and Electronics Eng.*, Bursa, Turkey, 2015.
21. T. S. A. Ragheb and A. M. Soliman, New square-root domain oscillators, *Analog Integrated Circuits and Signal Processing, 47(2),* 2006, 165-168.
22. Menekay, R. C. Tarcan and H. Kuntman, Novel high-precision current-mode circuits based on the MOS-translinear principle, *Int. J. Electron Commun. (AEU), 63,* 2008, 992-997.
23. M. S. Keserlioglu, Square-root-domain first order transadmittance and transimpedance type filter design, *ELECO-2010*, Bursa, Turkey, 2010.
24. M. S. Keserlioglu, Square-root-domain second order transadmittance type universal filter design, *Pamukkale University Journal of Engineering Science, 21(2),* 2015, 47-51.
25. A. Toker, O. Cicekoglu, S. Ozcan and H. Kuntman, High-output-impedance transadmittance type continuous-time multifunction filter with minimum active elements, *Int. J. Electronics, 88(10), 200*1, 1085-1091.
26. U. Cam, A new transadmittance type firs-order allpass employing single third generation current conveyor, *Analog Integrated Circuits and Signal Processing, 43(1),* 2005, 97-99.
27. S. Minaei, A new high performance CMOS third generation current conveyor (CCIII) and its application, *Electrical Engineering 85, Springer-Verlag, 85,* 2003, 147-153.
28. A. Carlosena and E. Cabral, Novel transimpedance filter topology for instrumentation, *IEEE Transactions on Instrumentation and Measurement, 46(4),* 1997, 862-867.
29. S. Kılınc and U. Cam, Transimpedance type fully integrated biquadratic filters using operational transresistance amplifier, *Analog Integrated Circuits and Signal Processing, 47(2),* 2006, 193-198.
30. S. Minaei, G. Topcu and O. Cicekoglu, Low input impedance trans-impedance type multifunction filter using only active elements, *International Journal of Electronics, 92(7),* 2005, 385-392.

MIMO ANN to Daylight Harvested Hybrid Lighting Control for an Office Building

Farhad Abdullayev[1], İsmail Kıyak[2]
[1] Institute of Pure and Applied Sciences, Marmara University, İstanbul, Turkey
[2] Department of Electric-Electronics Engineering, Faculty of Technology, Marmara University, Istanbul, Turkey
farhadabdullayev@marun.edu.tr

Abstract. Meteorological weather conditions may show sudden changes (cloud movements, dust) regionally, regardless of the seasons. In this study, control outputs are produced as a result of comparing both perceived sensors and national meteorological data so that temporary (sudden regional change) sky movements do not cause rapid changes in the system. The artificial intelligence supported design to control the light pipe (LP) & LED hybrid lighting system real-time application is modeled. The difference of this study from the others is the use of all natural light inputs. A system is developed to use artificial lighting armatures with the multi-input, multi-output artificial neural network (MİMO ANN) structure, which provides the required level of illumination in the environment by the amount of natural light entering through light pipes and windows. The whole space is partitioned according to natural light entrances and information from the light sensors which are placed in these areas is compared with the National Meteorological Centre (MGM) regional data, according to the office lighting strategy. MİMO ANN produces separate outputs for each area and dynamically illuminates the whole space. According to these results, the designed system is energy-efficient and sustainable with this designed control algorithm. Matlab Simulink circuit analysis performed using MGM regional data; it is calculated that the use of artificial lighting in the whole place decreased by 43.78% on a sunny day, by 32.17% on a cloudy day and by 24.36% in the overcast day.

Index Terms. Daylight harvesting; Hybrid lighting; MIMO ANN, Sustainable lighting.

1. INTRODUCTION

The transition from the traditional city concept to the concept of smart cities has become a necessity, with the rapid increase of urbanization. The need for energy is perhaps one of the biggest challenges ahead. It is assumed that global energy demand will increase by 36% by 2035, with the increasing population. The fact that the energy used in the production comes mainly from non-renewable sources, and carbon dioxide (CO_2) and similar gases emitted by using fossil fuels contribute to global warming, threatens the future of the world (Fig. 1). The most significant impact on greenhouse gases from 2000 to 2010 was exerted by global emissions of carbon dioxide from fossil fuels, which increased by 34% [1],[2]. Global urbanization has grown at a very fast pace, compared to the past. The combined use of natural and artificial light sources to save energy in lighting has increased the importance of controlling artificial lighting. The status of the energy conservation planned for Turkey shown in Fig.1.

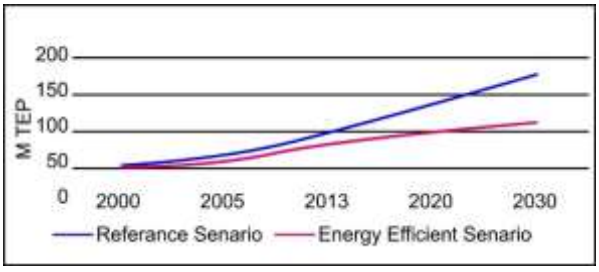


Fig.1. The status of the energy conservation planned for Turkey [3]

According to the source [3]; 20% of the electricity consumed in industrial enterprises, 30% in stores and approximately 40% in offices are spent for lighting purposes [4]. Reducing lighting consumption can be achieved by reducing installed power or reducing operating time through daylight using. Although a small percentage of commercial buildings (2% in the US) have daylight sensors [5], such controls can reduce lighting energy consumption by up to 60% [6] depending on space usage. The primary consumption calculated for the commercial office buildings in Istanbul using intelligent energy management from the State Institute of Statistics (SIS) [7] for 2014-2018 was 124.6 kwh / m^2 for heating, 92.7 kWh / m^2 for cooling and 203.5 for lighting. Therefore, any 40% reduction in lighting energy consumption (optimistic estimate) will reduce the total primary energy by about 19.4%. [8],[9]

Lighting costs can be reduced both by effective use of daylight and by eliminating unnecessary lighting, by making office lighting smart in buildings [5],[6],[10-12]. Side lighting of windows are widely used to allow daylight to enter

buildings. Daylight illumination can enter the building through upper illuminations, such as skylights or castle windows, especially for buildings with limited facades [3]. Intelligent office lighting is also of great importance for people's eye health and work efficiency [7],[13].

Different studies have been carried out on intelligent control of lighting control. In [14], they conducted a comparative study of adjustable LED lighting for efficient energy use. In [15],[16], they designed a conceptual framework based on artificial intelligence. In [17], they designed an artificial intelligence-based control system to control the lamp brightness circuitry driven by a bi-directional thyristor. In [18], [19], they have worked on a system that can automatically support the changing needs of people by integrating information technology into the field of lighting engineering [20]. In [4], [7], they have systematically examined the simulation of indoor lighting of buildings [21]. In [9], they compared the results of daylight illumination energy savings calculated by using EN1519 3: 2007 [14], [22] with DAYSIM 3 simulation results. In [23], they have evaluated reflectance values of walls, different arrangements and installation heights of historıcal buildings in terms of Lighting Energy Numerical Indicator and luminous integrity. In [24], they designed the power supply and consumption model from the lighting consumption data of six cities and towns in Northern Central Nigeria, in their paper. In [25], they analyzed the change in indoor light environment and energy consumption. In [26], they devised an energy-saving control method that could autonomously shape the light output of the LED lighting system according to the data received from the sensors. In [27], they have tested the applicability of a new concept that involves the use of different levels of dimming in the background area. In [28], they controlled the level of natural lighting in a building, simulated the brightness level with 3DS Max and processed the data with Matlab software. In [29], they analyzed the daylight-linked on-off controller of lighting, electrical lighting energy savings, and the switching frequency of an atrium building. In [30], they determined the illumination for the used and unused areas of the corporate building with the Deep neural network [30]. The differences we make with this study:

1. It has been studied only light pipe or only the light entering through the opening sıdes of the building, in the literature. There was made a design by combining all-natural light inputs in this study.
2. We simulated a real-time lighting system for light tubes and light inputs (windows) (Hybrid Lighting System) using the MIMO ANN architecture in this study.
3. There were compared annual daylight data and instant data obtained from the regional National Meteorological Center and was increased system reliability in this study.

There was a modeling of the real environment in a MIMO ANN program for a space used as an office building with a sustainable energy smart building lighting system in this study. It has been achieved the energy saving by utilizing a high level of daylight and it has been determined that the lighting control system created with MIMO ANN works correctly and fast by using this system.

2. MATERIAL AND METHOD

In the designed smart lighting system, high reflective light pipes are used for the transportation of natural light. LED retrofit luminaires with daylight color temperature (6500 ^{0}K) were used for the artificial lighting needed.

2.1 Light Pipe

The inner surface of the light tube is made of highly reflective metal or coated with materials such as Alcoa Everrbite or Silverlux, which have a reflectivity of approximately 99% [3], [7];

$$F_{in} = \pi r^2 E_{ex} \tag{1}$$

F_{in} - luminous flux falling on light pipes; F_{out} - luminous flux from light pipes; r – the radius of the light tube is (m).

$$\tau = \frac{F_{out}}{F_{in}} \tag{2}$$

Mathematically, it can be expressed as light pipe transmittance (τ) [20], [21]:

$$E_{in} = \frac{0.406 E_{ex} e^{-0.11 A_p} \pi r^2 cos^4\theta}{D^2} \tag{3}$$

E_{in} - internal brightness of a particular point; E_{ex} - external light falling into light pipes; *D* - reference point from light pipe diffuser; θ - is the angle between the normal of the light pipe diffuser and the desired point. There were used 9 light tubes with 53 mm diameter, 0.98 reflectance coefficient with inner surface coated by aluminum alloy, in the design.

TABLE 1. LIGHT PIPE SPECIFICATIONS

LP Diameter	Light Output (lumen)	Light Area(m^2)	Potential (m)
14in/350mm	6000~9100	23-28	9
21in/530mm	13900~20800	38-42	12

2.2 LED Armature

There were used 53 60x60 LED retrofit luminaires with daylight color temperature (6500 0K) as artificial lighting sources in the designed system. The technical information of the used LED luminaires is given in Table 2.

TABLE 2. TECHNICAL SPECIFICATIONS OF LED LUMINAIRE.

		Measurement unit
LED	SM134V PSD W60L60 1 xLED37S/830 NOC	-
Total lamb flux	3700	lm
Light output ratio	100.0	%
Colour Temperature	6500.0	K
Power	36	W
Dimensions	0.6*0.6*0.05	m

2.3 Method.

Learning typically occurs through the training. The training algorithm adjusts the connection weights (synapses) iteratively. Typical MLP network is arranged in layers of neurons, where each neuron in a layer computes the sum of its inputs $\mathbf{i} = [x\ y\ h\ v]^T$ and passes this sum through an activation function (f). The output of the network ($\mathbf{o}$) is defined as a matrix form;

$$o = f^2(W^2 f^1(W^1 x + b^1) + b^2) \tag{4}$$

Where; superscript defines the layer number $\mathbf{W}$ is weight matrices, **b** is bias vector, f are activation functions.

Figure 3 shows a realized one hidden layer MLP network for this work. MLP networks learn any input-output relation adjusting the weights using the backpropagation approach [19]. This algorithm adjusts the weights to minimize the mean square error as follows;

$$e = \frac{1}{2}\sum_{\gamma=1}^{p}(t^\gamma - o^\gamma)^2 \tag{5}$$

Where; t is the target, o is MİMO outputs, γ is the sample instant in q size.

The steepest descent algorithm iteratively decreases network error during learning phase at each epoch as given below;

$$w_{i.j}^{m}(k+l) = w_{i.j}(k) - \eta \frac{\partial e}{\partial w_{i.j}} \quad (6.1)$$

$$b_{i}^{m}(k+l) = b_{i}(k) - \eta \frac{\partial e}{\partial b_{i}} \quad (6.2)$$

Where η is the learning rate.

An ANN-based fitting model consisting of one input layer, hidden layer, and output layer (feed-forward network) can be used for modeling the behavior of any n×n lighting system, with n lights and n task tables. The equations governing the ANN model of the lighting system aregiven by:

$$H_f = Sig(W_h * U + b_h) \quad (7)$$

$$\bar{Y}_p = Lin(W_0 + H_f + b_0) \quad (8)$$

$\bar{Y}_p$ is the measured illuminance vector of dimension n, U is the luminaire power vector of dimension n. H_f is the output of hidden lawyer. W_h is the weight of hidden neurons and b_h is its corresponding bias. This two equations can be represented together as:

$$\bar{Y}_p(t) = M_f(U) \quad (9)$$

$M_f(U)$ is the input-output behaviour of the lighting system. The illuminance of the surface can be calculated by following formula:

$$E = \frac{d\Phi}{dA} \quad (10)$$

E is the illuminance of the surface; $d\Phi$ is the luminous flux; A is the illuminated area. The ratio of the diffuse illuminance can be calculated by following formula:

$$C = \frac{E_n}{E_w} * 100\% \quad (11)$$

C is the daylight factor; E_n is the illuminance produced by the diffuse light of the sky; E_w is the outdoorilluminance generated by the sky diffused light.

In office lighting, the space illuminated by the light pipe is grouped as edge and middle section. In three weather conditions (closed, cloudy and sunny), the brightness values of the middle and edges entering the space through light pipes are given in Table 3.

Table 3. THE LİGHT PİPE BRİGHTNESS VALUES for CLOSED, CLOUDY and SUNNY DAYS (LUX)

Center				Edges			
time& interval	avg. Overcast value(lux)	avg. cloudy value(lux)	avg. sunny value(lux)	time& interval	avg. Overcast value(lux)	avg. cloudy value(lux)	avg. sunny value(lux)
03:40	0	0	0	03:40	0	0	0
03:45	0	0	0,391	03:45	0	0	0,306
03:50	0	0	1,6675	03:50	0	0	1,305
03:55	0	0	2,944	03:55	0	0	2,304
04:00	0	0	5,6005	04:00	0	0	4,383
04:05	0	0	7,5785	04:05	0	0	5,931
.	.	.	.	.	.	.	.
12:00	25,828	54,648	102,143	12:00	21,132	44,712	79,938
12:05	25,861	54,692	102,166	12:05	21,159	44,748	79,956
12:10	25,949	54,769	102,189	12:10	21,231	44,811	79,974
12:15	26,048	54,791	102,2235	12:15	21,312	44,829	80,001
12:20	26,114	54,857	102,2465	12:20	21,366	44,883	80,019
12:25	26,246	54,912	102,2695	12:25	21,474	44,928	80,037
12:30	26,312	54,967	102,304	12:30	21,528	44,973	80,064

12:35	26,257	54,923	102,281	12:35	21,483	44,937	80,046
12:40	26,147	54,868	102,258	12:40	21,393	44,892	80,028
12:45	26,059	54,813	102,212	12:45	21,321	44,847	79,992
12:50	25,938	54,769	102,189	12:50	21,222	44,811	79,974
12:55	25,861	54,692	102,166	12:55	21,159	44,748	79,956
13:00	25,839	54,648	102,143	13:00	21,141	44,712	79,938
.	.	.	.	.	.	.	.
20:05	0	0	15,7205	20:05	0	0	12,303
20:10	0	0	13,6505	20:10	0	0	10,683
20:15	0	0	11,4195	20:15	0	0	8,937
20:20	0	0	8,2915	20:20	0	0	6,489
20:25	0	0	6,279	20:25	0	0	4,914
20:30	0	0	3,8295	20:30	0	0	2,997
20:35	0	0	1,472	20:35	0	0	1,152
20:40	0	0	0,782	20:40	0	0	0,612
20:45	0	0	0,138	20:45	0	0	0,108
20:50	0	0	0	20:50	0	0	0

Rooms with windows are grouped as window edges and middle part. The average brightness values of the sunny day are given in Table 4 due to the excessive data received in three weather conditions.

TABLE 4. LUMINOUS VALUES ENTERING THROUGH the WINDOW on SUNNY DAY (LUX)

time& interval	ISK	ISO	IS2	IS1	IS4	IS1& IS3 avg.	IS5	IS3
03:40	198,437	218,614	80,064	105,304	78,38266	98,98576	71,25696	92,66752
03:45	197,459	217,517	79,666	104,780	77,991	98,493	70,90094	92,20649
03:50	196,481	216,42	79,268	104,256	77,599	98,001	70,54492	91,74546
03:55	195,503	215,323	78,87	103,732	77,208	97,508	70,1889	91,28442
04:00	194,525	214,226	78,472	103,208	76,816	97,016	69,83288	90,82339
04:05	193,547	213,129	78,074	102,685	76,425	96,523	69,47686	90,36236
.	.	.	.	.	.	.	.	.
12:00	103,571	112,205	41,458	54,486	40,395	51,217	36,72302	47,94742
12:05	102,593	111,108	41,06	53,962	40,004	50,724	36,367	47,48638
12:10	101,615	110,011	40,662	53,438	39,612	50,232	36,01098	47,02535
12:15	100,637	108,914	40,264	52,914	39,220	49,739	35,65496	46,56432
12:20	99,659	107,817	39,866	52,390	38,829	49,247	35,29894	46,10329
12:25	98,681	106,72	39,468	51,866	38,437	48,754	34,94292	45,64226
12:30	97,703	105,623	39,07	51,342	38,046	48,262	34,5869	45,18122
12:35	96,725	104,526	38,672	50,818	37,654	47,769	34,23088	44,72019
12:40	95,747	103,429	38,274	50,295	37,262	47,277	33,87486	44,25916
12:45	94,769	102,332	37,876	49,771	36,871	46,784	33,51884	43,79813
12:50	93,791	101,235	37,478	49,247	36,479	46,292	33,16282	43,3371
12:55	92,813	100,138	37,08	48,723	36,087	45,799	32,8068	42,87606
13:00	91,835	99,041	36,682	48,199	35,696	45,307	32,45078	42,41503
.	.	.	.	.	.	.	.	.
20:05	9,683	6,893	3,25	4,1913	2,79961	3,939822	2,5451	3,688344

20:10	8,705	5,796	2,852	3,6674	2,407988	3,447356	2,18908	3,227312
20:15	7,727	4,699	2,454	3,1435	2,016366	2,95489	1,83306	2,76628
20:20	6,749	3,602	2,056	2,6196	1,624744	2,462424	1,47704	2,305248
20:25	5,771	2,505	1,658	2,0957	1,233122	1,969958	1,12102	1,844216
20:30	4,793	1,408	1,26	1,5718	0,8415	1,477492	0,765	1,383184
20:35	3,815	0,311	0,862	1,0479	0,449878	0,985026	0,40898	0,922152
20:40	2,837	0,21	0,464	0,524	0,3751	0,49256	0,341	0,46112
20:45	1,859	0,1	0,211	0,247	0,2244	0,23218	0,204	0,21736
20:50	0	0	0	0	0	0	0	0

The whole place is divided into natural and artificial lighting groups according to natural light inputs. The common working area is grouped into the middle and edge areas where the light is intense. Naturally lighted rooms are grouped as windows edge and interior areas. According to the office lighting strategy, with information from light and motion sensors placed in these areas, MİMO ANN produces separate outputs for each area and dynamically illuminates the whole space. The Matlab Simulink model of an intelligent hybrid lighting system is shown in Fig. 2.

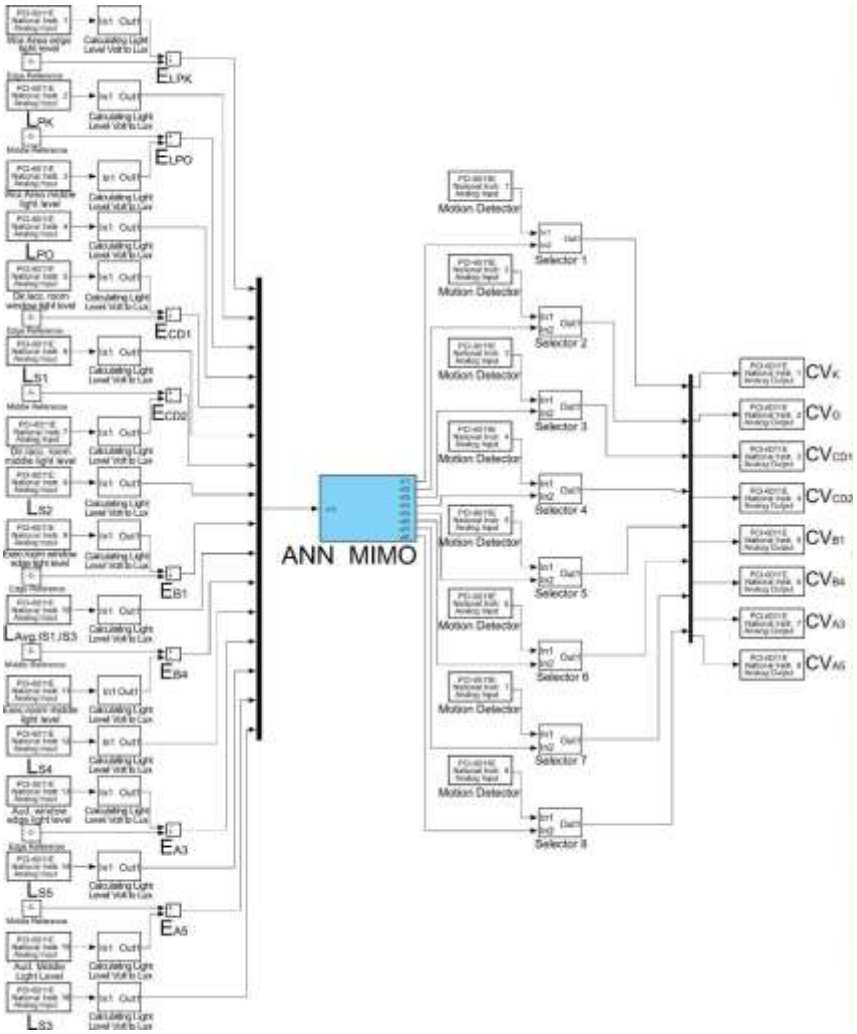


Fig.2. Smart Hybrid Lighting System Matlab Simulink model.

2.4 Design Components.

The application space measures 20 X 20 X 4 m and consists of 5 separate sections as executive room, auditorium, accounting room, director's room and common working area.

The one-year sunbathing and enlightenment data were obtained from the General Directorate of Meteorology (MGM) [31]. The weighted average values of cloudy, partially cloudy and cloudless weather days were taken from these data and a 24-hour average enlightenment database was created for all days of the year. It was assumed that no light is received from the light pipe of the space and 60 * 60 Retrofit LED luminaire was used as the lighting element in the artificial lighting design. According to office lighting standards, the level of illumination of the working plane is determined as 300 lx on average. The placement of natural and artificial lighting elements is shown in Fig. 3(a).

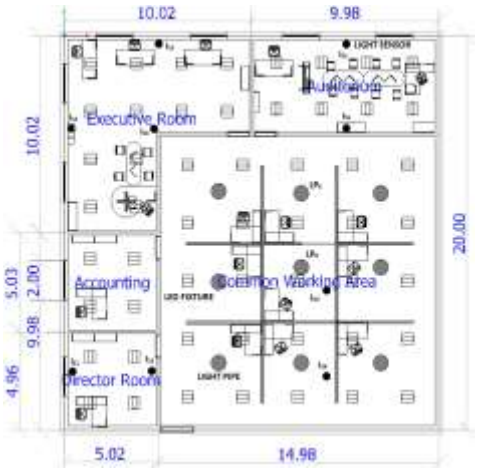


Fig. 3(a). Illuminated office and its parts.

The reference illumination values were 304 lux for edge sections and 318 lux for middle sections. There are twenty-five rows in the inner room with light pipe support, 4 rows in the director and accounting rooms, 12 rows in the executive room and 8 LED armatures in the auditorium room. The LED luminaire groups are dimmed with control voltages ranging from 0 to 10V. MIMO ANN's outputs are CV_K and CV_O for the interior room, CV_{CD1} and CV_{CD2} for the director and accounting rooms, CV_{B1} and CV_{B4} for an executive room and CV_{A3} and CV_{A5} for the auditorium. The realized MIMO network structure is shown in Fig. 4.

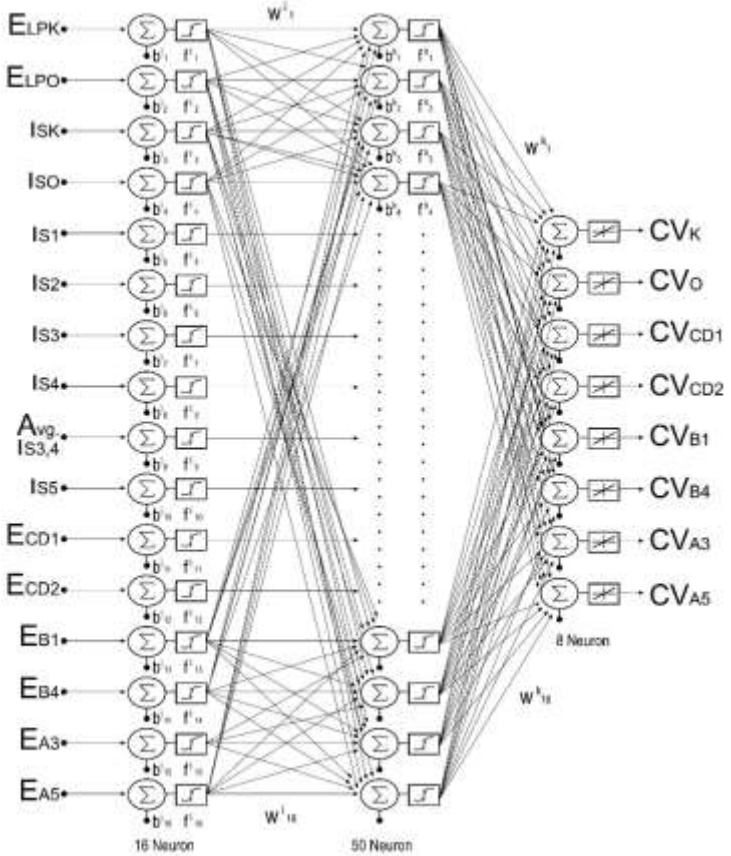


Fig. 4. The realized MIMO network structure.

There were used 16 neurons in the input layer, 50 neurons in the hidden layer and 8 neurons in the output layer in the designed MIMO ANN model. The hyperbolic tangent sigmoid function was chosen as an activation function for all layers. MIMO ANN models were trained for all rooms, respectively and the Levenberg-Marquardt (backpropagation) method was preferred. 143 samples (70% of total samples) were used for training, 30 samples were used for validation and the last 30 samples were used for testing. E_{LPK}, I_{SK}, E_{LPO}, I_{SO}, E_{CD2}, I_{S2}, E_{CD1}, I_{S1}, E_{B4}, I_{S4}, E_{B2}, I_{S1} and I_{S3}, E_{A5}, I_{S5}, E_{A3} and I_{S3} values measured for training of MIMO ANN are given in TABLE 5.

TABLE 5. INPUTS of MIMO ANN for TRAINING (LUX)

Num.	ELPK	ISK	ELPO	ISO	ECD2	IS2	ECD1	IS1	EB4	IS4	EB1	IS1&IS3 avg.	EA5	IS5	EA3	IS3
1	96,167	198,4	103,17	218,6	226,15	80,06	213,529	105,3	227,83	78,382	219,85	98,985	234,96	71,26	226,16	92,67
2	97,145	197,5	104,27	217,5	226,55	79,67	214,052	104,780	228,226	77,991	220,339	98,493	235,32	70,9	226,63	92,21
3	98,123	196,5	105,36	216,4	226,95	79,27	214,576	104,256	228,618	77,599	220,832	98,001	235,67	70,54	227,09	91,75
4	99,101	195,5	106,46	215,3	227,35	78,87	215,100	103,732	229,010	77,208	221,324	97,508	236,03	70,19	227,55	91,28
5	100,000	194,5	107,56	214,2	227,75	78,47	215,624	103,208	229,401	76,816	221,817	97,016	236,38	69,83	228,01	90,82
6	101,057	193,5	108,65	213,1	228,14	78,07	216,148	102,685	229,793	76,425	222,309	96,523	236,74	69,48	228,47	90,36
7	102,035	192,6	109,75	212	228,54	77,68	216,672	102,161	230,185	76,033	222,802	96,031	237,1	69,12	228,93	89,9
8	103,013	191,6	110,85	210,9	228,94	77,28	217,196	101,637	230,576	75,641	223,294	95,538	237,45	68,76	229,39	89,44
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
195	285,9	8,705	315,99	5,796	303,37	2,852	315,165	3,6674	303,81	2,408	315,39	3,44736	304,03	2,189	315,61	3,227
196	286,88	7,727	317,08	4,699	303,76	2,454	315,689	3,1435	304,2	2,016	315,88	2,95489	304,38	1,833	316,07	2,766

197	287,85	6,749	318,18	3,602	304,16	2,056	316,213	2,6196	304,59	1,6247	316,37	2,46242	304,74	1,477	316,53	2,305
198	288,83	5,771	319,28	2,505	304,56	1,658	316,737	2,0957	304,98	1,2331	316,86	1,96996	305,1	1,121	316,99	1,844
199	289,81	4,793	320,38	1,408	304,96	1,26	317,261	1,5718	305,38	0,8415	317,36	1,47749	305,45	0,765	317,45	1,383
200	290,79	3,815	321,47	0,311	305,36	0,862	317,785	1,0479	305,77	0,4499	317,85	0,98503	305,81	0,409	317,91	0,922
201	291,77	2,837	321,57	0,21	305,75	0,464	318,309	0,524	305,84	0,3751	318,34	0,49256	305,88	0,341	318,37	0,461
202	292,74	1,859	321,68	0,1	306,01	0,211	318,586	0,247	305,99	0,2244	318,6	0,23218	306,01	0,204	318,62	0,217
203	294,6	0	321,78	0	306,22	0	318,832	0,0001	306,22	0	318,83	0	306,22	0	318,83	0

CV_K, CV_O, CV_{CD1}, CV_{CD2}, CV_{B1}, CV_{B4}, CV_{A3} and CV_{A5} values measured for target of MIMO ANN are given in Table 6.

TABLE 6. OUTPUTS of MIMO ANN for TARGET (VOLT).

Num.	CVK	CVO	CVCD2	CVCD1	CVB4	CVB1	CVEA5	CVEA3
1	3,490	3,437438	7,478553	6,735013	7,530106	6,92775	7,749	7,12075
2	3,521	3,471074	7,490754	6,750985	7,542119	6,942784	7,760	7,135
3	3,552	3,504666	7,502956	6,766959	7,554133	6,957819	7,771	7,149
4	3,582	3,538215	7,51516	6,782934	7,56615	6,972855	7,782	7,163
5	3,612	3,571722	7,527366	6,798912	7,578168	6,987893	7,793	7,177
6	3,643	3,605187	7,539574	6,81489	7,590189	7,002933	7,804	7,191
7	3,673	3,63861	7,551784	6,830871	7,602211	7,017975	7,815	7,205
8	3,703	3,671993	7,563995	6,846853	7,614235	7,033018	7,826	7,219
.	.	.	.	.	.	.	.	.
195	9,6395612	9,811376	9,906872	9,883738	9,92135	9,890707	9,92849	9,89768
196	9,6794255	9,846939	9,91985	9,900332	9,934126	9,906308	9,94011	9,91228
197	9,7194468	9,882566	9,932833	9,916932	9,946908	9,921913	9,95173	9,92689
198	9,7596264	9,918257	9,945823	9,933536	9,959696	9,937522	9,96336	9,94151
199	9,7999653	9,954013	9,958819	9,950145	9,972491	9,953135	9,97499	9,95613
200	9,8404647	9,989835	9,971821	9,966759	9,985291	9,968753	9,98663	9,97075
201	9,8811257	9,993137	9,984829	9,983378	9,987736	9,984375	9,98885	9,98537
202	9,9219494	9,996733	9,993101	9,992166	9,992663	9,992637	9,99333	9,99311
203	10	10	10	10	10	10	10	10

Results clearly show that proposed MIMO ANN Models were successfully trained. MSE is 1.21673x10-2 for the interior room and 1.09264x10-2 for the room with a window. MIMO ANN models successfully predicted CVK, CVO, CVCD1, CVCD2, CVB1, CVB4, CVA3, CVA5 values. Overall, R is 99.76% for interior (with light pipe) room and 99.12% for rooms with windows. For the place where light pipe and artificial lighting are used together, measurement and luminous change graphs have been made according to three different weather conditions belonging to the DIM level and luminous levels of LED luminaires. In the TABLE 7, the brightness levels of LED fixture dimmable for the cloudy day are given.

TABLE 7. DIM LEVEL and LIGHT LEVEL VALUES of the ROOM wıth LIGHT PIPE for CLOUDY DAY (LUX).

DIM(VDC)	Illumination(lux)
1	27,6
2	54,8
3	91,66
4	125,08
5	159,17

6	187,69
7	217,54
8	256,21
9	291,03
10	322,165

Measurement and brightness change graphs have been made according to the three-different weather conditions of the DIM level and brightness levels of the LED fixtures, for the places where natural and artificial lighting entering through the windows, In the TABLE 8, the brightness levels for the LED armature DIM are given for cloudy days.

TABLE 8. DIM LEVEL and BRIGHTNESS VALUES of the ROOMS wıth WINDOWS for CLOUDY DAY (LUX)

Dim(VDC)	Illumination(lux)
1	27,6
2	54,8
3	91,66
4	125,08
5	159,17
6	187,69
7	217,54
8	256,21
9	291,03
10	316,42

MIMO ANN's interior light tube inputs are LP_K for the edges, LP_O for the middle sections and the difference between the reference value and the light pipe (error) E_{LPK}, E_{LPO}. Window light input for the director and accounting rooms is LS_1, the middle section is LS_2 and the difference between the reference value and the window lights (error) are E_{CD1}, E_{CD2}. Window light input of the executive room is average of LS_1 and LS_3, the middle section is LS_4 and the difference between the reference value and window lights (error) is E_{B1}, E_{B4}.

TABLE 9. RESULTS of MIMO-ANN MODELS

	Office Rooms			
	Samples	MSE	RMSE	R
Training	136	2.48903	0.0228403	1
Validation	28	5.15475	0.0525455	0.9997
Testing	28	4.39383	0.0489363	0.99955

TABLE 9 clearly shows that proposed MIMO-ANN models were successfully trained. MSE is 2.48903×10^{-6} for office rooms. Overall training R is 1, overall validation R % 99,97 and overall testing R is % 99,95 for office rooms. Best validation performance curves of MIMO-ANN model are shown in Fig. 5.

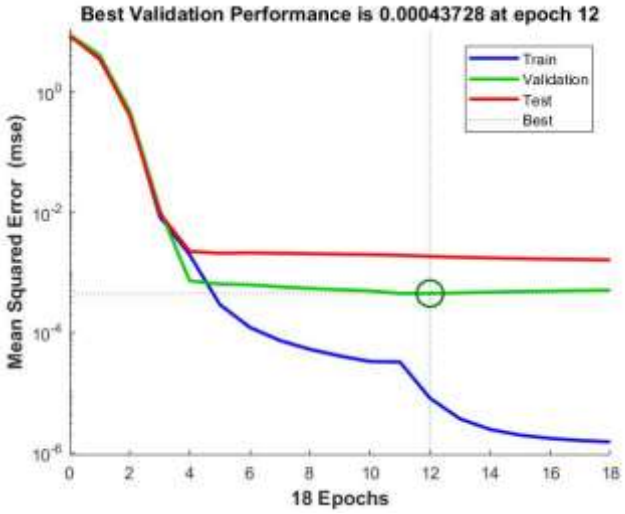


Fig.5 Best validation performance of MIMO-ANN model.

The regression analysis results are shown in Fig. 6.

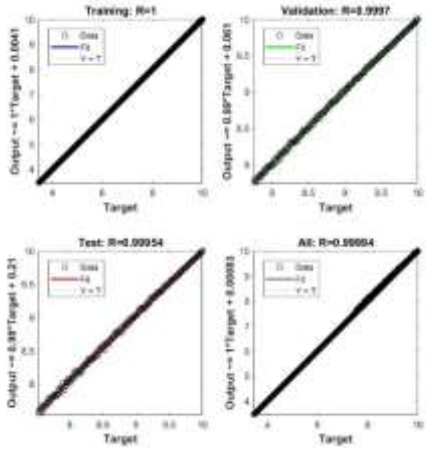


Fig. 6. Regression Analysis Results

The window light input for the auditorium is LS_3, the middle section is LS_5 and the difference between the reference value and the window lights (error) is E_{A3}, E_{A5}.

3. RESULTS AND DISCUSSIONS.

The office space was divided into 10 different lighting sections as the middle and edge sections for windowless rooms and window section and inner sections for the room with windows. The next step is to check the brightness levels of both rooms using the designed ANN models. The control voltages obtained with MIMO ANN for the whole place and the light level of the environment were analyzed according to three conditions where the air was overcast, cloudy and sunny.

The differences we make with this study (TABLE 10):

TABLE 10. THE DIFFERENCE BETWEEN METHODS.

Specifications	Proposed Method	In 15	In 19	In 23	In 27	In 29
A design by combining all-natural light inputs.	+	-	+	-	-	+

Increasıng system reliability	+	+	+	+	+	+
Daylight Harvesting	+	-	+	-	+	-
Closed Loop Control	+	+	+	+	+	+
Comparıng of sensor's values with data obtained from the MGM	+	-	-	-	-	-
Using the MIMO ANN architecture	+	-	-	-	-	-

3.1 Atificial Lighting System Results.

An artificial lighting system was developed and light distribution calculated using the Dialux Evo program.

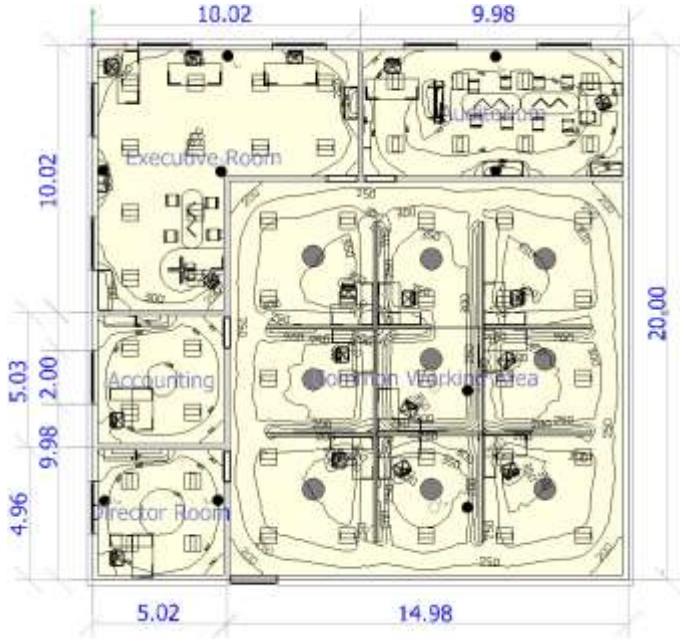


Fig. 3(b). Artificial system lighting distribution.

There were used 3700 lumens luminous flux and 36 watts LED retrofit armature for artificial lighting. The light distribution curves of the entire space are shown in fig. 3(b). According to the results of artificial lighting made with the Dialux Evo program, the lowest luminous value in the office is 301,45 lux and the highest luminous value is 318,86 lux.

3.2 MIMO Control System Results.

Values were taken every five minutes and the total number of data is 1380. When the motion sensor does not detect motion, the control voltages are reduced to zero and the LED luminaires are deactivated. In this case, the luminosity of the environment depends only on the amount of light transmitted through the light tube and windows. The control voltages of the edge luminaires for the overcast weather in the common working area with the light pipe are min 7.72 V, while medium luminaires is around 7.37 V. Since the difference between the light tube values is small, the control voltage values are very close to each other. The maximum control voltage for edge luminaires in cloudy weather at director and accounting rooms is around 7.24 V, and for medium luminaires is around 5.93 V. Fig. 7 shows the edge and medium LED luminaire control voltages, ambient light levels, light pipe values for indoor (with light pipe) area in overcast weather. Fig. 8 shows edge and middle LED luminaire control voltages, ambient light levels, light pipe values for cloudy weather for the director and accounting rooms (rooms with windows). The most significant difference is seen in sunny weather. The highest control voltage in the sunny weather for the auditorium is 4.69 V for edge fixtures, while for medium luminaires it is around 3.84 V. As can be seen from the graphs when taking the average value for an overcast day, the

illumination is measured at 114.6 lux for the edge sections from the light pipes for the common working area, and it is supplied with an adjustable voltage of 7.72 V from 186.4 lux LED lamps to provide the desired level of illumination.

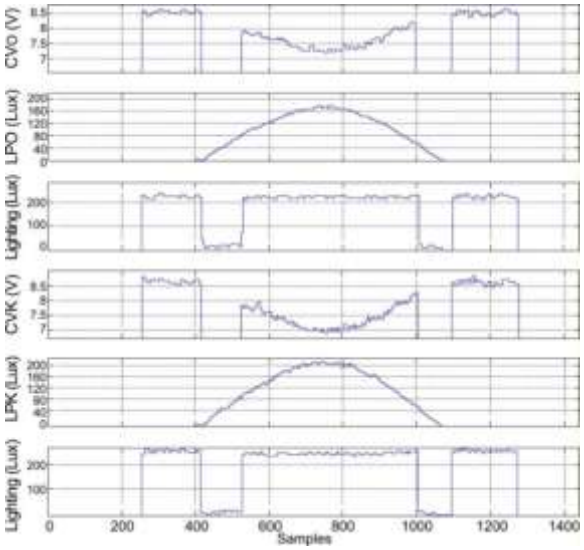


Fig. 7. The edge and medium LED luminaire control voltages, ambient light levels, light pipe values for indoor (with light pipe) area in overcast weather.

However, 148.6 lux illumination was measured for the central sections of the light tubes. The required 169.4 lux value according to the design scenario is provided by 7.37 V dim voltage from LED retrofit luminaires. When the average value of a cloudy day is taken, illumination of 172.3 lux is measured for windows and edge sections for director's and auditorium's rooms, and a DIM voltage of 5.93 V from 145.7 lux of LED lamps is applied to provide the desired level of illumination.

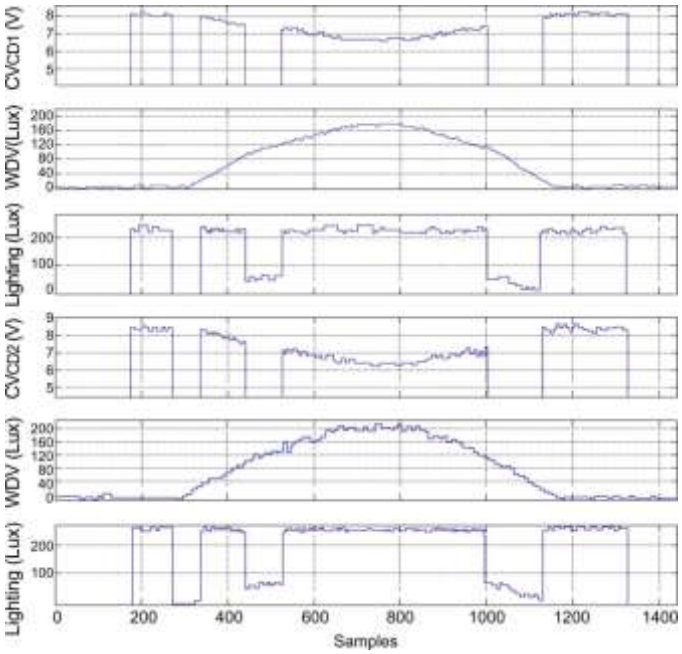


Fig. 8. LED and luminaire control voltages, ambient light levels, light pipe values in cloudy weather for the director and accounting rooms (rooms with windows).

However, 98.7 lux illumination was measured for the middle sections through the window and it was provided with 7.24 V dim voltage from the luminaires with 219.3 lux LED to provide the desired illumination level. Fig. 9 shows the control voltage, ambient light levels, light pipe values of the edge and middle LED luminaire in the open air for an auditorium (room with windows).

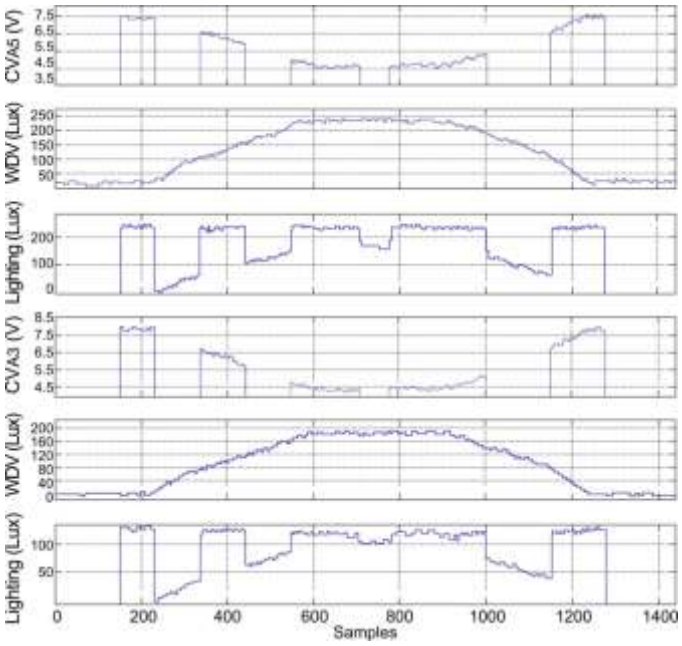


Fig. 9. The control voltage, ambient light levels, light pipe values of the edge and middle LED luminaire in the open air for an auditorium (room with windows)

When the average day is taken, a sunny day average of 208.68 lux is measured from the windows for the edge of the auditorium and was achieved a dimming voltage of 3.84 V from luminaires with LEDs of 92.3 lux to provide the desired level of illumination. However, 163.7 lux illumination was measured for the middle section from the window and it was provided with 4.69 V dim voltage from luminaires with 154.3 lux LED to provide the desired illumination level.

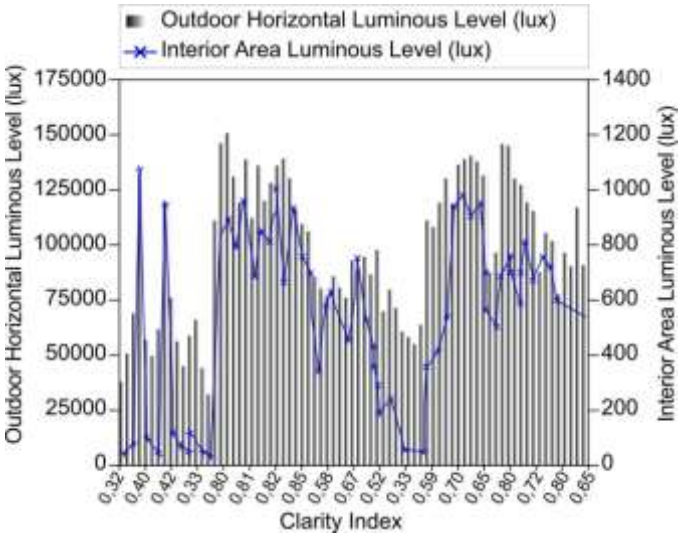


Fig. 10. Outdoor horizontal illumination level (lux) and working plane horizontal illumination level (lux) of common working area (with light pipe) depending on the aperture index.

The outdoor horizontal illumination level (lux) and working horizontal lighting level (lux) depending on the opening index for the common working area (with light pipe) are given in Fig. 10, and for the rooms with windows are given in Fig. 11.

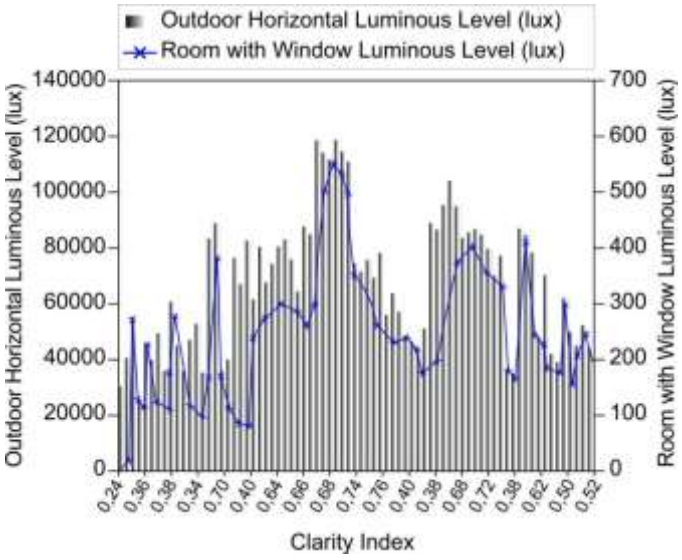


Fig. 11. Outdoor horizontal illumination level (lux) and working plane horizontal illumination level (lux) of the rooms with windows depending on the aperture index.

In bright daylight conditions, for 12: 30-14: 00 hours, the illumination level of the indoor working plane can reach an average value of 218.61 lux for the common working area with light pipe support and an average of 105.34 lux for the window assisted rooms.

4. CONCLUSIONS

The results of the study with MIMO ANN can be summarized as follows:

I. When the study data were analyzed, it was found that meteorological conditions had an effect on ANN results, and naturally, less artificial lighting was used on sunny days. The average illumination level of the whole common working area is kept constant between 301-308 lux in overcast weather, 302-314 lux in cloudy weather and 305-319 lux in sunny weather, while average light levels of all windows with rooms are 292-302 lux in overcast and 299- in cloudy, 307 lux and 302-312 lux in the sunny weather. Using motion sensors, MIMO ANN models can be activated and deactivated according to the movement in the environment, thus preventing unnecessary operation of LED luminaires.

II. Taking into consideration all the components of the daylighting system and determining the CIE color coordinates (x, y), it was observed that the system did not change the basic daylight spectral character under different standard daylight spectra (D55, D65, D75).

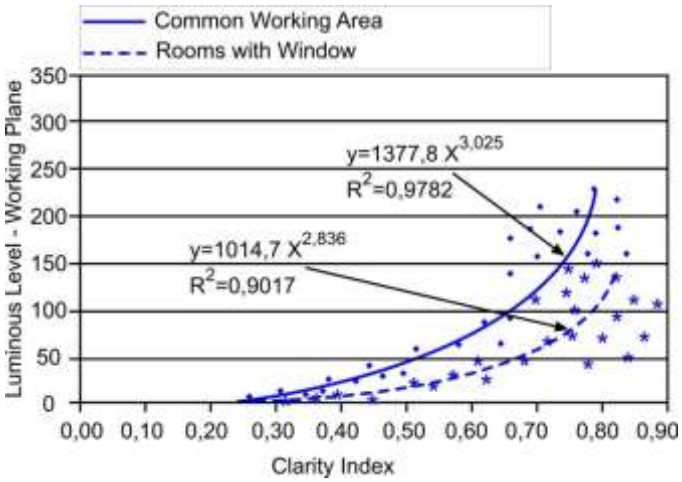


Fig. 12. The relationship between the aperture index and the horizontal illumination levels (lux) of the common working area and the working planes of the rooms with windows.

III. Fig. 12 shows the relationship between the aperture index and the horizontal illumination levels (lux) of the common working area and working rooms with windows. Similar to the working plane irradiation values, it has been shown that there is an exponential trend relationship between the aperture index and the working plane illumination level of the common working area, $R_2 = 0.9782$ affinity, and also the illumination levels of the working planes of the rooms with windows and the aperture index $R_2 = 0.9017$. As the aperture index increases, the horizontal plane level values of the working plane increase. As the aperture index increases, the light efficiency of the system increases. The increase in the aperture index means the increase of the external horizontal total radiation value at a given moment, and the light efficiency of the system increases due to the transmission of visible zone radiation of the fiber optic harness.

REFERENCES

[1] İ. Kıyak, B. Oral, V. Topuz, "Smart indoor LED lighting design powered by hybrid renewable energy systems" *Energy and Buildings*, vol. 148, pp. 342-347, 2017. Doi: 10.1016/j.enbuild.2017.05.016

[2] G. Gökmen, İ. Kıyak, "The System Design of Photovoltaic Supported Light Pipe LED Hybrid Illumination" *6. Europan Conference on Renewable Energy Systems ECRES*, İstanbul, 2018, pp. 1079-1088.

[3] T. Boyekin, İ. Kıyak, " Rooftop Solar Power Plant Based Electric Vehicle Charging Station" *6.Europan Conference on Renewable Energy Systems ECRES*, 2018, pp. 959-966.

[4] World on the edge – Energy Data Efficiency 2010, website (online), Avaliable: www.earth-policy.org/datacenter/xls/book_pb4_ch4-5_4.xls, 2018.

[5] M. Francesco, L. Fabio, P. Stefano, "Adaptive street lighting predictive control", *Energy Procedia* vol. 111, pp. 790 – 799, 2017.

[6] B. Christel, A. Myriam, K. Helianthe, R. Alexander, "Occupancy-based lighting control in open-plan office spaces: A state-of-the-art review", *Building and Environment*, vol. 112, pp. 308-321, 2016.

[7] M. Şahin, F. Büyüktümtürk, Y. Oğuz, "Yapay Sinir Ağları ile Aydınlık Kalitesi Kontrolü", *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, vol. 13, pp. 1-10, 2013.

[8] *Türkiye İstatistik Kurumu Yayın Kataloğu*, Türkiye İstatistik Kurumu (TUIK) Yayınları, Ankara, Türkiye, 2019

[9] M. Zinzi, A. Mangione, "The daylighting contribution in the electric lighting energy uses: EN standard and alternative method comparison", *Energy Procedia* vol. 78, pp. 2663-2668, 2015.

[10] A. Burçin, "Yapay Sinir Ağları İle Öngörü Modellemesi", *Öneri dergisi*, vol. 39, pp. 101-115, 2013.

[11] K. Nandha, K. Giridharan, S. Costas, J. King, H. Boon, "Smart lighting system using ANN-IMC for personalized lighting control and daylight harvesting", *Building and Environment*, vol. 139 pp.170-180, 2018.

[12] F. Juan, B. Javier, R. Sara, V. Gabriel, M. Juan, "Intelligent system for lighting control in smart cities", *Information Sciences*, vol. 372, pp. 241-255, 2016.

[13] İ . Kıyak, V. Topuz, B. Oral, "Modeling of Dimmable High Power LED Illumination Distribution Using ANFIS on the Isolated Area" *Expert System with Application*, vol. 38, pp. 11843-11848, 2011.

[14] I. Galkin, I. Milashevski, O. Teteryonok, "Comparative Estimation of Efficiency of LED Dimmers at Different Modulation Techniques", *Power Electronics and Applications (EPE 2011), Proceedings of the 2011-14th European Conference*, 2011.

[15] I. Wojnicki, A.Sedziwy, L. Katulski, "Towards AI-based Lighting Control Systems" Automatics, vol. 16, pp. 189-198, 2016.

[16] T. Boyekin, İ. Kıyak, F. Abdullayev (2019). "Artificial Intelligence Based Light Pipe Led Hybridtunnel Lighting Design" *10th International Conference of Strategic Research on Scientific Studies and Education (10th ICoSReSSE)*, İtaly, 2019, pp. 39-47.

[17] P. Anupam, J. Divya, "Mani S, Smart Control of Electric Lamp using Artificial Intelligence based Controller", *IEEE INDICON (2015)*, USA, 2017, pp. 157-164.

[18] T. Bayram, "Gravite Değerlerinin Geri Yayılımlı Yapay Sinir Ağları ile Hesaplanması", *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, vol.16, pp. 660-664, 2016.

[19] J. Eveliina, M. Esa, E. Juho, P. Henrika, Ö. Toni, "Smart and dynamic route lighting control based on movement tracking", *Building and Environment*, vol. 142, pp. 472-483, 2018.

[20] İ. Kıyak, "A Study on the Working Performance of Dimming Methods for Single and Multi Chip Power LEDs" *International Journal of Photoenergy*, vol. 6(1), pp. 1-8, 2012.

[21] A. Ashfaque, H. Pervez, S. Faheemullah, H. Zohaib, H. Nayyar, A. Muhammad, "Simulation tools application for artificial lighting in buildings" *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 3007-3026, 2018.

[22] İ. Kıyak, G. Gökmen "Increase Energy Savings with Lighting Automation using Light Pipes and Power LEDs". *ICCAE 2013: International Conference on Civil and Architectural Engineering, World Academy of Science and technology, Sweden*, 2013, pp.55-62.

[23] G. Ciampi, A. Rosato, M. Scorpio, S. Sibilio. "Retrofit Actions on the Envelope of an Existing Historical Public Building: Energy and Economic Analyses" *In proceedings of 12th International Forum of Studies -The Paths of the Merchants. Book series Fabbrica della Conoscenza 2014; 46:510-518. ISBN:978-88-6542-347-9.*

[24] S.O. Oyedepo, T. Ade,O.keye, R.O. Leramo, O. Kilanko, P. Babalol, A.O.Balogun, M.O. Akhibi. Assessment of energy saving potenti1als in covenant university, *Nigeria Energy Eng., vol.* 113 (3), pp. 7-26, 2016.

[25] Standard for Daylighting Design of Buildings. GB50033-2013. *Ministry of Construction of The People's Republic of China.* : pp．8－14. 2013.

[26] M. Rossi, A. Pandharipande, D. Caicedo, L. Schenato, A. Cenedese. Personal lighting control with occupancy and daylight adaptation.
Energy Build., vol. 105, pp. 263-272, 2015.

[27] M. Despenic, S. Chraibi, T. Lashina, A. Rosemann. Lighting preference profiles of users in an open office environment. *Build. Environ.*, vol.116, pp. 89-107, 2017.

[28] R. Sellami, M. Amirat, A. Mahrane, M.E.A. Slimani, A. Arbane, R. Chekrouni. Experimental and numerical study of a PV/Thermal collector equipped with a PV-assisted air circulation system: configuration suitable for building integration *Energy Build., vol.* 190, pp. 216-234, 2019. Doi:10.1016/J.ENBUILD.2019.03.007

[29] L. Dhw, A. Cheung, S. Chow, J.C. Lam "Switching frequency and energy analysis for photoelectric controls" *Build Environ,* vol. 85, pp. 205-210, 2015.

[30] M. Jia, R.S. Srinivasan, A.A. Raheem. "From occupancy to occupant behavior: an analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency". *Renew. Sustain. Energy Rev.*, vol. 68, pp. 525-540, 2017. Doi:10.1016/j.rser.2016.10.011

[31] Meteoroloji Genel Müdürlüğü(MGM) Yıllık Yıllık Yayınları, Meteorolojik Veri Bilgi Sunum ve Satış Sistemi, Ankara, Türkiye, 2019.

PLC Fuzzy PID Controller of MPPT of Solar Energy Converter

Abdullah J. H. Al Gizi[1]* , Cagday Dogu ATILLIA[2]* ,Salah M. Thajeel[2]*

[1]Thi-Qar Technical Collage, Southern Technical University, Iraq

[1]abdullah.algizi@stu.edu.iq

[2]Altinbas University, Istanbul, Turky

[2]cagdas.atilla@kemerburgazuniv.onmicrosoft.com

[2]eng.salahalmousoy@gmail.com

Abstract.

Development of the maximum power point tracking (MPPT) scheme for solar mounts and rectifiers remains interesting. We design of high-sensitive fuzzy (HSF) proportional integral derivative (PIDC) controller using Matlab and programmable logic controllers (PLCs) for a set point of the MPPT scheme. This proposal is founded on a synergistic mixture of the radial-basis function-neural network (RBF-NN), genetic algorithm (GA), and Sugeno fuzzy logic (SFL) systems. The finest limits of MPPT and PIDC are strong-minded through optimization, wherever RBF-NN is adjusted by means of GA to reach the best key. Also, RBF-NN is rummage-sale to improve the PID limits (got from GA) for scheming HSFL-PIDC of the MPPT scheme. The HSFL-PIDC controller is further designed to transfer in PLCs (STEP 75.5) for implementing the photovoltaic (PV) system. The entire scheme is further tuned by solar parameters under numerous operating conditions to advance the solar performance in terms of accusing and correcting. The performance of the projected analog-implemented MPPT controller is assessed by interfacing it with a hardware prototype of dual photovoltaic (PV) system. The hybrid controller conFig.s the control signal based on interaction and thereby reduces the voltage error and the oscillation in the terminal voltage control process.The achieved system is demonstrated to be efficient and robust in improving solar charging and rectifying capacity.

Keywords: GA, HSF, HSFL-PIDC, MPPT , PIDC ,PLC, RBF-NN, SFL

:

Nomenclature

MPPT	Maximum –Power- Point -Tracking	f_{best}	Highest suitability, value
K_p, K_d, *and* K_i	Proportion-coefficient, differential-coefficient, and integral-coefficient, respectively of PID controller	K_{best}	Individual with the highest suitability value in the current gen.
L	Amount of Persons Gen.	K_j^i *(g)*	a actual amount in the real-valued GA
(N),	Amount of Peers	*Rand( )*	an consistently dispersed chance number in [0,1]
(P_c)	Crossover Probability	*q*	separate with the highest suitability value using: Regulate the best values of *f* and *K* using the appearance
(Θ_r)	Generational Improvement	*L*	number of persons in a populace
(t_g)	Producing Period Constant	P_i	Replicated rate of the *I-the* separate Which can be written
(P_m)	Change Likelihood	$K_m(g)$ *and* $K_n(g)$	two persons from the present populace
fit(.)	Suitability function	σ_1 *and* σ_2	consistently dispersed random number in [0,1]
g-th	Gen.	f_{best}	Best fitness value
s_j $(K_j^{max} - K_j^{min})$	Penetrating variety for *j-th* limit $Kj \in [K_j^{max}, K_j^{min}]$ is an consistently dispersed chance number in [0,1]	*Vm*	Extreme voltage
K_{best}	Designated regulator by the future method founded on the got limit with the top suitability value of f_{best}	R_I ,L_O,C_I, V_O	Confrontation, inductance, capacitance and output voltage, correspondingly
PID parameters K_p , K_d *and* K_i	Robotically readopted by on-line learning procedure of RBF to keep the scheme error , $e(k)$ zero	k_m, a_m , b_m	Positive gain to ensure disapprovingly damped step response brought by the reference perfect
A and B	Fuzzy sets in the precursor, x and y are the input variables	*up (t) , yp (t) , ym (t)*	Control impartial design
f (x, y)	Crisp purpose in the resulting	(θ_r)	Nominal controller limits dead time from the tuning relative $\mathbf{r = \Theta_r = 0.5(\lambda + \Theta_o)}$
W_i	Lowest of the association purposes, input that got for each rule	(λ)	Shut loop time continuous for a setpoint change
RCC	Ripple Association Regulator	$\theta_o = 1$	Unique dead time
MRAC	Perfect Reference Adaptive Control construction is used to adjust the dynamics of the converter in answer to the responsibility cycle	Γ	Individuality matrix
(V_{PV}), (P_{PV})	Voltage and control	*yp (t)*	Asymptotically tracks

1. Introduction

In recent years, dedicated efforts have been made to achieve efficient solar photovoltaic (PV) systems with improved performance. In particular, as a strategy for performance improvement, solar tracking system is researched in-depth. Engin and Engin [1] projected a switch procedure that could recover the presentation and dependability of the two-axis solar follower. [2]. It was accomplished into two stages including the hardware and software development. Elagib and Osman [3] presented the design of a microcontroller based solar tracking system via solar maps, which could predict the exact apparent position (latitudes) of the sun. This system contained the solar tracking mechanical structure together with the associated electronic circuits. It established the way of regulatory a sun following panel with an entrenched microchip scheme [4].

Automation control systems are widespread in science and technology. The typical hardware device used in engineering control is Programmable Logic Controllers (PLC) that controls numerous industrial systems. Call of Simulink subsystems from MATLAB to STEP 75.5 for implementing PLC use a number of control loops responsible for upholding the actions excellence of the process are considerably increased with ever-growing difficulties of modern process plants. As a result, the modeling and simulation of the control systems became multifaceted. The major challenges in compound system are the extreme nonlinearities and the interaction between the control loops that make modeling difficult. PLC is exploited to control plants or industrial equipment's such as water and waste control, energy, oil and gas refining, etc. to cite a few [5]. Generally, a fuzzy PID controller is developed using PLC for tackling the problem of a set point pressure control in the main pressure collection system. An intelligent hierarchical coordinated control strategy is successfully applied by Hongbo et al. to a 300 MW boiler-turbine unit in China [6].

The theory introduced by Zadeh deals with the doubt and fuzziness related information concerning several parameters [7].The main objective of the AVR system is to control the terminal voltage by adjusting the generator exciter voltage. It must keep track of the generator terminal voltage all the time under any load condition by maintaining the voltage within pre-determined limits [8]. Despite much efforts in developing advanced control schemes, the control of classical integrated PLC-Fuzzy PID Simulink implemented AVR system is far from being understood [9]. The PID possessing differential, proportional and integral coefficients optimally controls the AVR system. Computational techniques such as GA and fuzzy logic are used for analytic solution[10]. A tuning fuzzy logic approach for determining the optimal PID controller parameters in AVR system is developed to obtain on-line PID parameters under various operating conditions [11].The development of an image-based sun position sensor and the algorithm was aimed to locate the sun precisely using image processing [12]. To verify the performance of the sun-tracking system an image tracking platform was established using an image-based sun position sensor, and a controller with embedded image processing algorithm. The performance testing was conducted in the laboratory. The results revealed that the proposed sun tracking system could overcome the problem of unstable tracking in cloudy weather and achieved a tracking accuracy of 0.04°. Ahmad et al [13] intended and automatic a device that could attain low power ingesting. Later, an intelligent solar following system based on entrenched microcontroller LPC2131 was implemented [14] which could recover the photoelectric change competence of the solar cell array by following the drive of the sun through 2-axis moving motors' revolution.

Xie and Zhang [16] considered a triple-junction solar cell to achieve the highest efficiency. Moreover, the system required the concentration and accurate tracking to maintain the light focusing on the solar cells during the sun movement throughout the day. The tracking system was built with two-axis driven by motor with angle sensor feedback. The two angles were determined by solar positions which were calculated from solar geometric algorithm. A new design of sun tracker for more transformation of solar energy was presented [17].

Over the years, the solar vigor has been enhanced in fresher way rendering to the cumulative rank of renewable energies by emerging a sample biaxial sun follower [20]. Three algorithms were used for tracking the sun. The first one allowed the plane to move in circular coordinates over small ranges for finding the point with the best voltage in terms of field of work and orientation repeatedly. Second algorithm aided to determine the slope of the voltage and used it to find its way. The last one is alike the second algorithm which assisted to find some appropriate points that are distinct in different times. Seme et al [21] careful two-axis sun following scheme for a PV system. The routes of this system were strong-minded via optimization. Furthermore, call of Simulink subsystems from MATLAB to STEP 75.5 can be integrated to PLC [22]. PLC control system is specially designed for industrial environment application with excellent stability and reliability. The attractive features of PLC such as simple, flexible, easy system configuration with low cost, low maintenance and running cost make them suitable for implementation. PLC being a specialized computer which interfaces a set of inputs to sensors and a set of outputs to

actuators can control the plant by performing various functions such as logic, sequencing, timing, counting and arithmetic. PLC control system is a versatile system consisting of several PLCs and computers coupled together for operation. The project of an lively solar panel dual-axis sun following system with all-out power point following fuzzy supervisor was future [23]. This system tracked the maximum solar power point and oriented the solar panel toward the Sun to enhance the efficiency of the PV generation system. The design of a two-axis standalone rotary sun tracker [24] was presented and implemented. Xiaofang and Wencheng [25] applied a new tracking system based on concentrated photovoltaic (CPV). The tracking accuracy and sensitivity of the solar tracker was found to have significant influence on the lighting rate and power generation rate of the developed CPV system. The CPV solar tracker based on ARM could track the sun mainly using the CMOS sensor. A sample of two-axis solar following system founded on a PIC microcontroller was also obtainable [26]. The parabolic reflector or parabolic dish is constructed around two feed diameter to capture the solar energy. The focus of the parabolic reflector was theoretically calculated up to an infinitesimally small point for achieving extremely high temperature. This two axis auto-tracking system was constructed using PIC 16F84A microcontroller. Dasgupta et al. [27] labeled the design and application of a original two axis sun following system which used no outside light sensors to brand PV cell opposite in the way of all-out irradiation to endorse system capability. The practical utilization of solar panels as the sensors was the main novelty of the proposed design. The hybrid of the solar cells electrical attributes was used to determine the insolation parameters. Jiao et al. [28] also proposed a two-axis sun-tracking system that kept the PV panel perpendicular to the sun light by absolute and relative position sensors signal analysis.

Diverse light sensing methods have been exploited such as LDR, image-based sun position sensor using image processing, Pyranometer, phototransistor, CMOS, and solar panels as a light sensor. Every research used different control strategy to perform tracking, where each of them was chosen to be suitable with its own system capabilities. Control is the way to lead the apparatus in a certain algorithm, where different control procedure was utilized following two major paths including closed or open loop control system. In open-loop tracking system, the tracker performed the calculation to identify the position of the sun and determined the rotational angles of the two tracking axes. A specific sun-tracking formula was utilized in order to drive the solar collector towards the sun.

This paper theme the project of a HSFL-PIDC and the best process in MPPT comptroller scheme of rules of rules useful for solar energy change and modification. This newly projected scheme joint the GA, RBF-NN and SFL schemes. The carrying into military action of the projected analog-implemented MPPT comptroller is assessed by implementing a computer hardware paradigm of dual PV system.

2. Radial Basis Function Networks

Moody et al. proposed a feed-forward two-layered RBF neural network with single hidden layer to mimic the systematic arrangement of restrictive readjustment in the human mind [29]. The RBF neural network produces the strongest response near the center of the Gaussian kernel function where each hidden node in the input data space can be regarded as a local detector and the RBF neural network is deliberative as local estimation model for the controlled processes [30-32]. The schematic of radial-basis function neural network is shown in Fig. 1. The updating algorithm for the adaptive PID based RBF can be formulated as,

$$\Delta k_p = \mu.e(k).e_p(k).\sum_{j=1}^{m} w_j h_j \frac{c_{j3} - u(k)}{\sigma_j^2} \quad (1)$$

$$\Delta k_i = \mu.e(k).e_i(k).\sum_{j=1}^{m} w_j h_j \frac{c_{j3} - u(k)}{\sigma_j^2} \quad (2)$$

$$\Delta k_d = \mu.e(k).e_d(k).\sum_{j=1}^{m} w_j h_j \frac{c_{j3} - u(k)}{\sigma_j^2} \quad (3)$$

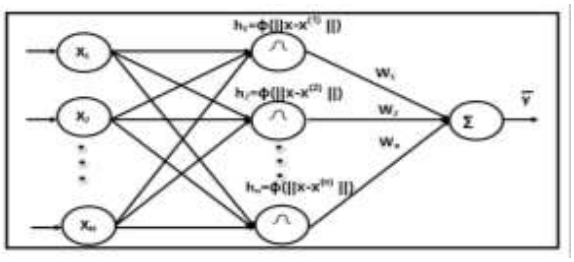


Fig. 1. Schematics of RBF neural network structure.

The PID parameters such as integral gain (K_i), the proportional gain (K_p) and the derivative gain (K_d) are automatically readjusted by RBF on-line learning algorithm to maintain the system error $e(k) = 0$. Two commands offered by Matlab namely Newrb and newrbe are used to design the RBF neural network in which Newrb adds neurons step by step until the goal is hit with long training time with minimal error and newrbe designs a network very quickly with zero error [33, 34]. In the training process, the achieved steps are: (i) neurons number in the hidden layer, (ii) the coordinates of the center of RBF function (iii) and the radius (spread) of each RBF functions in each dimension.

3. MPPT

3.1 Modeling Of Mppt System

Fig. 2 schematically illustrates the two-level adaptive control algorithm for the proposed MPPT [35] system. At the first stratum , the ripple association dominance (RCC) is used to compute the duty Hz of the convertor , which is predictable to bring all-out obtainable power to the load in the sweetheart state. In the second switch level, the perfect reference adaptive switch (MRAC) construction is used to control the subtleties of the converter in reply to the duty bicps intended from RCC, which banned the regalia voltage from fleeting fluctuation after vicissitudes in solar insolation. The most important change between these two time coefficients allowable us to decouple the analyses for RCC and MRAC, and thereby highly basic the superior general control aim . It is customary to describe briefly salient features of MRAC to validate its sexual union with RCC.

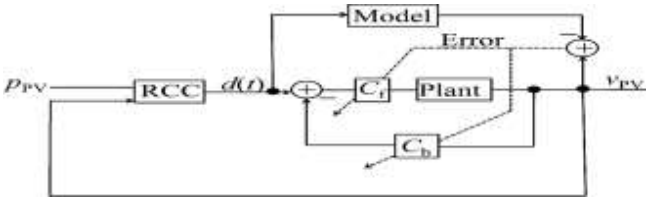


Fig. 2. Schematic diagram of the proposed MPPT system

As above-mentioned, our newly industrialized two-level MPPT switch procedure is contained of (RCC) [35-38] in the first level and MRAC [35] in the second level. Fig. 2 obviously shows that in the first switch level the collection voltage VPV and power PPV help as the inputs to the RCC unit. The RCC computes the duty cycle that brings the all-out power to the load in the stable national. The main novelty of RCC is to use the swapping ripple characteristic to the converter to trouble the system and thus path the MPP [35]. The RCC is fundamentally an better version of the P&O technique [35] except that the alarm is characteristic to the converter. Such a practice is beneficial because it refutes the need for external electrical system to vaccinate the alarm. Besides, RCC has been established to meet asymptotically to the MPP with negligible supervisor difficulty and straight-forward circuit application [35]. The RCC is based on the next observations: the product of the time-based offshoots of the collection voltage (V PV) and power (P PV) necessity be (i) greater than zero to the left of the MPP, (ii) less than zero to the right of the MPP, and (iii)precisely zero at the MPP obeying the control laws [35]:

$$\frac{dp_{pv}}{dt}\frac{dv_{pv}}{dt} > 0 \; when \; V_{PV} < V_M, \quad (4)$$

$$\frac{dp_{pv}}{dt}\frac{dv_{pv}}{dt} < 0 \; when \; V_{PV} > V_M \quad (5)$$

$$\frac{dp_{pv}}{dt}\frac{dv_{pv}}{dt} = 0 \; when \; V_{PV} = V_M \quad (6)$$

These observations lead to the control law derived in[36]

$$\frac{dd(t)}{dt} = k\frac{dp_{PV}}{dt}\frac{dp_{PV}}{dt} \quad (7)$$

where k is a constant called negative gain.

Referring to Fig.r.2, the control law in reckoning (7) can be qualitatively labeled as shadows: if *vPV* upsurges and there is a subsequent upsurge of *PV*, the scheme's working point is to the left of the MPP and so d should reduction, causing an upsurge of *vPV*. If *pPV* reductions after an upsurge in v*PV*, then the scheme's operating point is to the right of the MPP and thus d necessity increase in order to decrease v PV. Investigative equations (6) and (7), one can set the time-based copied of d to zero so that all-out power is attained. Fig. 3 displays PV the boost converter system.

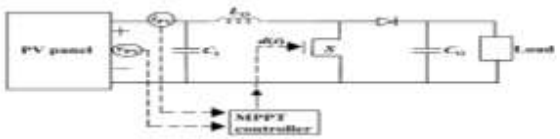


Fig. 3. Circuit diagram of MPPT controller designed for the PV boost converter system

3.2 Proposed MRAC Method

Fig. 4 displays the proposed architecture of MRAC system. The input to the overall system, r (t), is the change in duty cycle as calculated earlier using RCC. The work model corresponds to the carry-over function (equation 5). However, for convenience the sign is changed by multiplying (−single) so that the plant model has only positive coefficient. The input and output of the plant are denoted by up (t) and yp (t), respectively. Table 1 summarizes the time value of used boot converter parameter. Table 2 enlists the values of adaptive controller parameters Table 3 compares the values of nominal phrase with actual controller parameters. Table 4 summarizes the various transfer functions and the range of parameter values for solar components

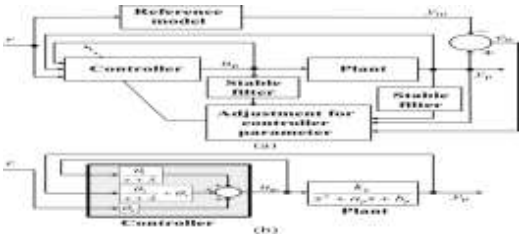


Fig. 4. Schematic presentation of the proposed: (a) MRAC structure and its (b) controller architecture

Table 1. Values of boost converter parameter

Circuit parameter	Value
R_I	45 Ω
L_O	600 μH
C_I	100 μF
V_O	350 V

Table 2. Values of adaptive controller parameters

Parameter	Value
$k_p=V_o/(L_o\times C_I)$	5.83×10^9 V (rad/sec)^2
$a_p=1(R_1\times C_I)$	222 rad/sec
$b_p=1/(L_o\times C_I)$	1.67×10^7 (rad/sec)^2
k_m	5.83×10^9 V (rad/sec)^2
a_m	8.17×10^3 rad/sec
b_m	1.67×10^7 (rad/sec)^2
ʎ	1
g	1
Γ	5×identity matrix

Table 3. Comparison between the values of nominal and actual controller parameters

	θ0	**θ1**	**θ2**	**θ3**
Nominal controller parameters	1.00	-7.95×10^3	-22.8	-3.00×10^{-4}
Updated controller parameters	1.00	-8.12×10^3	-20.4	-2.80×10^{-4}

Table 4. Various transfer functions and the range of parameter values for solar components

Components	**Transfer function**	**Parameter limits**
Plant	$TF_{plant}=K_p/s^2+a_ps+b_p$	$K_p=V_O/(L_O\times C_1)$ 5.83×10^9 V (rad/sec)^2 $a_p=1/(R_1\times C_1)$ 222 (rad/sec)^2 $b_p=1/(L_O\times C_1)$ 1.67×10^7 (rad/sec)^2 k_m, b_m, λ g, Γ 5.83×10^9 V (rad/sec)^2, 8.17×10^3 rad/sec, 1.67×10^7 (rad/sec)^2,1,1,5 identity matrix
Controller in r	$\theta_r=0.5(\lambda+\theta_0)$	$0<\theta_O\leq1, \lambda=1$
Controller feedback y_p	$y_p=(\theta_2/s+\lambda)+\theta_3$	$-22.8\leq\theta_2\leq-20.4$ $-3\times10^{-4}\leq\theta_3\leq-2.8\times10^{-4}$
Controller u_p	$u_p=\theta_1/s+\lambda$	$-7.95\times10^3\leq\theta_1\leq-8.12\times10^3$ $\theta_1\cong-8.035\times10^{-3}$

3.3 Optimization of Controller Parameters

The satisfactory operation of the system is determined by the option of the best PID control parameters. Moreover, the selection problematic of the PID controller limits is considered as an optimisation problem. The impartial function output ,

$$MinF(K_d, K_p, K_i)=(1-e^{-\beta})(O_{sh}+E_{ss})+e^{-\beta}(t_s-t_r) \quad (8)$$

The *MinF*(K_p, K_d, K_i) combines transient response counting rise metre overshoot, settling time and steady -body politic error. The satisfaction of the designer needs can be achieved by choosing suitable value of the weighting factor β. Therefore, the optimisation problem boils down to the following constraints,

$$K_p^{min}\leq K_p\leq K_p^{max},\ K_i^{min}\leq K_i\leq K_i^{max},\ K_d^{min}\leq K_d\leq K_d^{max} \quad (9)$$

Subsequent Devaraj et al. [39], GA is applied to enhance the values supervisor limits and the proposed GA is presented.

4. Proposed GA

GA is recognized as an effective and efficient technique to solve the optimization problems. In comparison to the optimization techniques, such as random search and simulated annealing, GA performance is superior that avoids local minima considered as a key issue in nonlinear systems [10, 11].

4.1 Genetic Algorithm Operators

The genetic algorithms are based on the natural selection mechanism that allows survival of the fittest and generate estimated solutions by exchanging information's to attain the optimum solution. After generating the initial population, the GA discovers new individuals by producing offspring's using the reproduction, crossover and mutation operators, which replace the old generation members and form the new generation. Once several generations are produced, the algorithm finds the best chromosome that represents the optimum or near optimum solution.

The major GA operators such as cross-over, reproduction and mutation are exploited. The convergence speed is controlled by applying various probabilities on these operators. The design of the crossover and mutation operators are carefully managed due to their immense impact on the performance of genetic algorithm [10, 11]. The details of the genetic operators used in the proposed GA are illustrated in Table 4.

4.1.1 Reproduction

In the process of reproduction, individuals are selected depending on their fitness function, the higher the fitness is, more chance for an individual to be selected for the next generation. Three main selection methods such as ranking method, fitness balanced selection and tournament selection are utilized [39]. In this work, we employ the tournament selection method, where 'n' individuals are randomly selected from the population and the best vale is chosen for additional genetic processing. This process is repeatedly performed until the mating pool is filled.

4.1.2 Crossover

The property of global search in GA is mostly determined by the crossover operator, which syndicates two-parent chromosomes to crop a new one. The variety of the designated likelihood is characteristically amid 0.6 – 1.0. One of the interesting features of the crossover operators is the relation between the generated chromosome and the location of both the parents. The generated new chromosome remains close to the parents in case both the parents are close to each other. Conversely, the search is more likely to be random [39].

4.1.3 Mutation

New chromosome is introduced into the populace for the variance procedure. Change arbitrarily makes an unimportant variety in the chromosome info. However, for unpredictable variance, the mutable takes a reliable random number amid the lower and upper bounds. In this study 'unchanging mutates' worker is used.

4.2 GA Application is adjusting limits of MMPPT PID

The best PID controller limits are got via GA tuning of PID. Two main melodies such as sign of the excellent variables (*variable symbol*) and preparation of the *fitness function* are used in this procedure.

4.3 Variable Representation

The answers of all applicants are made in the genetic populace. The answer elements of PID controller-tuning problematic comprise limits K_i, K_p and K_d. The direct picture of the answer variables decreases the processor space for storage the populace. The values of these limit got from direct change of GA into the RBF package for the best tuning of the PID manager limit are considerable for the thematic factory process of MMPPT scheme.

4.3.1 Fitness Function

The answer for the presentation of every applicant in the populace is assessed based on its suitability which is clear as a non-negative value to be exploited. Fitness is related in a conventional line with the value of impartial function. The limit set of the separate assessment can be a strong-minded using reckoning (6) for the presentation criteria. The value of separate fitness is intended by the consequence of the performance criteria via mutual calculation. The suitability, purpose is the performance of the sympathy criterion F ($KP_, K_d$,K_i) given in equation (6). Thus, the minimization of presentation criteria in eq. (6) can be distorted to the expansion of the fitness function as,

$$Fitness = \frac{k}{F(K_d, K_p, K_i) * ITAE} \tag{10}$$

Where *k* is a constant, ITAE is a time essential increased by the total error value. This is used to intensify the value of 1/F, which is usually small, so that, the genetic material suitability values happen in a wider variety.

5. Sugeno Fuzzy Model

Newly, the fuzzy set theory was used in which a adjustable is a associate of one or more sets with a association specified degree [39] . The fuzzy rule can be expressed as:

$$\text{If } x \text{ is A and } y \text{ is B then } z = f(x, y) \tag{11}$$

where A and B are fuzzy groups in the precursor, x and y are the participation variables and f (x, y) is a crisp occupation in the resulting.

Initially, the least of the association occupations input (w_i) is got for each rule, anywhere this value is the gunfire value for a specific rule. Furthermore, the general output is intended by a biased average of separate rule productions using the formulation:

$$z = \frac{\sum_{i=1}^{M} \omega_i z_i}{\sum_{i=1}^{M} \omega_i} \tag{12}$$

The limits of PIDC under numerous working settings are strong-minded by the SF system.

6. Functional Mechanisms For Application

Fig. 5 demonstrations the whole useful device for the request. The response of the HSFL-PIDC manager in MATLAB/Simulink to the runtime on PLC (target system) is tested [22].

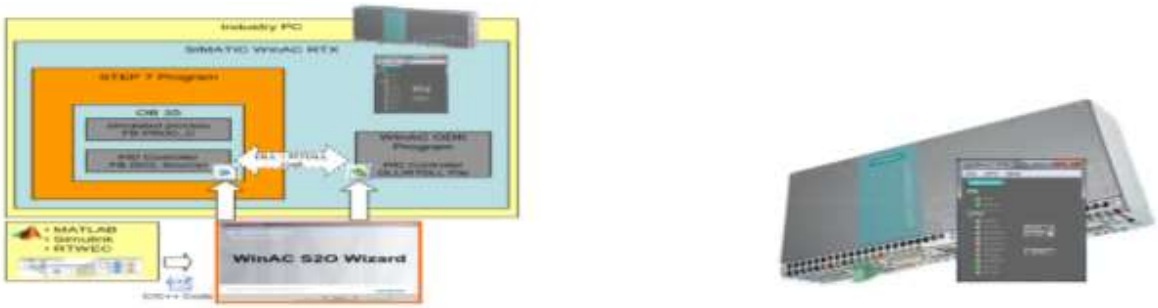


Fig. 5.Functional mechanism for the experimentation

6.1 Data types

Type change is prudently performed at the package borders due to dissimilar data type practice of STEP 7, WinAC ODK and the Real-Time Workshop Embedded Coder. The WinAC S2O Sorcerer knows the data types of the RTWEC C/C++ package and familiarizes the data in the WinAC ODK package and STEP 7 package (SCL source). Type change is achieved during the execution of WinAC S2O Wizard. The software used to transmission the data from Matlab to PLC are MATLAB V7.11 (R2010b), Simulink V7.6, Real-Time Workshop V7.6, Real-Time Workshop Embedded Coder V5.6, Microsoft Visual Studio 2008 Expert, STEP V5.5, S7-SCL V5.3,WinAC ODK V4.2 and WinAC RTX 2010 on a PC system (SIMATIC IPC427C). Hardware counting SIMATIC IPC427C (MICROBOX PC) is used in this trial. The S7 program shapes PG/PC interfaced Ethernet (192.168.2.200) and IPC427C interfaced PROFINET CP1616 (192.168.2.10) are employed.

The RGA procedure only needs the info of the suitability purpose value for each limit set. These two procedures (RGA and RBF-NN) are applied to choose a good PID control limit set for the MMPT system as labeled underneath.

7. Results And Discussion

7.1 Mythology to design fuzzy PLC PID controller

7.1.1 Development of a Sugeno Fuzzy Model to Design PID controller

The applied and final boundary to loop presentation can be submissive by understanding an indirect deceased time (θ_r) from the change relative $r = \theta r = 0.5(\lambda + \theta_0)$. The deceased time can be approached as the unique deceased time (θ_o) increased by a issue of 0.5 to the shut loop time constant (λ) for a setpoint alteration. For extreme trouble refusal by a PID, value of λ is set equal to the unique deceased time. In this state, the deceased time is equal to the unique deceased time. $r = \theta r = 0.5(\lambda + \theta_0)$, θ_o = 1, λ = *1*. The top and integrated errors for unmeasured step disturbances represents the worst case. The optimum PID parameters for actual process are got by evolving SF logic model, where, θ_r and *yp* (θ_1, θ_2) are the contributions with K_p, K_d and K_i are the outputs. Eight fuzzy sets such as 'very low (*VL*)', 'low(*L*)' ,'medium low (*ML*)', 'medium(*M*)', 'medium high (*MH*)' ,'high low (*HL*)', 'high medium (*HM*)' and 'high (*H*)' are clear for the mutable Θ_r . Likewise, the fuzzy sets clear for the mutable *yp* (θ_1, θ_2) are 'very low (*VL*)', 'low (*L*)', 'medium low (*ML*)', 'medium high (*MH*)', 'high (*H*)' and 'very high (*VH*)'. They are related with overlapping triangular involvement functions. To express the table for fuzzy rule, the values of Θ_r are varied from 0.7 to 1.0 in steps of 0.1 and the standards of *yp* are various from 1 to 2 in steps of 0.2. For each mixture of Θ_r and *yp*, the proposed RBF regulation via GA is applied to get the best values of K_p, K_d and K_i in each times. The fuzzy rule is expressed for K_p, K_d and K_i and abridged in Table 5(a)-(c), correspondingly. Throughout real-time process, the values of Θ_r and *yp* are strong-minded. Using these standards of Θr and *yp*, the best value of K_p, K_d and K_i are designed by the fuzzy rule table and the FIS editor Surgeon inference system. Contingent on the initialization (FIS editor), the standards of inputs of the fuzzy logic manager are Θ_r, *yp* and the outputs are (K_p, K_d and K_i).

The scheme with three fuzzy logic controllers (K_p, K_d and K_i) and rule viewer are established, in which each controller has two inputs (Θ_r, *yp*) and each input has fuzzy established associated with it. The input variety for Θr is $0.7 \leq \Theta r \leq 1$ and *yp* is $0.4s < yp < 1s$. The output standards of K_p, K_d and K_i be contingent on the difference of Θr with admiration to yp finished the output setting rule established on the tabulated values. The output has 72 fuzzy set rules for *Kp*, *Kd* and *Ki*, and 48 rules for each one limit as portrayed in Fig. 4 surface and rule viewer. For this value of Θ_r and y_p, the best value of K_p, K_d and K_i can be calculated using the fuzzy rule table and the sugeno inference scheme. The best results are got with the next control limits for GA: the number of generation = 50, population size = 30, crossover = 0.6, mutation probability = 0.001. Also, by choosing the correct value of the weighting factor β = 1, the presentation criterion can be made to content the designer obligation. The software for the proposed RBF and GA is written using MATLAB and performed on a laptop Intel core(TM)2 Duo CPU 5550@1.83GHz. The GA disbursed 23.79 s to reach the best solution.

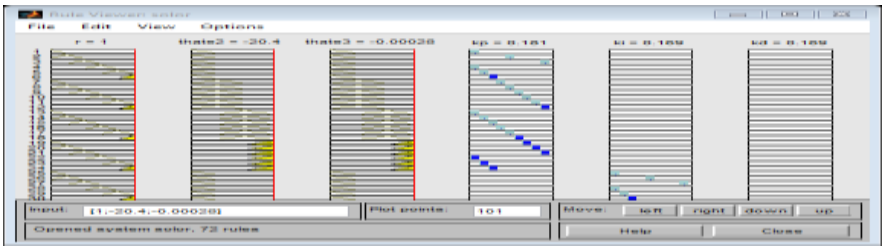


Fig. 6. Computer screen shot (the rule viewer) for the proposed Fuzzy controller of sun tracker system.

Table 5. Parameters used to generate the Sugeno fuzzy rule

	Very Low	**Low**	**Medium Low**	**Medium High**	**High**	**Very high**
$r = \theta_r = 0.5(\lambda + \theta_0)$	0.1	0.2	0.4	0.6	0.8	1
y_p , $-22.8 \le \theta_2 \le -20.4$ $-3\times10^{-4} \le \theta_3 \le -2.8\times10^{-4}$	0.55	0.6	0.7	0.8	0.9	1
$u_p = \theta_1 / s + \lambda$, $\theta_1 \cong -8.035\times10^{3}$						
(a) For proportional gain K_p						
Low (-22.8)-3,0.55	0.7947	0.7253	0.9814	0.8045	0.7292	0.9062
Medium low(-22) (-2.933)	0.9804	0.9822	0.5904	0.7195	0.9912	0.9541
Medium high (-21.2)(-2.866)	0.9062	0.6500	0.6520	0.8759	0.8416	0.6931
High(-20.4)(-2.8)	0.8651	0.9599	0.7977	0.9951	0.6197	0.8192
(b) For integral gain K_i						
Low (-22.8)	0.4946	0.4990	0.4585	0.4961	0.4990	0.4804
Medium low(-22)	0.5000	0.4883	0.4717	0.4907	0.4812000	0.4971
Medium high (-21,2)	0.4966	0.4861	0.4717	0.4995	0.4589	0.4834
High(-20.4)	0.4980	0.4844	0.4941	0.4990	0.4726	0.4956
(c) For derivative gain K_d						
Low (-22.8)	0.0428	0.0352	0.0757	0.0229	0.0150	0.0101
Medium low(-22)	0.0010	0.0199	0.18443	0.0000	0.00110	0.0528
Medium high (-21,2)	0.0613	0.1674	0.0126	0.0463	0.00422	0.1076
High(-20.4)	0.01660	0.0029	0.00023	0.0067	0.1049	0.270

7.1 MPPT of PV Using HSFL-PIDC

The MPPT scheme uses a DC to DC converter to recompense the production voltage of the solar board to save the voltage at the rate which maximizes the productive power. MPP FLC actions the values of the voltage and current at the production of the solar panel, then computes the power of the relative (P = VI) to excerpt the inputs of the manager. The crisp production of the controller signifies the duty cycle of the pulse width inflection to change the dc to dc converter. Fig. 7. Shows the block diagram of the future MPPT system.

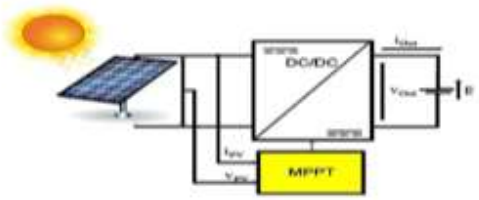


Fig. 7 . Diagram of MPPT scheme interfaced with PV solar board.

MPPT Fuzzy Logic Manager (FLC) inspects the output PV power at apiece example (time_k) and controls the variation in power compared to voltage (DP/dev). If this value is better than zero the supervisor variation the duty cycle of the pulse-width-modulation (PWM) to upsurge the voltage pending the power is all-out or the value (dp/dv)=0. Equally, dp/dv is less than zero the manager changes the duty cycle of the PWM to reduce the voltage until the power is all-out as exposed in Fig. 8.

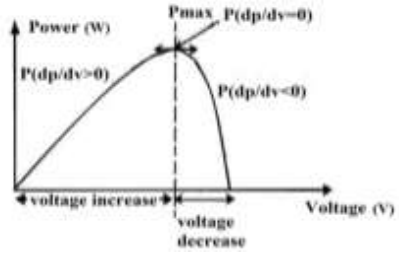


Fig. 8. Power-voltage characteristic of a PV module.

FLC owns two inputs θ_r and y_p (Θ_1,Θ_2). The applied and final limit to loop presentation can be submissive by understanding an indirect dead time (θ_r) from the change. The best PID limits for real-time process are got by emerging SFL models with *KP*, K_d and K_i as the productions, nourishing to the pulse width inflection to regulator the DC-to-DC converter. The two FLC input variables θ_r and y_p (Θ_1,Θ_2) at tested k times clear by:

$$\text{Error (k)} = \frac{\text{P(K)} - \text{P(K} - 1)}{\text{V(K)} - \text{V(K} - 1)}$$

$$\text{Change\_Error(k)} = \text{Error (k)} - \text{Error (k} - 1) \qquad (13)$$

Where P (k) is the prompt power of the photovoltaic producer. The input θ_r (k) means if the load process point at the prompt *k* is located on the left or on the right of the all-out power point on the *PV* typical. Also, the input y_p *(*Θ_1,Θ_2*)* couriers the moving way of this point. In this imitation, originally the battery-operated is first cleared and then invigorated both at a continuous current of 5.5 A. A humble current model is rummage-sale to duplicate the cordless fever. It is expected that refrigeration is chiefly via convection, and the boiler is majorly from the cordless interior resistance (R2). A normal 14 V lead-acid battery is modeled by linking seven reproductions of the 2 V cordless cell block in series [40].

The best PID limits of the fuzzy implication is approved out via real-time process and got by emerging SFL perfect for the MPP follower as showed in Fig.. 9(a,b,c). It is obvious that the signs are not flat, but they carried a constituent of the all-out power amid voltage and current. The voltage range altered from 12.5 V to 14 V and the current IL is varied amid 0 and 10 A throughout the inductance and -5, +5 as throughout the capacitor at time 0.05

sec. The voltage signal (Fig. 10) ranges between 13.3 -14.3 V at gazing and steady state at 13.4 V alike to the orientation signal, where the present signal advanced is varied amid 4.8-13 at staring and stable sate at 5.5 alike to the orientation signal showed this. The fuzzy logic supervisor with PWM sign has two compensations to the inverter. First, it crops a smooth, error-free sine-wave. Second, it attains a flat transition for the current signal and constant (no) change for the voltage signal (in variable-load case). The smooth change saves the load from obliteration by high-voltage pulses or turbulences.

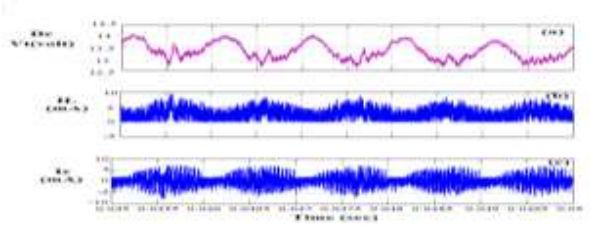


Fig. 9. DC-DC output with FC.

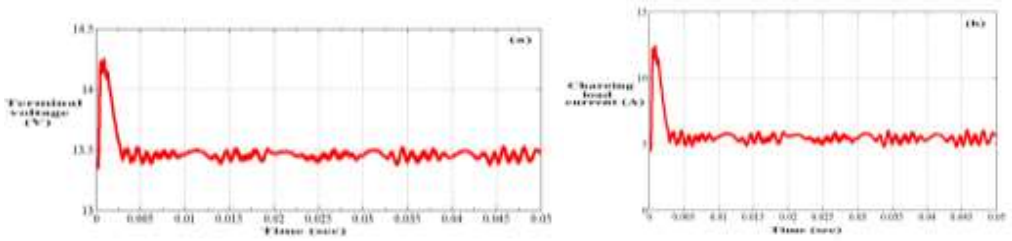


Fig. 10. The battery-operated charging output.

The design for HSFL-PIDC information transmission from MATLAB to PLC is approved out in the next way:

7.2 C/C++ Generation with RTWEC

RTWEC is an add-on of Mathworks which makes C/C^{++} code from Simulink subsystems and unlocks the Simulink model. In the blackboard by snapping on "Tools – Real -Time Workshop- Options…" a window is showed with a steering bar. The entry "Real-Time Workshop" is chosen by evasion. The next limits are required to enter: Navigation item "Real-Time Workshop" – Scheme target file: ert.tlc – Language: C++ – Click the button "Set objects" Shift "Traceability" and "Execution competence" with the "->" button to the correct field. Navigation article "Code Placement" – File wrapping format: Compact. With the correct mouse-button click on the subsystem "PID_HSFL-PIDC_disc", Select "Real-Time Workshop Build Subsystem…" in the setting menu, in the window "Build code for Subsystem" click on the "Build" button. Once the C/C^{++} code is shaped, the window mechanically closes "Build Code for Subsystem". The produced code is located in the directory of the Simulink model"...\ PID_HSFL-PIDC_disc_rtw" as shown in Fig. 11.

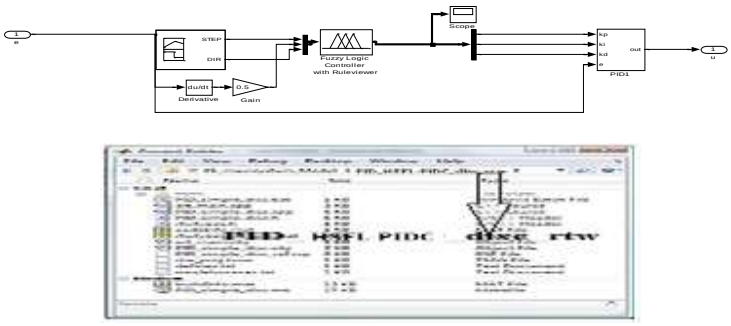


Fig. 11. Typical code for designing PID_HSFL-PIDC controlled systems.

7.3 Wines Simulink to ODK (S2O) Wizard

Wines S2O Wizard mechanically makes all obligatory chunks, and files for the addition of the Simulink subsystem into a STEP 7 project. An SCL source and a DLL or RTDLL file are created from the made C/C++ code of RTWEC. The integration is done with STEP 7 V5.x packages through WinAC S2O Wizard. Lastly, the charging of STEP 7 program with DLL or RTLL is managed to a PC system using WinAC RTX. In adding, the WinAC ODK Library is combined in the SIMATIC Administrator that needs blocks SFB65001 (CREA_COM) and SFB65002 (EXEC_COM) for performing DLL/RTDLL.

7.4 Configuration with SIMATIC Manager STEP 7 V5.5

The folder "… \ HSFL-PIDC _STEP7_V5x_Project" covers the scheme for STEP7 V5.5 with the next contents:

i. WinAC RTX is the only the hardware formation of PC position with WinAC RTX serves as a pattern and the package is empty.
ii. PID_HSFL-PIDC delivers a ready program for addition finished the WinAC S2O Wizard. The next blocks are concurrently shaped:
   - OB35 (CYC_INT5) as cyclic OB with 100 ms cycle. The fake skillful scheme and the PID manager (PID_HSFL-PIDC_disc) are named in this block.
   - DB35 (Data) as a worldwide data chunk that covers all obligatory variables such as Setpoint [Real] and PID_output [Real].
   - Process_value [Real], crea_status [Word] (delivers the position via CREA_COM), exec_status [Word] (delivers the position via EXEC_COM) and initialize [Bool].
   - FB100 (PROC_C) with Instanz-DB100 Simulated PT3 process (parameterized like the procedure in MATLAB/Simulink).
   - SFB65001 (CREA_COM) for initialization of the DLL/RTDLL file.
   - SFB65002 (EXEXC_COM) for implementation of the DLL/RTDLL file.
   - VAT_1 Mutable table with the variables of DB35 (DATA).
iii. PID_HSFL-PIDC_dll_final contains a whole software design with addition of the PID manager from MATLAB/Simulink finished the WinAC S2O Wizard. A DLL call is used for this program.
iv. PID_HSFL-PIDC_rtdll_final include a complete software design with the integration of the HSFL-PIDC controller from MATLAB/Simulink through the WinAC S2O Wizard. An RTDLL call is used in this package.

The steps obligatory to whole the package "PID_HSFL-PIDC" with the "PID_HSFL-PIDC _disc" block from the WinAC S2O Wizard are portrayed in Fig. 12.

Fig. 12.The FB PID_HSFL-PIDC_disc" into the "Network 2: PID Controller

7.5 SIMATIC Manager Opening

To start with, the instance scheme "S2O_PID_HSFL-PIDC" in the almanac "...\ HSFL-PIDC _STEP7_V5x_Project" is opened. As labeled before, throughout the implementation of the WinAC SO2 Wizard, the SCL source "PID_HSFL-PIDC_disc" is previously been combined and the purpose block ("PID_HSFL-PIDC_disc") is formed. Initial of OB 35 and supplement of "FB PID_HSFL-PIDC_disc" into the “Network 2: PID Controller” is done. The inputs and productions intersect are exposed in the Fig. 12. Few steps such as filling the package to the WinAC RTX, repetition the DLL (C_ODK.dll) or RTDLL (C_ODK.rtdll) into the almanac C:\ of the PC system with the WinAC RTX, registration RTDLL files and location the WinAC into the RUN mode are wanted to be approved out.

7.6 Implementation

The next steps are performed for ordering after charging the package to WinAC RTX: Open the "VAT_1" variable table, Click the icon "Display Variable", Change the value of the mutable "initialize" to "TRUE" and click the icon "Modify Mutable". Alteration the value of the mutable "initialize" to "FALSE" and then click the icon "Adapt Variable". Alteration the value of the mutable "Set point" to a wanted charge, e.g. 50, and click the icon "Modify Variable". The HSFL-PIDC manager makes an output due to the alteration of the set point. The procedure therefore fluctuations until the "Set point" is rehabilitated as shown in Fig. 13.

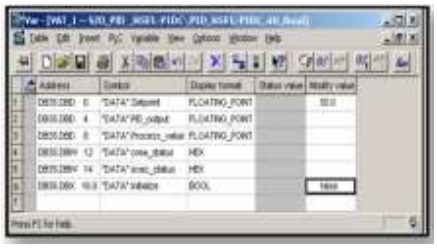


Fig. 13. The variable table VAT_1

7.7 HSFL-PIDC-PLC Controller

The PLC (HSFL-PIDC) controller hardware fitting is carried out using SIMATIC IPC427C (MICROBOX PC) and S7 program is loaded with the quantified formation, such as PG/PC interface: Ethernet (192.168.2.200) and IPC427C border: PROFINET CP1616 (192.168.2.10). For a comparison of our results, secure, HSFL-PIDC manager are active without HSFL-PIDC (following with Arduino)manager.

Table 6. The experimental results

Time Hour	Fixed voltage	Tracking voltage with Arduino	Tracking voltage with HSFL-PIDC controller
6 am	3.7	5.81	7.55
6.30 am	4.2	6.61	7.7
7 am	4.7	6.78	7.73
7.30 am	5	7.4	8.24
8 am	5.2	7.61	8.55
8.30 am	5.33	7.65	8.7
9 am	5.81	7.67	8.73
9.30 am	6.22	7.71	8.76
10 am	7.14	7.82	8.85
10.30 am	7.27	7.88	8.92
11 am	7.61	7.9	8.95
11.30 am	7.88	7.93	8.96
12	8	8.04	8.98
12.30	8.11	8.12	9
13pm	7.93	8	8.97
13.30 pm	7.79	7.97	8.96
14 pm	7.68	7.8	8.95
14.30 pm	7.49	7.78	8.87
15 pm	6.95	7.71	8.84
15.30 pm	6.59	7.62	8.82
16 pm	5.64	7.54	7.88
16.30 pm	5.17	7.47	7.75
17 pm	4.53	7.31	7.6

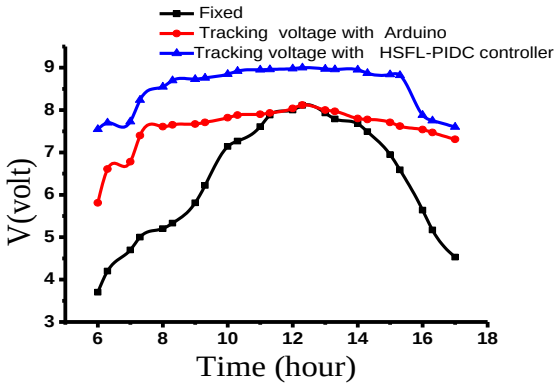


Fig. 14. Seen comparison of the overall method of Fixed (manual tracking), Tracking voltage with Arduino and Tracking voltage with HSFL-PIDC controller

7.8 Comparison of HSFL-PIDC with Other Related Works

Seen evaluation of the overall technique of Fixed (manual tracking), Following voltage with Arduino and Following voltage with HSFL-PIDC controller a shown in Fig. 14 . It is obvious that the projected HSFL-PIDC located the maximum power point (MPPT) of the PV module very close to the theoretic value as compared to the traditional PSO process and ANN as shown in Fig. 15.

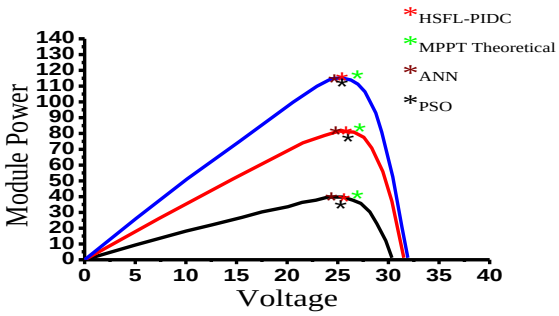


Fig. 15. Comparing P-V features of dissimilar methods

Fig. 16 demonstrations the battery charging voltage difference of the 1A-3P PV, the fixed PV and following with HSFL-PIDC throughout the day.It is shown that the upsurge of monthly-total power generation from HSFL-PIDC sun following PV is observed to be in between -13.41% and 303.26%. The efficiency of solar cells is dropped from 11.6% to 10.4% while the temperature is augmented from 38 to 48 oC and the density of dust per month was estimated at 0.0618. Humidity affects the solar cells in one way or another associated to dust collectors.[28] Thus, the proposed HSFL-PIDC procedure is efficient under dissimilar values of solar radiation as in Fig. 17. Table 6 enlists the presentation of the proposed method when compared with five other established MPPT methods.The variation of daily power generation is increased for solar tracking with HSFL-PIDC and solar radioactivity in the long-term test as shown in Fig. 18. It is further observed that power generation is augmented above 30.1829% as predictable in sunny days [14].The battery voltage of the fixed PV is a slight bit smaller than that of the 1A-3P PV meanwhile the power generation of the fixed PV is less than the 1A-3P PV, whereas the proposed HSFL-PIDC comparison with overall approaches gives development of MPPT [1].[25, 28]It is experiential that the response of an MPPT system with new HSFL-PIDC controller is highly sensitive for a very minor change as depicted in Fig.14,15,16.17,18. [28]By comparing our results with the following voltage of Arduino manual and HSFL-PIDC it is found that the proposed HSFL-PIDC controller has least settling time, less rise time and steady-state error.

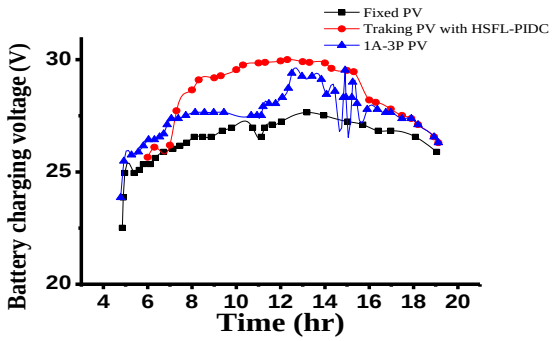


Fig. 16. Contrast of battery charging voltage of overall technique of Fixed PV[1], tracking voltage with 1A-3P PV[1] and following voltage with HSFL-PIDC.

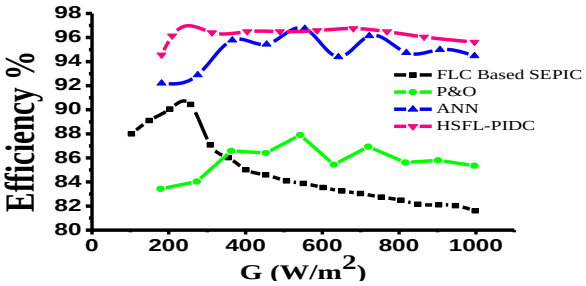


Fig. 17. Comparing efficiency of the proposed PV scheme with other approaches.

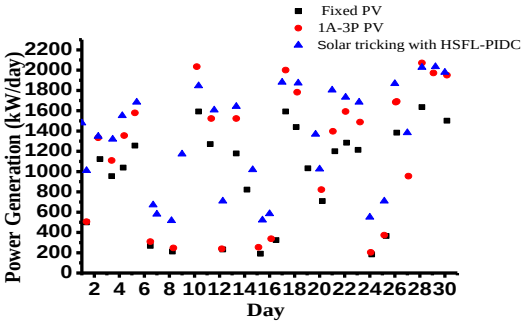


Fig. 18. Long-term test results of 1A-3P following PV, fixed PV and following voltage with HSFL-PIDC approaches.

8 Conclusion

A novel solar following system is projected and applied to improve the energy achievement of solar power plants. This dual-axis tracking system achieved the highest energy gain in every region and offered as the most versatile systems. Moreover, results exposed that it can be fitted wherever with assured high energy gain. The PLC (HSFL-PIDC) controller hardware fitting is carried out with the stated configuration such as PG/PC interface. The performance of the proposed analog-implemented MPPT with HSFL-PIDC controller is evaluated by interfacing it with a hardware prototype of dual photovoltaic (PV) system. HSFL-PIDC controller is used to regulate the location of servo motors which ensured the convenient post on the panel and progress the solar performance in full term of charging and rectifying. The use of LDR sensors and high accuracy angle sensors produced more accurate and the

efficient following scheme. The proposed automatic solar following scheme is established to achieve an overall king increase from about 10% to 50% more than the fixed-angle PV system. The achievement system is demonstrated to be efficient and robust in improving solar charging and rectifying capacity.

Acknowledgements
Abdullah is thankful to Dr.S.K.Ghushal for many valuable suggestions and critical readings of the manuscript.

References

1. Engin, M. and D. Engin. Optimization mechatronic sun tracking system controller's for improving performance. in 2013 IEEE International Conference on Mechatronics and Automation. 2013. IEEE.
2. Zolkapli, M., et al. High-efficiency dual-axis solar tracking developement using Arduino. in Technology, Informatics, Management, Engineering, and Environment (TIME-E), 2013 International Conference on. 2013. IEEE.
3. Elagib, S. and N. Osman. Design and implementation of dual axis solar tracker based on solar maps. in 2013 INTERNATIONAL CONFERENCE ON COMPUTING, ELECTRICAL AND ELECTRONIC ENGINEERING (ICCEEE). 2013.
4. Paul, M.A.J. Design and performance analysis of automated two axis solar tracking system for steam generation. in Energy Efficient Technologies for Sustainability (ICEETS), 2013 International Conference on. 2013. IEEE.
5. Chen, W., M. Xing, and K. Fang. A PLC-based fuzzy PID controller for pressure control in Coke-oven. in Control Conference (CCC), 2012 31st Chinese. 2012.
6. Liu, H., S. Li, and T. Chai, Intelligent coordinated control of power-plant main steam pressure and power output. Journal of Systems Engineering and Electronics, 2004. 15(3): p. 350-358.
7. A, Z.L., Fuzzy sets,Information and Control. , 1965. 8:338-353.
8. Farouk, N. and B. Tian. Application of self-tuning fuzzy PID controller on the AVR system. in Mechatronics and Automation (ICMA), 2012 International Conference on. 2012.
9. A, O.D., Handbook of PI and PID controller tuning rules, ed. L. Imperial College Press. 2003.
10. Baskarn, R.A.a.D.K., Speed Control of Induction Motor Using Fuzzy PI and Optimized Using GA. International Journal of Recent Trends in Engineering, 2009. 2(5).
11. Neenu Thomas, D.P.P., Position Control of DC Motor Using Genetic Algorithm-based PID Controller. London, U.K Proceedings of the World Congress on Engineering, 2009. 2(WCE 2009).
12. Lee, C.-D., H.-C. Huang, and H.-Y. Yeh, The development of sun-tracking system using image processing. Sensors, 2013. 13(5): p. 5448-5459.
13. Ahmad, S., S. Shafie, and M.Z.A. Ab Kadir. A high power generation, low power consumption solar tracker. in Power and Energy (PECon), 2012 IEEE International Conference on. 2012. IEEE.
14. Zhang, X., X. Li, and K. Lu. Research on an intelligent solar tracking system based on LPC2131. in 2012 3rd IEEE International Conference on Network Infrastructure and Digital Content. 2012. IEEE.
15. Kivrak, S., M. Gunduzalp, and F. Dincer, Theoretical and experimental performance investigation of a two-axis solar tracker under the climatic condition of Denizli, Turkey. Przegląd Elektrotechniczny, 2012. 88(2): p. 332-336.
16. Xie, F. and G. Zhang. Simulation and Analysis of Computer Control Tracking System Based on Solar Geometric for Fresnel Lenses Concentration Photovoltics. in Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2011 International Conference on. 2011. IEEE.
17. Li, Z. and R. Zhou. Sun tracker design based on AVR. in Power and Energy Engineering Conference (APPEEC), 2011 Asia-Pacific. 2011. IEEE.
18. Barsoum, N., Fabrication of dual-axis solar tracking controller project. Intelligent Control and Automation, 2011. 2(02): p. 57.
19. Kassem, A. and M. Hamad. A microcontroller-based multi-function solar tracking system. in Systems Conference (SysCon), 2011 IEEE International. 2011. IEEE.
20. Shama, F., et al. A Novel Design and Experimental Study for a Two-Axis Sun Tracker. in Power and Energy Engineering Conference (APPEEC), 2011 Asia-Pacific. 2011. IEEE.
21. Seme, S., G. Štumberger, and J. Voršič, Maximum efficiency trajectories of a two-axis sun tracking system determined considering tracking system consumption. IEEE Transactions on power electronics, 2011. 26(4): p. 1280-1290.
22. Siemens, SIMATIC WinAC S2O Wizard February 2012
23. Lu, H.-C. and T.-L. Shih. Fuzzy system control design with application to solar panel active dual-axis Sun tracker system. in Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on. 2010. IEEE.
24. Al Nabulsi, A., et al. Efficiency optimization of a 150W PV system using dual axis tracking and MPPT. in Energy Conference and Exhibition (EnergyCon), 2010 IEEE International. 2010. IEEE.

25. Xiaofang, S. and G. Wencheng. A Sun Spot Center Orientation Method Based on Ellipse Fitting in the Application of CPV Solar Tracker. in Intelligent System Design and Engineering Application (ISDEA), 2010 International Conference on. 2010. IEEE.
26. Oo, L.L. and N.K. Hlaing. Microcontroller-based two-axis solar tracking system. in Computer Research and Development, 2010 Second International Conference on. 2010. IEEE.
27. Dasgupta, S., et al. Dual axis sun tracking system with PV cell as the sensor, utilizing hybrid electrical characteristics of the cell to determine insolation. in 2010 IEEE International Conference on Sustainable Energy Technologies (ICSET). 2010. IEEE.
28. Jiao, H., et al. Design of automatic two-axis sun-tracking system. in 2010 International Conference on Mechanic Automation and Control Engineering. 2010.
29. Yao-Lun, L., et al. Design an Intelligent Neural-Fuzzy Controller for Hybrid Motorcycle. in Fuzzy Information Processing Society, 2007. NAFIPS '07. Annual Meeting of the North American. 2007.
30. S.N. Qasem, S.M.S., Memetic Elitist Pareto Differential Evolution algorithm based Radial Basis Function Networks for classification problems. Applied Soft Computing, 2011. 11(1): p. 5565–5581.
31. S.N. Qasem, S.M.S., Radial basis function network based on time variant multi-objective particle swarm optimization for medical disease diagnosis. Applied Soft Computing, 2011. 11(1): p. 1427–1438.
32. N. A. Al-geelani, M.A.M.P., R. Q. Shaddad Characterization of acoustic signals due to surface discharges on H.V. glass insulators using wavelet radial basis function neural networks Applied Soft Computing, 2012. 7(2): p. 1327-1338.
33. Sang Jeen, H., G.S. May, and P. Dong-Cheol, Neural network modeling of reactive ion etching using optical emission spectroscopy data. Semiconductor Manufacturing, IEEE Transactions on, 2003. 16(4): p. 598-608.
34. Shu-Kun Zhao, M.-W.K., Yi-Seul Han, Se-Youn Jeon, Yun-Keun Lee, and Seung-Soo Han, Radial Basis Function Network for Endpoint Detection in Plasma Etch Process. Springer-Verlag, 2010. 67: p. 253–263.
35. Khanna, R., et al., Maximum power point tracking using model reference adaptive control. IEEE Transactions on power Electronics, 2014. 29(3): p. 1490-1499.
36. Krein, P.T. Ripple correlation control, with some applications. in Circuits and Systems, 1999. ISCAS'99. Proceedings of the 1999 IEEE International Symposium on. 1999. IEEE.
37. Logue, D. and P. Krein. Optimization of power electronic systems using ripple correlation control: A dynamic programming approach. in Power Electronics Specialists Conference, 2001. PESC. 2001 IEEE 32nd Annual. 2001. IEEE.
38. Kimball, J.W. and P.T. Krein, Discrete-time ripple correlation control for maximum power point tracking. IEEE Transactions on Power Electronics, 2008. 23(5): p. 2353-2362.
39. Devaraj, D. and B. Selvabala, Real-coded genetic algorithm and fuzzy logic approach for real-time tuning of proportional-integral - derivative controller in automatic voltage regulator system. Generation, Transmission & Distribution, IET, 2009. 3(7): p. 641-649.
40. Jackey, R.A., A simple, effective lead-acid battery modeling process for electrical system component selection. 2007, SAE Technical Paper.

Improved Features of Intelligent PLC-Fuzzy Generator parameters Controller

Abdullah J. H. Al Gizi1*

1 Thi-Qar Technical Collage , Southern Technical University, Iraq

1abdullah.algizi@stu.edu.iq

Abstract:

The quality of an intelligent control system with certain degree of autonomy is prerequisite for effective functioning. We design a fuzzy proportional integral derivative (PID) controller using Matlab and programmable logic controllers (PLCs) for a set point voltage and frequency. The proposed controller aims to maintain the terminal voltage and frequency continually under any loads and operational conditions which can be attained to the desired range via the regulation of the Generation Governor (GG) system. The main voltage control system uses PLCs to implement the AVR action. The existing algorithm being based on controllers including the radial basis function neural networks (RBF-NN), real-valued genetic algorithm (RGA), and Sugeno fuzzy logic (SFL) requires optimal tuning for thematic factory operation of the GG system. The newly developed controller combines the RGA, RBF-NN , generator parameters (Kg and tg) and fuzzy logic control (FLC) to determine the optimal PID controller in the AVR system. Combination of the Generator Parameters (Kg and tg) , Fuzzy PID (GPFPID) controller and industrial governor speed controller (KS 92) is used as hybrid full control system for the voltage and frequency. Optimal PID gains obtained by a combined RGA and RBF for various operating conditions are utilized to develop the rule base of the Sugeno fuzzy system. The GPFPID controller is further designed to transfer in PLCs (STEP 75.5) for implementing the AVR system with improved system response. The hybrid controller conFig.s the control signal based on interaction and thereby reduces the voltage error and the oscillation in the terminal voltage and frequency control process. An excellent voltage and frequency control performance is achieved when the proposed hybrid controller is impremented on a practical AVR and governor system in synchronous generator to improve the transient response. The suitability of the hybrid controller is demonstrated via experiment.

Keywords: GG, AVR, GPFPID, RBF, SFL, FLC.

1. Introduction

It is well known that the main function of Generation Governor (GG) is the hybrid control, where the terminal voltage and frequncy are combined by regulating the generator parameter voltage and govenor speed. The PID inside the GG is responsible for its optimal control. It is comprised of three coefficients including differential, proportional, and integral. For an industrial control system is it important to obtain high-quality performances. Thus, design of an effective and efficient generator parameters PID (FOPID) controller as a generalization of a standard PID controller based on generator parameters (FO) calculus is required [1]. A FO PID (FOPID) controller is designed for an Automatic Voltage Regulator (AVR) system with broader performance objectives [2]. Furthermore, FOPID controller is an application of fractional calculus theory in PID controller [3]. Fractional-order proportional-integral-derivative (FOPID) controllers are designed for load-frequency control (LFC) of two interconnected power systems [4]. A stochastic multi-parameters divergence method for online parameter optimization of fractional-order proportional–integral–derivative (*PID*) controllers was presented [5]. An adaptive optimal control design approach was followed for automatic voltage regulator (AVR) system where policy iteration technique based adaptive critic scheme was utilized [6]. It dealt with the design, implementation and analysis of an integer order (IO) as well as generator parameters (FO). A system was proposed based on Proportional Integral Derivative (PID) controller for speed regulation in a chopper fed Direct Current (DC) motor drive [7]. A chaotic ant swarm optimization (CASO) was utilized to tune the parameters of both single-input and dual-input power system stabilizers (PSSs) [8]. The additional parameters of differ-integral orders on one hand rendered more flexibility TO the generator parameters elements which were essentially infinite dimensional filters [9].The optimum PID parameters required to formulate the fuzzy rule table were generated via the real-coded genetic algorithm (RGA) [10]. Zhang and Zhang [11] proposed an vehicle stability adaptive PID control algorithm with single neuron network (NN). Kun et al. [12] designed an optimum PID controller using radial basis function (RBF) for direct-drive permanent magnet linear synchronous motor (PMSM). Moody and Darken [13] introduced a feed-forward two-layered RBF NN with single hidden layer. In this controller, the load frequency control (LFC) and automatic voltage

regulator (AVR) was installed in each generator to regulate the real and reactive power flows. A best PID controller for a universal second-order system was improved using a linear-quadratic regulator (LQR) method [22]. This approach required an appropriate weighting functions for acceptable performance. Minglin [23] proposed a method for designing PID-like fuzzy controller with FPGAv. The feed forward fuzzy PID (FFFPID) controller was used to improve the performance of high pressure common rail system [24]. Sinthipsomboon et al. [5–8] used computational techniques such as GA and fuzzy for analytic solution of FFFPID controller. A combined fuzzy and fuzzy self-tuning PID controller was proposed to surmount the limitations of the current hybrid fuzzy PID controller performance, where system parameters alterations required a new PID controller adjustment variable [25]. An improved Fuzzy PID controller was used [26, 27] to control the speed of brushless DC motor. PLC was used to develop a fuzzy PID controller for a set point pressure control in the main pressure collection system [28]. The design and stability analysis of Takagi-Sugeno-Kang (TSK)-type full-scale fuzzy proportional-integral-derivative (PID) controller was performed [29]. Parameter self-setting fuzzy PID control algorithm was applied to control drying temperature for improving the temperature fluctuations [30]. An improved fuzzy PID controller algorithm was proposed based on DSP [31]. In comparison to NSGA-II-based FOPID design algorithm [32], the proposed MOEO algorithm adopted individual-based iterated optimization mechanism with only mutation operation called polynomial mutation. From the perspective of algorithm design, the proposed MOEO algorithm was relatively simpler than NSGA-II [32] and reported competitive single objective evolutionary algorithms such as GA [33, 34], PSO [33-35], CAS [34] due to its fewer adjustable parameters and single individual base dictated optimization mechanism with only mutation operation. The PID inside the AVR being the incharge of the optimal control contains differential, proportional, and integral coefficients. Despite of many efforts an intelligent control system with optimum autonomy and efficient function is far from being achieced. In this view, the present paper proposed the design of a combined Generator Parameters (Kg and tg),RGA, RBF-NN and SFL approach to determine the optimal PID controller of AVR system. This novel GPFPID voltage and industrial speed controller is further used to develop an AVR and Governor system. Besides, the proposed algorithm could search a high-quality solution effectively via full control system with improved transient response. This paper is organized as follows. Section 2 describes the RGA and RBF-NN algorithm. Section 3 introduces the concept of AVR system in terms of modeling and optimization of the controller parameters. Section 4 highlights the notion of the Sugeno fuzzy system that is used to determine the parameters of the PID controller under different operating conditions. Design of Fuzzy PID Controller using generator parameters (Kg and tg) is depicted in Section 5. The detailed design of the PLC-based fuzzy controller with industrial controller KS 92, the experimental results, and discussions are provided in Section 6. Section 7 concludes the paper.

2. Real-Valued GA and RBF-NN Algorithm

This paper proposed a real-valued GA(RGA) and a RBF-NN algorithm to select an appropriate control parameter set K = (kp, kd, ki) of the PID controller. A defined fitness function was used to guide these two algorithms for finding an appropriate control parameter set. In the RGA, the control parameter set K = (kp, kd, ki) is viewed as an individual and each parameter value is coded by a real number [36]. For a set of L individuals in a generation, the procedure of the proposed RGA is described by the following steps:

Step 1:Initialize RGA by setting L, the number of generations (N), the crossover probability (Pc), generator gain (Kg), generator time constant (tg) and the mutation probability (Pm). The i-th individual of the population with L individuals in the g-th generation is denoted by:

$$Ki(g) = (K_1^i(g), K_2^i(g), K_3^i(g)) \quad (1)$$

where the number of parameters in the parameter set is 3 and

$K_j^i(g)$,i ϵ{1, 2, …, L}, j ϵ[37], g ϵ{1, 2, …, N} is the j-th parameter of the i-th individual in the g-th generation.

Kgϵ{0.7, 0.8,0.9,1.0}, tg ϵ{1, 1.2,1.4,1.6,1.8,2}. Note that $K_j^i(g)$ is a real number in the real-valued GA.

Step 2: Set g = 1 for the first generation and randomly generate the initial population with L individuals as:

pop(1) = {K1(1), K2(1), …, KL(1)} by $K_j^i(1) = K_j^{min} + (K_j^{max} - K_j^{min}).rand()$, (2)

where i = 1,2,…,L; j = 1,2,3, the searching range of the parameter Kj is $[K_j^{min}, K_j^{max}]$ (i.e.,Kj ϵ $[K_j^{min}, K_j^{max}]$) and rand( ) is an uniformly distributed random number in [0,1].

Step 3:Calculate the fitness value of each individual in the g-th generation using:

$$fi = fit(Ki(g)),\ i = 1,2,3,\ldots,L \quad (3)$$

where fit(.) is the fitness function.

Step 4: Find an index q of the individual with the highest fitness value by:

$$q = \arg\max fi\ i \in \{1,2,\ldots\ldots,L\} \quad (4)$$

$$fbest = fq = \max fi\ i \in \{1,2,\ldots\ldots,L\} \quad (5)$$

and

$$Kbest = Kq \quad (6)$$

where fbest is the highest fitness value in the current generation and Kbest is the individual with the highest fitness value in the current generation.

Step 5: If g > N, then go to Step 11. Otherwise, go to Step 6.

Step 6: Reproduce each individual in the reproduction process by:

$$ni = L.Pi,\ i = 1,2,\ldots..,L \quad (7)$$

where ni is the reproduced number of the i-th individual, L is the number of individuals in a population,and Pi is the reproduce rate of the i-th individual and is determined by:

$$Pi = \frac{f^i}{\sum_{i=1}^{L} f^i},\ i = 1,2,\ldots.,L \quad (8)$$

where f i is the fitness value of i-th individual.

Step 7: Choose two individuals Km(g) and Kn(g) from the current population (m, n {1, 2, …, L}) to be the parents and generate two new individuals in the crossover process (the crossover probability Pc) by

$$\begin{cases} K_j^m(g) = K_j^m(g) + \sigma_1(K_j^m(g) - K_j^n(g)) \\ K_j^n(g) = K_j^n(g) + \sigma_2(K_j^m(g) - K_j^n(g)) \end{cases},\ j = 1,2,3 \quad (9)$$

where σ_1 is an uniformly distributed random number in [0,1].

Step 8: Generate a new individual in the mutation process (the mutation probability Pm) for each individual by:

$$K_j^i(g) = K_j^i(g) + (\sigma_2 - 0.5).s_j, \text{ with } j = 1,2,3 \quad (10)$$

where s_j (K_j^{max}- K_j^{min}) is a range value for the searching range Kj $\epsilon K_j^{max}, K_j^{min}$] of the j-th searching parameter Kj. σ_2 is an uniformly distributed random number in [0,1].

Step 9: Bound each updated parameter K_j^i in its searchingrange by:

$$K_j^i(g) = \begin{cases} K_j^{max} & \text{if} K_j^i(g) > K_j^{max} \\ K_j^i(g) & \text{if} K_j^{min} \le K_j^i(g) \le K_j^{max} \\ K_j^{min} & \text{if} K_j^i(g) < K_j^{min} \end{cases} \quad \text{with} \quad i = 1,2,\ldots,\ L,\ \ j = 1,2,3 \quad (11)$$

Step 10: Let g = g + 1 and go to Step 3.

Step 11: Determine the selected controller by the proposed method based on the obtained parameter set Kbest with the best fitness fbest.

In a synchronous generator, the terminal voltage is maintained constant at various levels by using an AVR. Therefore, the outer loop acted as a self-tuning PID voltage controller based on RBF NN to offer the adaptive capability for uncertain load and system conditions. Following Moody and Darken [38], the FF two-layered NN RBF with one single hidden layer is considered. The PID parameters (Kp, Kd, Ki) are automatically readjusted by on-line learning algorithm of RBF to keep the system error $e(k) = 0$. The RBF-NN algorithm is achieved through the following steps:

Step 1:Initialize RBF-NN by setting the xite=0.25;alfa=0.05;belte=0.01;x=[0,0,0]';
ci=30*ones(3,6);bi=40*ones(6,1);w=10*ones(6,1); h=[0,0,0,0,0,0]';
ci_1=ci;ci_3=ci_1;ci_2=ci_1;bi_1=bi;bi_2=bi_1;bi_3=bi_2;w_1=w;w_2=w_1;w_3=w_1;
u_1=0;y_1=0;xc=[0,0,0]';error_1=0;error_2=0;error=0;
%kp=rand(1);%ki=rand(1); %kd=rand(1); (12)

Step 2: Set the input data of PID parameter from step 11 after impalement data to math lab Simulink for each combination of Kg and tg, applied to obtain the optimal values of K_p , K_d and K_i if dot get the output try again kp0= Kbest ; ki0= Kbest ; kd0= Kbest; kp_1=kp0;kd_1=kd0;ki_1=ki0; xitekp=0.20;xitekd=0.20;
xiteki=0.20; ts=0.001; (13)

Step 3: Set the k=1:1:2000 time(k)=k*ts; rin(k)=1.0*sign(sin(2*pi*k*ts)); yout(k)=(-0.1*y_1+u_1)/(1+y_1^2); (Nonlinear system) Set j=1:1:6 h(j)=exp(-norm(x-ci(:,j))^2/(2*bi(j)*bi(j))); (14)
Step 4: Calculate ymout(k)=w'*h; d_w=0*w; set j=1:1:6 for d_w(j)=xite*(yout(k)-ymout(k))*h(j); (15)
Step 5: Calculate the w=w_1+d_w+alfa*(w_1-w_2)+belte*(w_2-w_3);
d_bi=0*bi; by set the j=1:1:6 calculate
d_bi(j)=xite*(yout(k)-ymout(k))*w(j)*h(j)*(bi(j)^-3)*norm(x-ci(:,j))^2; (16)
Step 6: bi=bi_1+ d_bi+alfa*(bi_1-bi_2)+belte*(bi_2-bi_3); Set the j=1:1:6 and i=1:1:3 to evaluate
d_ci(i,j)=xite*(yout(k)-ymout(k))*w(j)*h(j)*(x(i)-ci(i,j))*(bi(j)^-2); (17)
Step 7: After the above instructions done ci=ci_1+d_ci+alfa*(ci_1-ci_2)+belte*(ci_2-ci_3);
the Jacobean set are yu=0; j=1:1:6 yu=yu+w(j)*h(j)*(-x(1)+ci(1,j))/bi(j)^2; after done do dyout(k)=yu;step 8:start of Control system by calculate PID controller parameters by set
error(k)=rin(k)-yout(k); kp(k)=kp_1+xitekp*error(k)*dyout(k)*xc(1);
kd(k)=kd_1+xitekd*error(k)*dyout(k)*xc(2); ki(k)=ki_1+xiteki*error(k)*dyout(k)*xc(3);
chick the error if kp(k)<0, if kd(k)<0, ki(k)<0 and set kp(k)=0;kd(k)=0; ki(k)=0; to avoid it step 9: set M=0; switch M , case 0 case 1 Only PID Control parameters kp(k)=kp0; ki(k)=ki0; kd(k)=kd0; go to step 2 (18)
Step 8: Calculate
du(k)=kp(k)*xc(1)+kd(k)*xc(2)+ki(k)*xc(3);
u(k)=u_1+du(k);Return of parameters
x(1)=du(k); x(2)=yout(k); x(3)=y_1; u_1=u(k);
y_1=yout(k); ci_3=ci_2; ci_2=ci_1; ci_1=ci;
bi_3=bi_2; bi_2=bi_1; bi_1=bi;
w_3=w_2; w_2=w_1; w_1=w; (19)
Step 9: Calculating P xc(1)=error(k)-error_1;
Calculating D xc(2)=error(k)-2*error_1+error_2;
Calculating I xc(3)=error(k);
Calculate errors error_2=error_1; error_1=error (k);
kp_1=kp(k); kd_1=kd(k); ki_1=ki(k);
Go to step 2 (20)
Step 10: Determine the selected controller based on the input parameter obtained from RBF-NN
kp_1=kp(k); kd_1=kd(k); ki_1=ki(k); (21)
The RGA algorithm only require the information regarding the fitness function value of each parameter set. These two algorithms (RGA and RBF-NN) are applied to obtain a good PID control parameter set for AVR system. For each combination generation parameters of Kg and tg, the proposed RBF-NN tuning by RGA is applied to obtain the optimal values of Kp, Kd and Ki .Following [39] the generator exciation paramters are adopted. The SF rule table is formulated for Kp, Kd and Ki using the above approach. Table 2 as (a), (b) and (c) enlists these parameter as used to design GPFPID voltage controller.

3. PID Controller Design for AVR System

For the stable electric power service it is essential to develop a highly efficient and speedy AVR of the synchronous generator. So far, the analog PID inflammatory disease controller (PIDIDC) is mostly used for the AVR due to its simplicity and economy. However, the parameters of PIDIDCs cannot be tuned easily. Gaing [37] proposed a technique to determine such parameters by using a particle swarm optimisation (PSO) algorithm. The AVR system model controlled by the PID controller is depicted in Fig. 1&2 with Vs the output voltage of sensor model, Ve the error voltage between Vs and the reference input voltage Vref (S), VR the amplified voltage by the electronic equipment model (amplifier model), VF the output voltage by exciter model, and Vt the output voltage by the generator. There are 5 models for the AVR system such as PID controller, electronic equipment (amplifier model), exciter, generator and (e) sensor. The transfer function of each model is represented by:

(i) PID controller model

$$G_s(s)=K_p+K_d\ s+\frac{K_i}{s} \quad (22)$$

where Kp, Kd, and Ki are the proportion coefficient, differential coefficient, and integral coefficient, respectively.

(ii) Amplifier model

$$\frac{V_R(s)}{V_e(s)} = \frac{K_A}{1+\tau_A S} \tag{23}$$

where KA is a gain and τ_A is a time constant.

(iii) Exciter model

$$\frac{V_F(s)}{V_R(s)} = \frac{K_E}{1+\tau_E S} \tag{24}$$

where KE is a gain and τ_E is a time constant.

(iv) Generator model

$$\frac{V_t(s)}{V_F(s)} = \frac{K_G}{1+\tau_G S} \tag{25}$$

where KG is a gain and τ_G is a time constant.

(v) Sensor model

$$\frac{V_S(s)}{V_t(s)} = \frac{K_R}{1+\tau_R S} \tag{26}$$

where KR is a gain and τ_R is a time constant.

The transfer functions of the AVR components are depicted in Table 1.

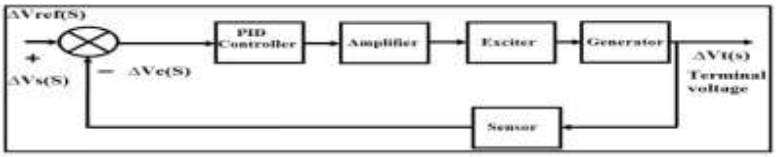


Fig. 1. Typical AVR system model controlled by the PID controller

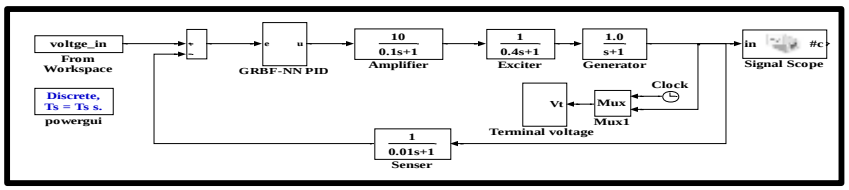


Fig. 2. MATLAB-Simulink model of AVR system with PID controller

Table 1. The transfer functions of AVR components.

Components	Transfer function	Parameter range
Amplifier	$TF_{amplifier} = K_a/1+\tau_a s$	$10 < K_a < 40$ $0.02s < \tau_a < 1s$
Exciter	$TF_{exciter} = K_e/1+\tau_e s$	$1 < K_e < 10$ $0.4s < \tau_e < 1s$
Generator	$TF_{generator} = K_g/1+\tau_g s$	K_g depend on the load (0.7-1.0) $1s < \tau_g < 2s$
Sensor	$TF_{sensor} = K_s/1+\tau_s s$	$0.001s < \tau_s < 0.06s$

The best choice of PID controller parameters is essential for the acceptable operation of the system. Thus, the problem of PID controller parameter selection is devised as an optimization problem wherein the objective function is given by:

$$MinF(\mathrm{K_d}, K_p, K_i) = (1 - e^{-\beta})(O_{sh} + E_{ss}) + e^{-\beta}(t_s - t_r) \tag{27}$$

where $MinF(\mathrm{K_d}, K_p, K_i)$ uses a combination of transient response counting rise time overshoot, settling time and steady-state error. By choosing the suitable value of the weighting factor β, the presentation principle can be made to please the designer requirements. The above optimization problem is subjected to the following restrictions.

$$K_p^{\min} \le K_p \le K_p^{\max},\ K_i^{\min} \le K_i \le K_i^{\max},\ K_d^{\min} \le K_d \le K_d^{\max} \tag{28}$$

Here, RGA is applied to the above optimization problem to search for the optimum value of the controller parameters.The detail of proposed RGA and RBF-NN are depicted hereunder.

3.1 GA Implantation

The optimal PID controller parameters are achieved by applying RGA which involved two major aspects:

i) Variables repersentation and
ii) Fitness function

3.2 Variable Representation

For PID controller tuning, candidate solutions in the genetic population are represented. The elements of the solution consisted of the variables such as the integral gain (K_i), proportional gain (K_p) and derivative gain (K_d) which are reprented by point number in the proposed GA population. Using this repersentation, an individual in the prposed GA is computed to achieve the optimal PID gain. Interestingly, the computer memory requirement to strore the population is remarkably reduced due to the direct representation of the solution variables. The value of parameter set $(\mathrm{K_p}, K_i, K_d)$ that is obtained from RGA is directly tuned into the RBF-NN to obtain the optamim value of PID controller parameter $(\mathrm{K_p}, K_i, K_d)$. This optimum value is essential to design FPID controller for the thematic factory operation of AVR system.

3.3 Fitness Function

It is defined as the non-negative Fig. of value to be maximized so that the performance of each individual in the population can be evaluated. The fitness function is the mutuality of the presentation criterion $F(\mathrm{K_d}, K_p, K_i)$ given in equation (27). Hence, the minimization of performance criteria is transformed to a fitness function to be maximized as,

$$Fitness = \frac{k}{F(\mathrm{K_d}, K_p, K_i) * ITAE} \tag{29}$$

where k a is constant, ITAE is an integral of time multiplied by the absolute error value. This is used to amplify 1/F (usually small) so that the fitness value of the chromosome occurs in a wider range.

4. Sugeno Fuzzy Model

Devaraj *et al.* [10] used a fuzzy set theory, wherein a variable was a member of one or more sets with a membership specified degree. The fuzzy rule was expressed as:

If x is A and y is B then z = f(x,y) (30)

where *A* and *B* are the fuzzy sets in the antecedent, *x* and *y* are input variables and *f (x, y)* is a crisp function in the consequent. Each variable in the fuzzy set is represented by suitable membership functions. The core of the fuzzy logic system is formed by a set of such rules. For an exact input signal condition, the fuzzy system defined the rules to be fired and then calculated the efficient output in two steps. First, the minimum of the membership functions input (w_i) was obtained for each rule, where this value is regarded as the firing value for a particular rule. Second, the overall output was calculated by a weighted average of individual rule outputs given by:

$$z = \frac{\sum_{i=1}^{M} \omega_i z_i}{\sum_{i=1}^{M} \omega_i} \tag{31}$$

The PID controller parameters under various operating conditions were determined by the Sugeno fuzzy system.

5. Design Of Fuzzy PID Controller Using Generator Parameters (Kg And Tg)

The optimum PID parameters for real-time operation are obtained by developing SFL model, where the parameters for the inputs are K_g and t_g and outputs are K_p, K_d and K_i. Eight fuzzy sets including 'very low (VL)','low(L) ,'medium low (ML)', 'medium(M)', 'medium high (MH)' ,'high low (HL)', 'high medium (HM)' and 'high (H)' are defined for the variable K_g . Likewise, the fuzzy sets defined for the variable τ_g are 'very low (VL)', 'low (L)', 'medium low (ML)', 'medium high (MH)', 'high (H)' and 'very high (VH)'. They are linked with overlapping triangular membership functions. To formulate the table for fuzzy rule, the values of K_g are varied from 0.7 to 1.0 in steps of 0.1 and τ_g are varied from 1 to 2 in steps of 0.2 as shown in Fig. 3 . For each combination of K_g and τ_g, the proposed RBF tuning via GA is applied to obtain the optimal values of K_p, K_d and K_i at each time. The table for the fuzzy rule for K_p, K_d and K_i is formulated (Table 2). During real-time operation, corresponding to the present operating condition, the values of K_g and t_g (in seconds) are determined. For this value of K_g and t_g, the optimal value of K_p, K_d and K_i is computed using the fuzzy rule table and the Sugeno inference system.The system with three fuzzy logic controllers (K_p, K_d and K_i) and rule viewer are set where each controller has two inputs (K_g, τ_g) and each input has fuzzy set associated with it.

The input range for K_g is $0.7 \leq K_g \leq 1$ and τ_g is $0.4 < \tau_g < 1$. The output of K_p, K_d and K_i depended on the variation of K_g with respect to τ_g through the output setting rule based on the table. The output has 72 fuzzy set rules for K_p, K_d and K_i, and 48 rules for each parameter as depicted in the form of surface and rule viewer (Fig. 4). For this value of K_g and t_g, the optimal value of K_p, K_d and K_i is computed using the fuzzy rule table and the Sugeno inference system. The best results are obtained with following control parameters of GA: number of generation = 50, population size = *30*, crossover = *0.6*, mutation probability = *0.001* and β = 1. The software for the proposed RBF and GA is written in MATLAB code and executed on a laptop Intel core(TM)2 Duo CPU 5550@1.83GHz and 3 GB. The GA took 23.79 s to reach the optimal solution.

Fig. 3. A thematic factory operation of AVR system with Kg = 0.7 and τg= 1

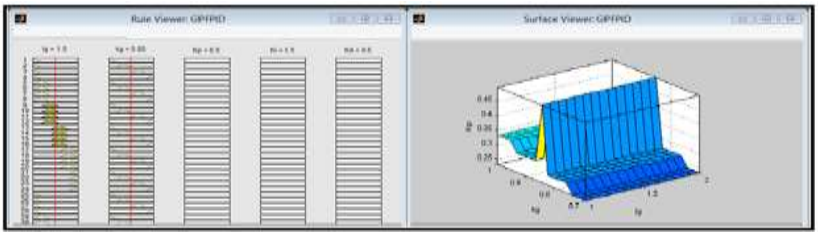


Fig. 4. Surface and rule viewer

Table 2. The fuzzy rule table formulated for Kp, Kd and Ki using the above approach

τ_g	**Very low**	**Low**	**Medium Low**	**Medium high**	**High**	**Very high**
K_g	1	1.2	1.4	1.6	1.8	2
		(a)	*For proportional gain (K_p)*			
Low (0.7)	0.6431	0.8664	0.6338	0.6217	0.8119	0.6638
Medium low (0.8)	0.5951	0.7111	0.83	0.768	0.7295	0.7601
Medium high (0.9)	0.4428	0.6446	0.6900	0.6983	0.7178	0.547
High (1.0)	0.4156	0.6135	0.64	0.5418	0.58	0.470
		(b)	*For integral gain (K_i)*			
Low (0.7)	0.4388	0.5349	0.5402	0.4969	0.5381	0.5410
Medium low (0.8)	0.3845	0.4855	0.5248	0.4527	0.5195	0.5266
Medium high (0.9)	0.3466	0.4372	0.4799	0.3980	0.4654	0.4826
High (1.0)	0.3247	0.4034	0.4379	0.4129	0.4237	0.4255
		(c)	*For derivative gain (K_d)*			
Low (0.7)	0.1223	0.1192	0.1113	0.2003	0.1059	0.1278
Medium low (0.8)	0.1855	0.1167	0.1201	0.1888	0.1189	0.1317
Medium high (0.9)	0.2220	0.2110	0.1058	0.1898	0.1481	0.1287
High (1.0)	0.1019	0.2192	0.1866	0.1754	0.1943	0.1017

6. Experimental Setup

Fig. 5. Shows the schematic representation of the experimental setup in terms of modified model. The test model consisted of STC ac synchronous alternator generator (three-phase, four-wire type, and adopting star connection with neutral point). The rated line voltage is *400 V*, phase voltage *230* V, frequency *50 Hz*. power factor *0.8 (lag)*, synchronous generator (400 V 3-phase 30 KVA, PF 0.8, 45.6 A, and *1500 RPM)* coupling with par mover internal composition gas engine *Kia motors 2701*. The control circuit consisted of PLC type SIMATIC IPC427C, accurate with gearbox, convertor, speed sensor, voltage sensor, AVR (*400 V and 10 A*). Measurement devices are the Voltmeter, frequency meter, digital oscilloscope. The supplementer devices are the transient device,

load resistances, step down transformers *(220-6v, CB 3-phase 60 A* and two *DC* power supplies *(0-5,0-30* and *0-24).*

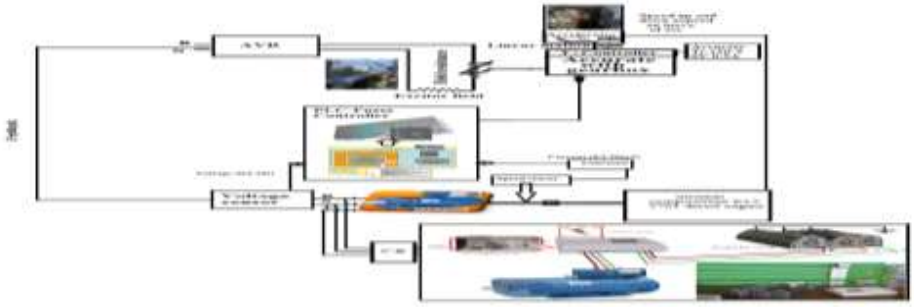


Fig. 5. The experimental setup

The data transfer in the proposed GPFPID controller from MATLAB to PLC involved the following:

6.1 GENERATION OF C/C++ WITH RTWEC

The RTWEC being an add-on of Mathworks generated C/C++ code from Simulink subsystems (opened the Simulink model) wherein the "Tools–Real-Time Workshop –Options" in the menu is clicked on. Then, a window is displayed with a navigation bar in which the entry "Real-Time Workshop" is selected by default. Next, the following parameters are entered: Navigation item "Real-Time Workshop" – System target file: ert.tlc – Language: C++ – Click the button "Set objectives" Shift "Traceability" and "Execution efficiency" with the "->" button to the right field. Navigation item "Code Placement" – File packaging format: Compact. Afterward, the right mouse-button is clicked on the subsystem "PID _GPFPID_disc". The "Real-Time Workshop Build Subsystem" is selected in the context menu. In the window menu of "Build code for Subsystem" the "Build" button is clicked on. After the creation of the C/C++ code, the window "Build Code for Subsystem" is closed automatically. Later, the generated code is located in the directory of the Simulink model"...\ PID_GPFPID_disc_rtw" as illustrated in Fig. 6.

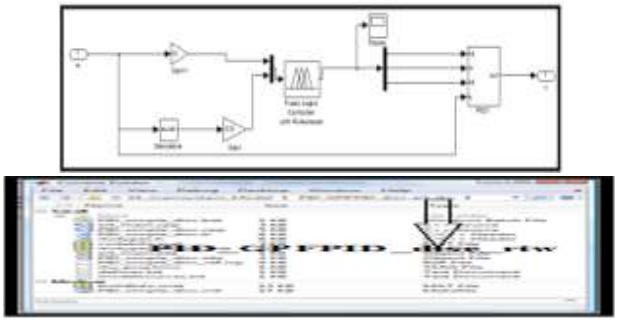


Fig. 6. Generation C/C++ using RTWEC.

6.2 WINAC SIMULINK TO ODK (S2O) WIZARD

The WinAC S2O Wizard automatically generated all required blocks and files for the integration of the Simulink subsystem into a STEP 7 project. An SCL source and a DLL or RTDLL file is created out of the generated C/C++ code of the RTWEC. With STEP 7 V5.x programs, integration became feasible due to the presence of WinAC S2O Wizard. Finally, the STEP 7 program and the DLL or the RTLL is loaded to a PC system with WinAC RTX. Meanwhile, the WinAC ODK Library is integrated in the SIMATIC Manager because of the requirement of SFB65001 (CREA_COM) and SFB65002 (EXEC_COM) blocks for the execution of the DLL/RTDLL.

6.3 CONFIGURATION WITH SIMATIC MANAGER STEP 7 V5.5

The folder "… \GPFPID _STEP7_V5x_Project" contained the project for STEP7 V5.5 with the following items: WinAC RTX is only the hardware configuration of a PC station with a WinAC RTX where the program is empty. This configuration just served as a template. PID _GPFPID provided a prepared program for integration through the WinAC S2O Wizard. The following blocks are created: OB35 (CYC_INT5) as cyclic OB with 100 ms cycle. The simulated controlled system and the PID controller (PID _GPFPID_disc) are called in this block. DB35 (Data) as a global data block. It contained all the required variables such as - Setpoint [Real], and - PID_output [Real], Process_value [Real], - crea_status [Word] (provides the status via CREA_COM), - exec_status[Word] (provides the status via EXEC_COM) - initialize [Bool]. FB100 (PROC_C) with Instanz-DB100 Simulated PT3 process (parameterized like the process in MATLAB/Simulink). SFB65001 (CREA_COM) for initialization of the DLL/RTDLL file. SFB65002 (EXEXC_COM) for execution of the DLL/RTDLL file.VAT_1 Variable table with the variables of DB35 (DATA) PID _GPFPID _dll_final enclosed a complete program with integration of the PID controller from MATLAB/Simulink through the WinAC S2O Wizard. A DLL call is used in this program.PID _GPFPID _rtdll_final contained a complete program with integration of the GPFPID controller from MATLAB/Simulink through the WinAC S2O Wizard. An RTDLL call is used in this program. Fig. 7. displays the complete program of "PID_ GPFPID" inclusion with the "PID_ GPFPID _disc" block from the WinAC S2O Wizard.

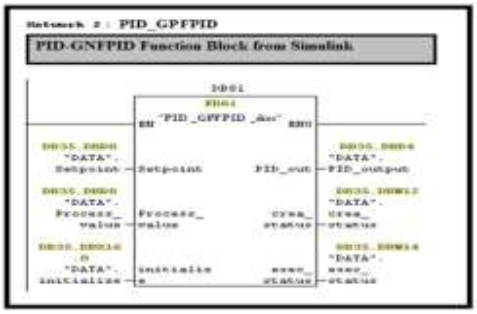


Fig. 7. The "FB PID_ GPFPID_disc" insertion into the "Network 2: PID Controller

6.4 OPEN THE SIMATIC MANAGER

The example project "S2O_PID_GPFPID" is opened in the directory "...\ GPFPID _STEP7_V5x_Project". During the execution of the WinAC SO2 Wizard, the SCL source "PID_ GPFPID _disc" is integrated already and the function block "PID_ GPFPID_disc" is created. The OB 35 is opened and the "FB PID_ GPFPID _disc" is inserted into the "Network 2: PID Controller". The inputs and outputs are interconnected as shown in the Fig. 7. The programis loaded to the WinAC RTX. The DLL (C_ODK.dll) or RTDLL (C_ODK.rtdll) is copied into the directory C:\ of the PC system with the WinAC RTX. Finally, the RTDLL files are registered and WinAC is set into the RUN mode.

6.5 IMPLEMENTATION

After loading the program to the WinAC RTX, the following steps are executed for commissioning. The "VAT_1" variable table is opened, the icon "Monitor Variable" is clicked on, the variable "initialize" is chanced to "TRUE" and the icon "Modify Variable" is clicked on. The value of the variable "initialize" is changed to "FALSE" and the icon "Modify Variable" is clicked. The value of the variable "Set point" is changed to a desired value say 50, and the icon "Modify Variable" was clicked on. Due to the change of the set point, the GPFPID controller produced an output value. The process is changed accordingly until it corresponded to the "Set point". The entire execution is shown as screen-shot in Fig. 8.

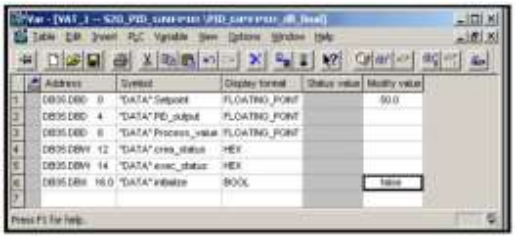


Fig. 8. Illustration of the VAT_1 variable table.

6.6 GENERATOR AND GOVERNOR CONTROL SYSTEM TEST AND RESULTES

As aforementioned, the PLC (GPFPID-PLC) controller and Industrial speed (KS 92) controller are used to develop the AVR and Governor system. This algorithm searched a high-quality solution effectively and provided fully controlled system. Besides, the PLC (GPFPID-PLC) controller gathered the data of control signal (0-10 V) from voltage sensor that connected between any two lines (R and S) of the generator terminal voltage. It allowed sending the control signal (0-10 V) depending on the voltage difference between line to line (R-S or S-T or R-T) generator terminal to GPFPID-PLC controller. This in turn ensured that the value of field excitation resistance (excitation current) in the gearbox (gearbox to make the motion of change variable resistance value smooth and accuracy) is varied accurately. Conversely, the speed data control signal (4-20 mA) appeared from the speed sensor that is fixed on the generator shaft coupled with the shaft of the internal composition KIA 2701 diesel engine. The Industrial controller KS 92 processed to maintain the constant speed (1500 rpm) accurately through the gearbox. The relation between the speed of generator rotor and excitation current is established to get the constant terminal voltage and frequency as shown in Fig. 9. The experiment is performed using SIMATIC IPC427C (MICROBOX PC) hardware system where S7 program is loaded with the desired configuration mentioned earlier.

Fig. 9: Three AVR system without and with GPFPID-PLC controller.

Fig. 9 depicts three types of AVR used in the experiment for comparison. These include AVR 1 with GPFPID-PLC controller as well as the normal AVR2 and AVR3 without GPFPID-PLC controller. The performance of fuzzy PID controller and industrial KS 92 controller is evaluated by implementing on a practical collecting main speed and voltage system. Following our earlier design [39], the fault device is used. The generator terminal voltage collection is controlled via the fuzzy PID control algorithm and the flow of fuel by the Industrial controller KS 92. The terminal voltage and frequency (speed) are adjusted independently. Fig. 10 represents the sensor current dependent terminal voltage, frequency, and speed of the proposed GPFPID. This hybrid control system combined AVR1 and adjusted the GPFPID-PLC controller. The speed governor is adjusted by Industrial PID controller to keep the terminal voltage andfrequncy under control. The AVR1 system under unsymmetrical disturbance (single line, line to line, and 3 lines to ground fault) is applied at the generator termainal and response of system is monitored.

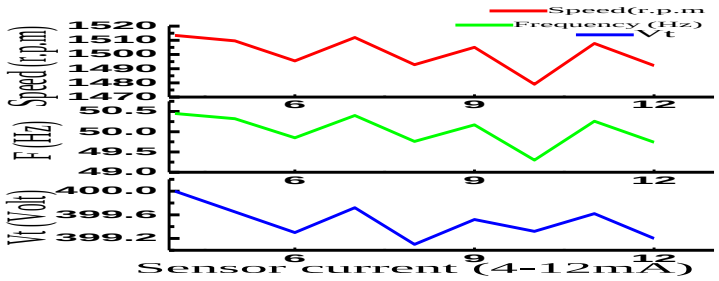


Fig. 10: The related of terminal voltage, frequency and speed with AVR1 sensor controlled GPFPID.

Fig. 11-14 illustrate the system rsponse for above contingency with GPFPID-PLC. It is observed that the GPFPID-PLC with AVR1 is able to suppress the oscillation in the terminal voltage and revealed good damping characristics compared to AVR2, and AVR3 without controller (Fig. 15). Fig. 11-14cleary displayed that AVR1 is adjusted by the GPFPID-PLC controller and speed governor is tuned by the Industrial PID controller to keep the termanal voltage and frequncy under control.

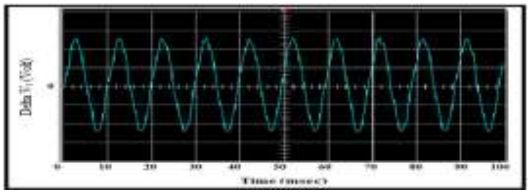


Fig. 11: Voltage control curve for AVR1 adjusting with GPFPID.

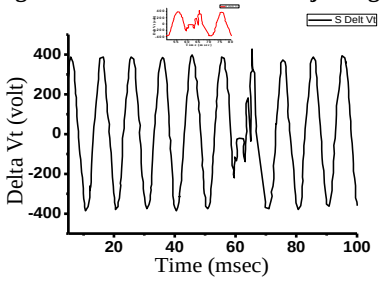


Fig. 12: Voltage control curve for a single line fault.

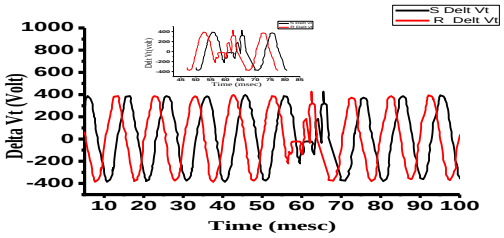


Fig. 13. Voltage control curve for 2 line to ground fault

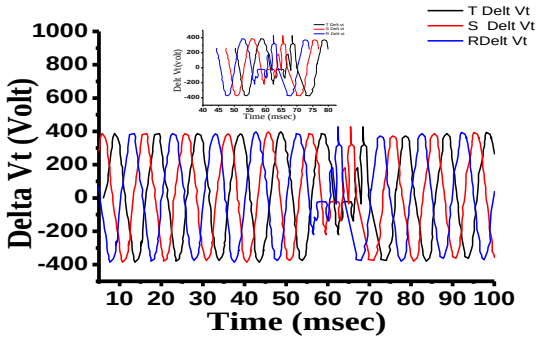


Fig. 14: Voltage control curve for a severe disturbance fault.

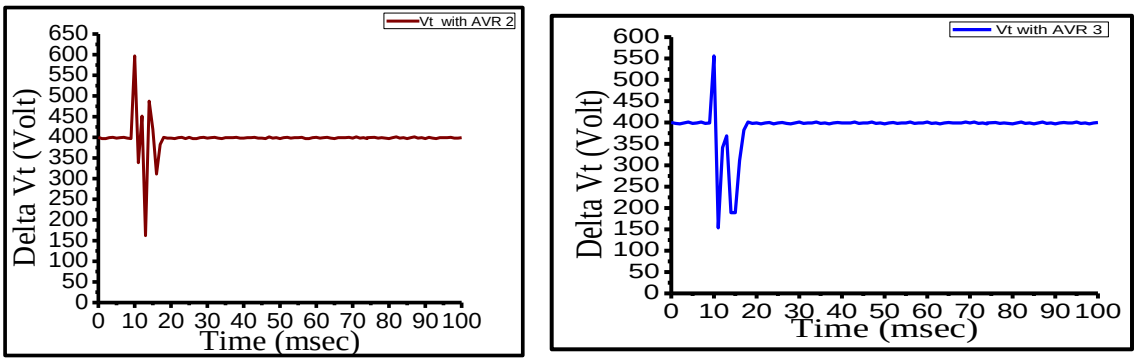


Fig. 15: Voltage control curve for a disturbance with AVR2 and AVR3.

It is clear from Fig. 15 that the AVR2 and AVR3 under severe disturbance fault exhibited the voltge swing between 600-150 V and 550 -152 V, respectively. Thus, AVR1 is demonstrated to be better in terms of voltge swing (200-410 V) as shown in Fig. 16.

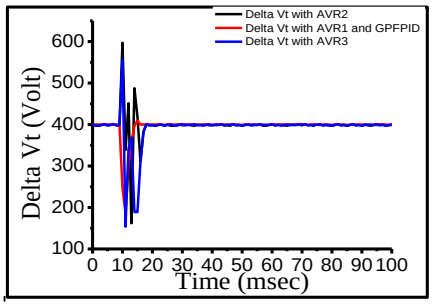


Fig. 16: Voltage response comparison under severe disturbance of AVR1, AVR2 and AVR3.

Fig. 17 clearly reveals that the AVR2, and AVR3 under severe disturbance produced the frequncy swing between 57 – 40 Hz and 56 – 40 Hz, respectively. Furthermore, AVR1 revealed better frequency siwng (47-55 Hz) than AVR2 and AVR3.

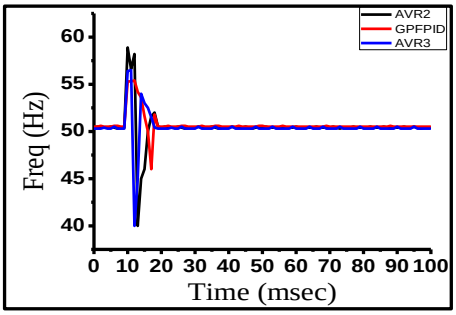


Fig. 17. Frequency response comparison under severe disturbance of AVR1, AVR2 and AVR3.

The experimental results (Fig. 10-17) clearly demonstrate that the developed fuzzy PID controller is able to obtain a desired control effect even with the unbalanced fault. Thus, it fulfilled the requirement of collecting main voltage and grid frequency control under the industrial conditions. Furthermore, the performance of the newly proposed GPFPID is remarkably improved in terms of the desired voltage and speed range attainment. This finding is attributed to the use of hybrid control where AVR and Gg (GPFPID-PLC) controller is combined and the Industrial speed governor is exploited to provide good control performance in diverse operating conditions.

6.7 COMPARATIVE OF GPFPID WITH OTHER RELTED WORKS

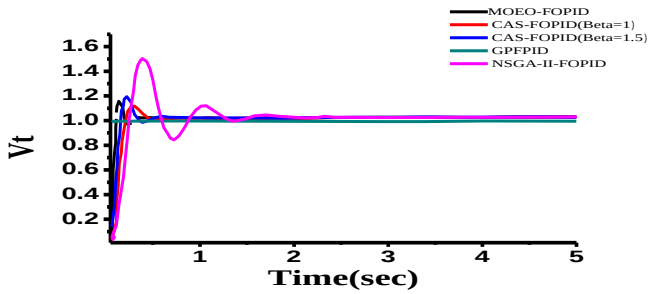


Fig. 18. Comparison with GPFPID data with other methods

The performance of the proposed GPFPID was compared (Table 3) with other related system such as NSGA-II-FOPID , GA-FOPID , PSO- FOPID, CAS-FOPID and competitive single-objective evolutionary algorithms-based FOPID controllers. The terminal voltage step response of AVR system is presented in Fig. 18. The proposed GPFPID was found to be more robust and better than these reported NSGA-II-FOPID,GA-FOPID [32], PSO- FOPID [32], CAS-FOPID with β¼=1 and β¼=1.5 [2, 32] under the uncertainty of amplifier model parameters. However, from the perspective of engineering design and system operation the performance of GPFPID was accepted by engineers under the variation of parameters KA and τA are enlisted in Table 1. Fig. 18 demonstrated the terminal voltages step response for AVR system with MOEO-FOPID controller when KA was varied from K_g depending on the load (0.7-1.0), and τA (1.0 to 2.0). Clearly, as the value of parameter KA was increased, the overshoot (Mp) was also increased but the rising time (tr) and settling time (ts) became shorter; and the steady-state error (Ess) appeared smaller. A comparison with other calculation revealed that the proposed GPFPID controller achieved minimal settling time, short rise time, and overshoot (0.001, 0.0008, 0). In short, the developed GPFPID controller can be viewed as robust for the uncertainty of amplifier model parameter swith in the range defined (Table 1).

Table 3 Comparative performance of different evolutionary algorithms

Algorithm Mp (%)	Mp (%)	tr (sec)	ts (sec)	Ess
GA-FOPID [14]	17.27371	0.12	0.26	4.05E-04
PSO-FOPID [14]	10.46691	0.12	0.24	4.63E-04
CAS-FOPID(β¼1) [14]	9.079521	0.17	0.32	2.35E-04
CAS-FOPID(β¼1.5) [14]	8.941498	0.17	0.36	1.89E-04
NSGA-II-FOPID [30]	46.71605	0.2	1.12	6.92E-0
MOEO-FOPID [1]	14.07733	0.07	0.23	2.98E-05
Proposed Method GPFPID	0	0.0008	0.001	0.0000
Parameterization sets	The optimum PID parameters for real-time operation are obtained by developing Surgeon fuzzy logic model. Here, Kg and τg are the inputs and Kp, Kd and Ki are the outputs. Eight and six fuzzy sets are defined for the variable Kg and τg, respectively.			
Simulation results	The response of the system with new fuzzy PID controller is highly sensitive to a very small change (0.005). The proposed methodology achieves good performance in the rise time, peak overshoot and the steady state error. The novel GPFPID controller improved the transient response by minimizing the swing in terminal voltage between the up and down swing wave and kept the system stable as reported by Devaraj, Ching-Chang, M.S. Rahimian and N. Madinehi [10, 40-42]			
Standardize deviation Steady Stat Error [201]	(400-399)/400= 0.0025			

7. Conclusion

We combined GA, RBF-NN and SFL to achieve the optimal PID controller parameters in AVR system with industrial speed governor (ISG). In the proposed system, RBF-NN was used to enhance the PID parameters that obtained from GA to design Sugeno fuzzy PID controller tuned by generator parameter (Kg, tg). In this hybrid controller the AVR control system was united with governor speed control system of prime mover that relied on the grid frequency. The developed algorithm provided a high-quality solution effectively and offered full control of voltage and frequency of proposed system compared with other existing art of the techniques. The suitability of the proposed GPFPID and ISG was demonstrated through experiments, where PLC controller and industrial KS 92 controller was utilized. The improvement of the collecting voltage-frequency control performance of proposed system was attributed to the inherent interacting effect in the collecting control loops. Furthermore, the method was based on the oscillation. This led to the coupling as reflected by the minimal error on the control parameters mostly in the collecting main voltage and grid frequency system. The GPFPID with ISG KS 92 PID controller revealed combinative structure in the presence of a conventional basic PID and a fuzzy compensating controller and industrial speed controller. It is established that this GPFPID-KS 92 PID compensating controller can conFig. the control signal based on interacting effect. Experimental results demonstrated the improvement in attaining to the desired voltage range using the proposed fuzzy PID controller and frequency to provide good control performance in difference operating conditions. An excellent voltage and frequency control performance is achieved upon testing the proposed hybrid controller on a practical AVR and governor system in synchronous generator for improving the transient response.

Acknowledgements

Abdullah is thankful to Dr. S.K. Ghoshal (UTM) for many valuable suggestions and critical readings of the manuscript. The authors wish to thank to UTM and Ministry of Higher Education for the financial support through GUP/RU grants No. 13H50 and 17H19.

References

[1] G.-Q. Zeng, J. Chen, Y.-X. Dai, L.-M. Li, C.-W. Zheng, M.-R. Chen, Design of generator parameters PID controller for automatic regulator voltage system based on multi-objective extremal optimization, Neurocomputing, 160 (2015) 173-184.
[2] I. Pan, S. Das, Frequency domain design of generator parameters PID controller for AVR system using chaotic multi-objective optimization, International Journal of Electrical Power & Energy Systems, 51 (2013) 106-118.
[3] Z. Chen, X. Yuan, B. Ji, P. Wang, H. Tian, Design of a generator parameters PID controller for hydraulic turbine regulating system using chaotic non-dominated sorting genetic algorithm II, Energy Conversion and Management, 84 (2014) 390-404.
[4] I. Pan, S. Das, Fractional-order load-frequency control of interconnected power systems using chaotic multi-objective optimization, Applied Soft Computing, 29 (2015) 328-344.
[5] C. Yeroğlu, A. Ateş, A stochastic multi-parameters divergence method for online auto-tuning of generator parameters PID controllers, Journal of the Franklin Institute, 351 (2014) 2411-2429.
[6] L.B. Prasad, H.O. Gupta, B. Tyagi, Application of policy iteration technique based adaptive optimal control design for automatic voltage regulator of power system, International Journal of Electrical Power & Energy Systems, 63 (2014) 940-949.
[7] A. Rajasekhar, R.K. Jatoth, A. Abraham, Design of intelligent PID/PI λ D μ speed controller for chopper fed DC motor drive using opposition based artificial bee colony algorithm, Engineering Applications of Artificial Intelligence, 29 (2014) 13-32.
[8] A. Chatterjee, S. Ghoshal, V. Mukherjee, Chaotic ant swarm optimization for fuzzy-based tuning of power system stabilizer, International Journal of Electrical Power & Energy Systems, 33 (2011) 657-672.
[9] I. Pan, S. Das, Intelligent generator parameters systems and control: an introduction, Springer, 2012.
[10] D. Devaraj, B. Selvabala, Real-coded genetic algorithm and fuzzy logic approach for real-time tuning of proportional-integral - derivative controller in automatic voltage regulator system, Generation, Transmission & Distribution, IET, 3 (2009) 641-649.
[11] a.H.Z. J. Zhang, algorithm of vehicle stability adaptive PID control with single neuron network was proposed, IEEE Int., Auto. Robotics,, (2010) 978-971-4244-5194-4244/4241, 2010.
[12] W.M. L. Kun, and Z. Jianmin, An optimal PID controller using radical basis function (RBF) for the so-called direct-drive permanent magnet linear synchronous motor, IEEE 3 (2009) 978-971-4244-4507.
[13] J.M.a.C. Darken, Learning with Localized Receptive Fields, Proceedings of the 1988 Connectionist Models Summer School, (1988) 133-143.
[14] S.M.S. S.N. Qasem, Memetic Elitist Pareto Differential Evolution algorithm based Radial Basis Function Networks for classification problems, Applied Soft Computing, 11 (2011) 5565–5581.
[15] M.A.M.P. N. A. Al-geelani, R. Q. Shaddad Characterization of acoustic signals due to surface discharges on H.V. glass insulators using wavelet radial basis function neural networks Applied Soft Computing, 7 (2012) 1327-1338.
[16] L. Yao-Lun, T. Chia-Chang, J. Wu-Shun, L. Shuen-Jeng, Design an Intelligent Neural-Fuzzy Controller for Hybrid Motorcycle, in: Fuzzy Information Processing Society, 2007. NAFIPS '07. Annual Meeting of the North American, 2007, pp. 283-288.
[17] M.G. Simoes, B.K. Bose, R.J. Spiegel, Fuzzy logic based intelligent control of a variable speed cage machine wind generation system, IEEE Transactions on Power Electronics, 12 (1997) 87-95.
[18] J. Jiang, Design of an optimal robust governor for hydraulic turbine generating units, IEEE transactions on energy conversion, 10 (1995) 188-194.
[19] J.M. Guerrero, M. Chandorkar, T.-L. Lee, P.C. Loh, Advanced control architectures for intelligent microgrids, part I: decentralized and hierarchical control, IEEE Transactions on Industrial Electronics, 60 (2013) 1254-1262.
[20] N. Higgins, V. Vyatkin, N.-K.C. Nair, K. Schwarz, Distributed power system automation with IEC 61850, IEC 61499, and intelligent control, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41 (2011) 81-92.
[21] A. Soundarrajan, S. Sumathi, Fuzzy-based intelligent controller for power generating systems, Journal of Vibration and Control, (2010) 1077546310371347.
[22] a.R.C.H. G.R.YU, ptimal PID speed control of brushless DC motors using LQR approach, IEEE Int. Conf.Syst., Man Cybern., , (2004) pp. 473–478.
[23] Y. Minglin, Realization of Fuzzy PID controller used in turbine speed control system with FPGA, in: Future Information Technology and Management Engineering (FITME), 2010 International Conference on, 2010, pp. 261-264.
[24] H. Su, G. Hao, P. Li, X. Luo, Feed Forward Fuzzy PID Controller for Common-Rail Pressure Control of Diesel Engine, in: Measuring Technology and Mechatronics Automation (ICMTMA), 2010 International Conference on, 2010, pp. 264-267.
[25] K. Sinthipsomboon, I. Hunsacharoonroj, J. Khedari, W. Pongaen, P. Pratumsuwan, A hybrid of fuzzy and fuzzy self-tuning PID controller for servo electro-hydraulic system, in: Industrial Electronics and Applications (ICIEA), 2011 6th IEEE Conference on, 2011, pp. 220-225.
[26] R. Arulmozhiyal, Design and Implementation of Fuzzy PID controller for BLDC motor using FPGA, in: Power Electronics, Drives and Energy Systems (PEDES), 2012 IEEE International Conference on, 2012, pp. 1-6.
[27] R. Arulmozhiyal, R. Kandiban, Design of Fuzzy PID controller for Brushless DC motor, in: Computer Communication and Informatics (ICCCI), 2012 International Conference on, 2012, pp. 1-7.
[28] W. Chen, M. Xing, K. Fang, A PLC-based fuzzy PID controller for pressure control in Coke-oven, in: Control Conference (CCC), 2012 31st Chinese, 2012, pp. 4754-4758.
[29] K. Jinwook, C. Oh-Kyu, J.S. Lee, Design and stability analysis of TSK-type full-scale fuzzy PID controllers, in: Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on, 2012, pp. 1-8.

[30] X. Man-chen, W. Ling-long, Intelligent fuzzy- PID temperature controller design of drying system, in: Information Management, Innovation Management and Industrial Engineering (ICIII), 2012 International Conference on, 2012, pp. 54-57.
[31] M. Zhang Xiao, Y. Long Shi, Simulation Study on Fuzzy PID Controller for DC Motor Based on DSP, in: Industrial Control and Electronics Engineering (ICICEE), 2012 International Conference on, 2012, pp. 1628-1631.
[32] I. Pan, S. Das, Chaotic multi-objective optimization based design of generator parameters PIλDμ controller in AVR system, International Journal of Electrical Power & Energy Systems, 43 (2012) 393-407.
[33] Y. Tang, M. Cui, C. Hua, L. Li, Y. Yang, Optimum design of generator parameters PI λ D μ controller for AVR system using chaotic ant swarm, Expert Systems with Applications, 39 (2012) 6887-6896.
[34] M. Cui, Low carbon dispatch of distribution network containing microgrid using chaotic ant swarm, in: Control Engineering and Communication Technology (ICCECT), 2012 International Conference on, IEEE, 2012, pp. 818-821.
[35] H. Ramezanian, S. Balochian, A. Zare, Design of optimal fractional-order PID controllers using particle swarm optimization algorithm for automatic voltage regulator (AVR) system, Journal of Control, Automation and Electrical Systems, 5 (2013) 601-611.
[36] S.-A.L.a.H.-Y.W. Ching-Chang Wong*, optimal PID Controller Design for AVR System

Tamkang Journal of Science and Engineering, 12 (2009) 259-270.
[37] G. Zwe-Lee, A particle swarm optimization approach for optimum design of PID controller in AVR system, Energy Conversion, IEEE Transactions on, 19 (2004) 384-391.
[38] C.-C.T. Yao-Lun Liu , Wu-Shun Jwo , Shuen -Jeng Lin Design an Intelligent Neural-Fuzzy Controller for Hybrid Motorcycle, . IEEE, 1 (2007) 4244-1214.
[39] A.J. Al Gizi, M. Mustafa, K.M. Al Zaidi, M.K. Al-Zaidi, Integrated PLC-fuzzy PID Simulink implemented AVR system, International Journal of Electrical Power & Energy Systems, 69 (2015) 313-326.
[40] S.-A.L.a.H.-Y.W. Ching-Chang Wong*, Optimal PID Controller Design for AVR System, Tamkang Journal of Science and Engineering, 12 (2009) 259-270.
[41] N. Madinehi, K. Shaloudegi, M. Abedi, H.A. Abyaneh, Optimum design of PID controller in AVR system using intelligent methods, in: PowerTech, 2011 IEEE Trondheim, 2011, pp. 1-6.
[42] M.S. Rahimian, K. Raahemifar, Optimal PID controller design for AVR system using particle swarm optimization algorithm, in: Electrical and Computer Engineering (CCECE), 2011 24th Canadian Conference on, 2011, pp. 000337-000340.

Data Bus for Computer Based Assessment via Micro Services

Mehmet Karakoç[1] and Melih Günay[2]
[1] Department of Computer Engineering, Alanya Hamdullah Emin Paşa University, Antalya / Turkey
[2] Department of Computer Engineering, Akdeniz University, Antalya / Turkey
[1] mehmetkarakoc@ahep.edu.tr ; [2] mgunay@akdeniz.edu.tr

***Abstract*—** In this study, we introduce a secure *data bus* for *test management*, *test delivery*, and *assessment* in computer based system. The developed *RESTful web services* software program includes the base settings, controllers, and views with a large number of APIs regarding the core *business logic* and relevant data retrieval (*selection*) queries. However, to keep data consistency no *data manipulation* (insert, update, and delete) action is done via this program while the current data is checked and/or verified before and after making manipulation on it when needed. During our implementations and experiments, we encountered with very good developer/tester and end-user experiences in both managing and applying tests with the following benefits: (i) separating the common *back-end* issues from the *test management* and *test delivery* web applications. (ii) dividing the *information services* into many small services in which each service is responsible for its own with the separation of concerns on entity/model level. (iii) providing RESTful APIs over HTTPs that were created based on the *micro services* architecture with developer independence, isolation, low complexity, and simple integration. (iv) testing APIs easily without requiring the use of any other application while maintaining scalability. (v) redirecting the *data bus* to either test or production database.

***Keywords*—** data bus, information services, RESTful web services, API, computer based assessment, and micro services

1. INTRODUCTION

Classical *paper-and-pencil* based testing systems include three main steps: (1) *exam preparation* for each evaluation criteria of any required/elective course given by an instructor, (2) *exam administration* for applying and grading the exams prepared, and (3) *exam storage* for keeping and maintaining the results obtained. The following is hereby needed: (i) secretarial time for the creation of exam items manually or in an auto-formatting with shuffling and grouping issues including photocopying per each exam, (ii) manual grading or optical mark recognition, and (iii) papers/reports regarding the exam. Since this hard and complex process takes a very long time, it is an alternative to develop a complete *computer based assessment* (CBA – or *computer based testing* (CBT)) system. However, in developing such a successful system, a large number of key factors become the case with *test base*, *test management*, *test delivery*, and *assessment*.

In applying a test via simple *user interfaces* (UIs) successfully, there are many critical operational procedures such as assessing the achievement based on scorings and specifying the passing/failing cases for test takers. For a perfect assessment with accurate statistics, a robust data collection mechanism becomes a mandatory. When a test is applied on computer, it is critical to obtain information from the stored data. Whereas *information retrieval* procedures (*information services*) are implemented once, it would be possible to focus on the implementation of the aforementioned issues.

Integrity may not be maintained when connected to a database with a different web interface. Therefore, it is a good approach in a flexible software architecture to implement a generic *data bus* system for *information services*. In this study, we introduce a secure *data bus* for *test management*, *test delivery*, and *assessment* in computer based system.

The remainder of this paper is organized as follows. In Section 2, the literature research done is presented. In Section 3, the details about the technology used for the implementation are given. In Section 4, the implemented *data bus* is introduced. The results obtained are presented in Section 5, and the final thoughts are given in Section 6.

2. LITERATURE REVIEW

Since CBA has many advantages such as security, cost and accuracy, it is becoming a main part of *electronic learning and assessment* systems in higher education institutions [1]. However, when CBA is deployed, a variety of aspects need

to be taken into consideration such as software quality, secure delivery, reliable network, maintenance, software development cost and *test delivery* [2]. In the rapidly emerging field of CBA, the design of assessment task is complex with preparing questions, testing, scoring, reporting and feedback issues including interactivity, the flow between items, and assessment specifications [3].

CBT refers to any test taken on any type of computer with the benefits such as embedding interactive media, applying frequently, reducing human-error associated with grading, and receiving rapid feedback [4]. CBT methods include many benefits such as rapid access to test results and feedback, ability to rescoring or adjusting answers on exams when needed, and availability of longitudinal data for long-term performance assessment with the potential of technology-related difficulties [5]. The advantage of using CBT is to take a CBT test motivates a test taker to take another CBT test when compared to *paper–pencil testing* that includes serious testing effect problem [6].

Nikou and Economides investigated the effect of three modes of assessment based on *paper-and-pencil*, computer-web and mobile devices on the learning motivation and achievement of students [7]. Ricketts and Wilks suggested that the mode of presentation of assessments can significantly affect student performance, and an appropriate screen design, in which questions are presented e.g. by displaying one question at a time without scrolling need, is perhaps the most important factor in online assessment [8].

Singh and Tiwari designed and implemented a secure *computer based online examination* system with the support of multi-language questions, displaying random question at a time, and solution to the security and cheating issues that may be deployed on either internet or intranet [9].

To use in online testing it requires instructors to invest additional time and effort to design high-quality exam items [10]. Instead of instructor-made *paper test*, this issue can be handled with *computer based test delivery* even if the initial steps may take a long time. In *exam preparation* with a large question bank, an instructor can easily create a test by selecting questions or a randomly selected test with an automated process [11]. When creating a test and for many other processes, CBA *meta-data* can be used to facilitate *item retrieval* [12].

To create, manage, apply, and assess tests, quizzes and surveys we had created a complete and flexible database schema [13] for question storage and retrieval with comprehensive question banks. The designed hierarchical structure was based on (1) question, (2) user, and (3) test, and consisted of question banks, modules, topics, question groups, questions and answers, as follows: (1) *question bank* includes a set of modules with its owner(s) in which each module includes a set of topics. Each *module topic* includes a set of *question groups* while each question is included by a *question group* with a set of answers. (2) *test user group* includes a set of *test users* with its manager. (3) *test* includes a set of *test formats* with the total number of questions to be asked, scorings for right/wrong answered and unanswered questions, maximum score, score threshold, and etc. Each *test format* includes a set of *test question formats* with the number of questions to be asked for the *test topic* selected, question type, difficulty, the number of answers to be provided, shuffling options for both questions and answers, option for including *question groups*, and etc.

Based on the database schema [13] created in MySQL on phpMyAdmin, we had developed an integrated Java based CBA system [14] with the *test base system* including the *test management* and *test delivery* web applications implemented by using *Spring Framework* technologies. After integrating it into the current *information system* and deploying on a *Tomcat 8* application server at *Akdeniz University*, any test with a set of questions from different topics could be created to apply for a number of *test user groups* with different *exam booklets* prepared via the shuffling algorithm developed. *Information retrieval* in such a schema becomes critical when more than one instructor work on the same question bank while both academic staff and students can make change on the same database. Therefore, for *information retrieval*, we implemented a *data bus* to our CBA system based on the *test base system* established onto the database schema.

3. METHODOLOGY

In this section, the details about the technology used to implement the *data bus* are given.

3.1. Object-relational mapping

ORM (*object-relational mapping*) is an approach making data exchange (*database actions*) by laying a bridge between relational database management system and OOP (*object-oriented programming*). This approach links the *object*

models designed for managing the data in database with the corresponding tables. In performing database queries, the methods of the OOP language used are mapped onto the queries by using the specifiers created based on the *object models*. Moreover, ORM tools provide solutions to many problems such as automatic transaction management, caching, clustering, concurrent batch processing, locking, pagination, and polymorphism. For ORM issues *Spring Framework* can be used.

3.2. Spring Framework and Spring Boot

Spring Framework is a Java platform providing an infrastructure support to developing Java applications. When developing an application, it comes with many mandatory dependencies (required JAR (*Java ARchive*) files and helper libraries). *Spring Boot* combines all of these standard dependencies within a simple library that can be incorporated into applications as a *Maven* dependency. *Spring Framework* configures properties based on a result JAR file with auto-configuration to use the default values for each configuration, or the configuration can be overridden as desired.

3.3. Maven and dependency context

Maven is a software project management tool that can be used to develop any type of Java project. With its help, *Spring Framework* easily manages all the dependencies in a simple POM (*project object model*) file used to configure *Maven* when developing a web application or a console application.

When class *A* needs class *B* within itself, class *A* then depends on class *B*, and the class *B* therefore has to be directly injected as a dependency into the class *A*. This is called as *dependency injection* and can be done via the *annotation* "@Autowired" in *Spring Framework*. Moreover, the instances of all the dependencies (*objects*) injected into a class are automatically created by the auto-wiring constructor while a new instance of the class, the object own, is created. When this action is done by standard getter-setter methods, the dependencies are then injected after creating a new instance of the class. Besides this, when an object *obj1* requires another object *obj2*, the *obj1* should get the *obj2* from an external source such as configuration file. This is called as *inversion of control* and can easily be managed by *Maven*.

3.4. MVC framework

MVC (*model–view–controller*) software architectural *design pattern* gives the opportunity of separating different aspects easily that are related to the input/data logic, *business logic* and output/UI logic when implementing the UIs of an application. This design creates a *loose-link* between three main interconnected parts with the help of developing applications in this manner while describing the points where each type of logic is to be located within the application. Input logic and output logic belong to controller and view, respectively, while model is the case for the implementation of *business logic* in the application. This separation helps to manage the complexity in the development of applications with the facility of focusing on one certain aspect of the implementation at a time for many developers.

Model includes the simple classes to represent the data required for *information retrieval* from database and *data manipulation* on the tables in database. View is the collection of the classes representing the application elements in UI that is created with the model data. View only displays information and does not include UI-related information. Controller includes the components that handle *user interaction* (e.g. buttons and display-boxes), works with model, and displays the processed state (response: e.g. a list of any objects) on UI. Controller takes the input/request (how information is to be accepted) from the user, passes the related values to model e.g. for a query statement to be created, and returns the output/response (how information is to be presented) to the user after querying the database by these values. User may see the model as JSP (*Java Server Page*) or JSON (*JavaScript (JS) object notation*) that is provided via controller used to establish communication between the classes connecting model and view.

In this design, (i) user triggers controller (writing on the current state) e.g. by clicking on a button, (ii) controller manipulates and processes model, (iii) model updates view, and (iv) view is shown to the user (reading the final state). Events such as change warnings are reflected to the user via controller. As a lightweight and testable *design pattern*, MVC associates UI with basic *data models* successfully in the implementations of OOP, and benefits in reusing *object-code* and reducing required time significantly when developing the applications with UIs. The abstraction of individual elements from the rest of the framework makes it easy to test an individual page while focusing on a singular part of the application without requiring a web server.

3.5. Using Spring Framework with MVC

Using *Spring Framework* MVC applications can be developed as follows. A request is handled (via the *annotation* "@Controller"), a new model is created, and the response (output data in the model) is returned to the user within a view e.g. by using JSPs. When running a page, one single class can be used to both respond to the user input and display output without instancing and initializing many that page dependent classes and child-controls. Moreover, MVC framework supports making both parallel development and test-driven development.

3.6. Micro services architecture and RESTful web services

Micro services are monolithic and small applications developed all in one piece so that it is easy to understand to develop an application as a set of modular components (collection of *loosely-coupled* services) with the minimum responsibility per each component. In the *loose-coupling* design, while supporting the separation of concerns with single-responsibility, reusability and flexibility, the degree of dependency between two *mostly independent* components is very low. A dependency that a component needs is incorporated into it indirectly without providing all the information of the dependency. Since the changes in the component *comp1* do not affect the actions of the component *comp2*, it becomes easier to manipulate, test, and maintain *comp1* independently.

Micro services are different with the sharing and reusing aspects while they can be seen as an evolution of SOA (*service-oriented architecture*) [15]. This architecture can provide significant benefits such as the possibility to design, develop, test, and release services with great agility while supporting autonomy and isolation [16]. Moreover, the *micro services* architecture is the case for the development of applications as suites of autonomous, conversational, and small services in understanding, deploying and scaling easily later [17].

A *micro service* provides an API (*application programming interface*) mostly such as REST (*REpresentational State Transfer*) API over HTTP. With developer independence, isolation, low complexity, and simple integration of *micro services*, the most common example of a *micro service* is a *RESTful* (REST architecture based) *web service*. The *annotation-based* MVC framework of *Spring Framework* simplifies the creation of *RESTful web services* [18]. Since *RESTful web service* controller (via the *annotation* "@RestController") corresponds to how *response body* is created, it simply returns the object whose data is directly written into the HTTP response as JSON or XML (*eXtensible Markup Language*). In *RESTful applications*, data retrieval, creation, update, and deletion can be done over HTTP requests. Since *RESTful web services* are based on a *resource-oriented* architecture, it is important to design the related *resource model* with all the classified resources for the interaction between *client* and *server* when implementing. Using REST APIs applications can communicate with each other over network. To easily manage the *dataflow* with web-related technology a generic *data bus* system may be implemented.

3.7. Data bus technology

The *data bus* standard is a technology used to share information in a *data-centric* manner in retrieving the needed information from database, and ensures interoperability among the integrated software applications to manage the *dataflow* in a secure manner.

4. IMPLEMENTED DATA BUS

To provide data to the *test management* and *test delivery* web applications we implemented a *data bus* to our CBA system based on the *test base system*. The design of the developed CBA system is given in Figure 1.

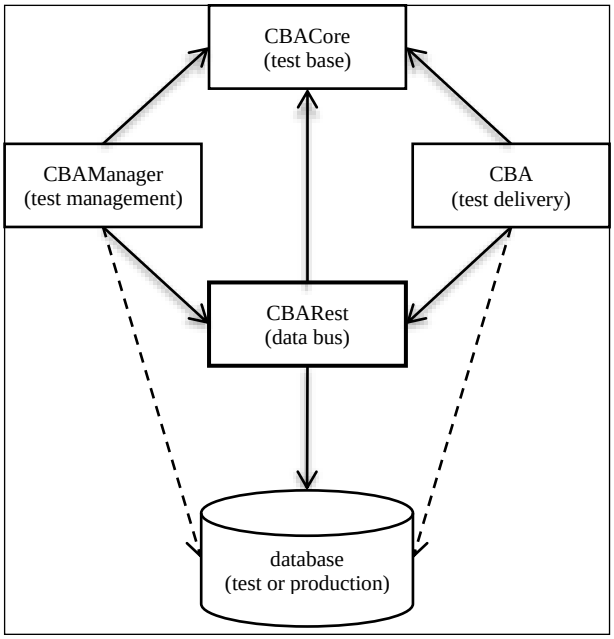


Figure 1. The design of the developed CBA system with the implemented *data bus*

The root packages in the *test base system* (the application *CBACore*) are given in Table 1.

Table 1. The root packages in *CBACore*

Package	**Task**
configuration	definitions for the *beta* and *production* releases
constant	*constant* and *enum* definitions
entity	*object models* regarding the tables in database
mapper	interfaces including database queries
model	*intermediate* and *custom* classes
utility	helper classes for the *data bus*

When writing *object models* based on the database schema, we created *base classes* for each table in database and the *extended ones* for making relations between them to keep a record individually or with its children. While the entity *module* includes its own attributes with a topic list belonging to that, its base entity *module-base* does not include any *module topics*. The entity *test-format* is the extension of its base entity *test-format-base* and includes a *test-format-question* list belonging to that. To design, implement and test the UIs of the *test management* and *test delivery* web applications *CBAManager* and *CBA* easily all database queries were structured in an *object-oriented* manner without using automatic SQL query and code generation or any database views.

For the design in Figure 1 including four *Maven* projects, we incorporated *CBACore* into other applications as a *Maven* dependency as given in Table 2.

Table 2. *Maven* dependency injection

CBARest*	**Definition in "pom.xml"**
Artifact	<modelVersion>4.0.0</modelVersion> <groupId>edu.akdeniz</groupId> <artifactId>cbarest</artifactId> <name>cbarest</name> <version>0.0.1</version>
CBACore dependency	<dependency> <groupId>edu.akdeniz</groupId> <artifactId>CBACore</artifactId> <version>0.0.55</version> </dependency>

* (the implemented *data bus* – the application *CBARest*)

The following statements are also incorporated into the *artifacts* in the POM files of the applications *CBAManager* and *CBA*:

- <version>beta</version>
- <packaging>war</packaging>

The version *beta* is used when testing the applications, and at the end of a successful build a WAR (*Web Application Resource* or *Web application ARchive*) file is created to run the application on an application server.

To easily change the versions of Java, *Spring Framework* and *Spring Security* we incorporated the essential *version descriptions* into the POM files of the applications as given in Table 3.

Table 3. The *version descriptions* for the applications

Definition in "pom.xml"
CBACore, *CBAManager* and *CBA* <properties> <java-version>1.8</java-version> <org.springframework-version>4.3.2.RELEASE</org.springframework-version> <spring.security.version>4.1.3.RELEASE</spring.security.version> </properties>
CBARest <properties> <java-version>1.8</java-version> </properties> <parent> <groupId>org.springframework.boot</groupId> <artifactId>spring-boot-starter-parent</artifactId> <version>1.4.3.RELEASE</version> </parent>

The applications *CBAManager* and *CBA* include *project-specific* configuration files and controllers, and make only their own *data manipulation* (insert, update, and delete) actions individually by following the execution of the needed *information services* of the *data bus*. The *data bus* includes the base settings, controllers (with the core *business logic* to return the raw data in database), and views (for the packaged information to be used) merging the procedures of the related controllers. Some views of the *data bus* are given in Table 4 with the controllers.

Table 4. Some views in *CBARest*

View	**Controllers**
module	▪ module-controller ▪ module-topic-controller
question-group	▪ question-group-controller
question	▪ question-controller ▪ question-answer-controller ▪ question-definition-view
question-definition	▪ question-definition-controller
test-format	▪ course-exam-test-format-controller ▪ test-format-controller
test-question-format	▪ test-question-format-controller ▪ test-topic-controller ▪ module-view
exam-booklet	▪ exam-booklet-controller ▪ exam-booklet-question-view ▪ test-format-view ▪ test-question-format-view
exam-booklet-question	▪ exam-booklet-question-controller ▪ exam-booklet-question-answer-controller ▪ question-group-view ▪ question-view
test-log	▪ test-log-controller ▪ test-log-answer-controller ▪ exam-booklet-view

Any *test question format* is retrieved over its own controller and the *test topic* belonging to that is set over its own while the *module description* and *module topic description* of the *test topic* are set over *module-view*. Any *exam booklet* with a set of questions is retrieved over *exam-booklet-view* while the *exam booklet* is returned over its own controller and the

exam-booklet-question list belonging to that with a set of answers per each question is returned over *exam-booklet-question-view*. *Test-log-view* is used to prepare different *exam booklets* (*test log* lists) for *test users* by using an intelligent shuffling algorithm that we developed before [14]. A service implemented for an entity or a model is only responsible in retrieving that entity/model related information.

To redirect the *data bus* to test or production database three classes were created with the *beta* and *production* releases extending the *default one* given in Table 5.

Table 5. The release configuration in *CBACore*

ReleaseConfigurationCBA.java

```
public class ReleaseConfigurationCBA {

 // attributes…
 // CBA_REST_URL
 // MYSQL_USERNAME
 // MYSQL_PASSWORD
 // MYSQL_DBNAME
 // …attributes

 public static ReleaseConfigurationCBA getReleaseConfiguration() {

  String rc = System.getProperty("RELEASE");
  if (rc == null) {
   return getDefaultReleaseConfiguration();
  }

  if (rc.equalsIgnoreCase("beta")) {
   return BetaReleaseConfiguration.getBetaReleaseConfiguration();
  } else if (rc.equalsIgnoreCase("prod")) {
   return ProdReleaseConfiguration.getProdReleaseConfiguration();
  }

  return getDefaultReleaseConfiguration();
 }

 private static ReleaseConfigurationCBA getDefaultReleaseConfiguration() {
  return new ReleaseConfigurationCBA();
 }
}
```

The head definitions in *CBARest* are given in Table 6, Table 7 and Table 8.

Table 6. The entrance of *CBARest*

Application.java

```
@SpringBootApplication
public class Application {

 public static ReleaseConfigurationCBA configuration;
 static {
  configuration = ReleaseConfigurationCBA.getReleaseConfiguration();
 }

 public static void main(String[] args) {
  SpringApplication.run(Application.class, args);
 }
}
```

Based on the "Application.configuration" given in Table 6, the test or production database is set to work on it while for the *beta-release-configuration* the "CBA_REST_URL" given in Table 5 is only changed.

Table 7. The container configuration of *CBARest*

ContainerConfig.java

```
@Component
public class ContainerConfig implements EmbeddedServletContainerCustomizer {

 @Override
 public void customize(ConfigurableEmbeddedServletContainer container) {
  container.setPort(7100);
 }
}
```

The port is set with "7100" as seen in Table 7 and the *data bus* is used over that port after deploying *CBARest* on the relevant machine.

Table 8. The database configuration of *CBARest*

DatabaseConfigMySQL.java

```
@Configuration
public class DatabaseConfigMySQL {

 @Bean
 public DataSource sqlDataSource() {
  BasicDataSource dataSource = new BasicDataSource();
  dataSource.setDriverClassName(*);
  dataSource.setUsername(*);
  dataSource.setPassword(*);
  dataSource.setUrl(*);
  dataSource.setDefaultAutoCommit(*);
  return dataSource;
 }

 @Bean(name = "sqlSessionFactoryMySQL")
 public SqlSessionFactory sqlSessionFactory() {
  final SqlSessionFactoryBean sessionFactory = new SqlSessionFactoryBean();
  sessionFactory.setDataSource(sqlDataSource());
  try {
   return sessionFactory.getObject();
  } catch (Exception e) {
   e.printStackTrace();
  }
  return null;
 }

 @Bean
 public DataSourceTransactionManager transactionManager() {
  return new DataSourceTransactionManager(sqlDataSource());
 }

 @Bean
 public MapperScannerConfigurer sqlMapperConfigurer() {
  MapperScannerConfigurer msc = new MapperScannerConfigurer();
  msc.setBasePackage("edu.akdeniz.cbacore.mapper");
  msc.setSqlSessionFactoryBeanName("sqlSessionFactoryMySQL");
  return msc;
 }
}
```

The fields stated with *stars* (*) in Table 8 are set via the corresponding attributes of the "Application.configuration" given in Table 6. For the database queries, the *mapper* package of *CBACore* is set as seen in Table 8.

In this scope, the mapper for the entity *module-topic* is given in Table 9.

Table 9. The mapper for the entity *module-topic*

ModuleTopicMapper.java

```
public interface ModuleTopicMapper {

 /* Record */

 @Select("SELECT * FROM ModuleTopic WHERE ModuleTopicID = #{moduleTopicId}")
 public ModuleTopicBase getModuleTopicBase(@Param("moduleTopicId") int moduleTopicId);

 /* List of Records */

 @Select("SELECT * FROM ModuleTopic WHERE ModuleId = #{moduleId}")
 public List<ModuleTopicBase> getModuleTopicBaseListByModuleId(@Param("moduleId") int moduleId);

 @Select("SELECT * FROM ModuleTopic WHERE ModuleId = #{moduleId} AND ModuleTopicEnabled = 1")
 public List<ModuleTopicBase> getActiveModuleTopicBaseListByModuleId(@Param("moduleId") int moduleId);
}
```

The controller for the *module-topic-mapper* is given in Table 10.

Table 10. The controller for the *module-topic-mapper*

ModuleTopicController.java

```
@RestController
public class ModuleTopicController implements ModuleTopicMapper {

 @Autowired
 private ModuleTopicMapper;

 @Override
 public ModuleTopicBase getModuleTopicBase(int moduleTopicId) {
  return moduleTopicMapper.getModuleTopicBase(moduleTopicId);
 }

 @Override
 public List<ModuleTopicBase> getModuleTopicBaseListByModuleId(int moduleId) {
  return moduleTopicMapper.getModuleTopicBaseListByModuleId(moduleId);
 }

 @Override
 public List<ModuleTopicBase> getActiveModuleTopicBaseListByModuleId(int moduleId) {
  return moduleTopicMapper.getActiveModuleTopicBaseListByModuleId(moduleId);
 }
}
```

Using the *annotation* "@Autowired" as seen in Table 10, the dependency *module-topic-mapper* is directly injected into the *module-topic-controller*.

5. RESULTS

To implement the *data bus* we used Java on *Spring Tool Suite* (STS) development environment on Ubuntu and developed a *RESTful web services* software program with a large number of APIs. Each API in this program corresponds to a resource in the *resource model*, and returns the response as JSON based on the request submitted with the URI (*uniform resource identifier*) and the input data required by the API. Moreover, any resource here is a Java method in its enclosing *controller class* that is associated with the corresponding HTTP procedure via the *annotation* "@RequestMapping" including the *value* (service URL) and the *method* (the *annotation* "RequestMethod", e.g. GET/POST). The output data may be provided as JSON over the *data bus* by using the given URLs via any web browser, or as JSP with the JSON data via the UIs of the applications *CBAManager* and *CBA* making requests over these URLs by adding the *annotations* "@ResponseBody" for the relevant Java method and "@RequestParam" (*not essential*) for its each parameter.

Project life-cycle tools are given in Table 11.

Table 11. The tools used for the implementation

Issue	Tool
Software project management tool	Apache Maven
Revision control system	Apache Subversion (SVN)*
Continuous deployment center	Jenkins^
Developer reference sources	JavaDoc
Test application server	Tomcat

* ("http://software.akdeniz.edu.tr/svn/CBACore")
^ ("http://software.akdeniz.edu.tr:8080")

To use the *CBACore* dependencies on the developer/tester *local* machines not including the application own, *test* machine or *production* machine we established an *artifactory* at "http://software.akdeniz.edu.tr:9081". For the final version of *CBACore*, the following *artifacts* exist at the *repository browser* "/artifactory/simple/ext-release-local/edu/akdeniz/CBACore/0.0.55/" of the *artifactory*:

- CBACore-0.0.55-javadoc.jar // documentation
- CBACore-0.0.55.jar // application
- CBACore-0.0.55.pom // POM file

We configured the integration server *Jenkins* so that the successful build of *CBACore* automatically triggers the builds of other applications after deploying the *artifacts* of the final version of *CBACore* on the *artifactory*.

To handle the "Application.configuration" given in Table 6 we defined an *environment variable* in *Jenkins* with the following statements:

- java -D=beta // http://software.akdeniz.edu.tr
- java -D=prod // https://bys.akdeniz.edu.tr

With the help of this definition, the APIs of the *data bus* are provided from the *test* machine or *production* machine. However, when working on a *local* machine during development and/or test, the relevant machine ("http://localhost:7100") or another machine (for the *default-release-configuration* by changing only the "CBA_REST_URL" given in Table 5 easily) provides the APIs. Besides this, to use the APIs in a secure manner we set the properties "security.user.name" and "security.user.password" in the file "application.properties" of *CBARest*.

The applications are tested as follows: (i) the final version of the application in SVN is compiled on *Jenkins*. (ii) when the compilation succeeds, the Java documentation is generated and published, and the application is automatically deployed on the application server. (iii) until no known bug exists in the *beta release*, the application is tested, the caught bugs are fixed with refactoring, and the new releases are published. (iv) when the application becomes stable, its *production release* is published by changing only the names of the database connections in the *project-specific* configuration files of the applications *CBAManager* and *CBA*.

To test the *data bus* when working on a *local* machine the "deploy.sh" scripts were created for each application for the aforementioned compilation and installation issues and the steps of the algorithm in Table 12 were targeted.

Table 12. The algorithm of testing the *data bus*

Algorithm: data-bus-test
▪ *CBACore* is compiled and installed into the "~/.m2/" *repository folder** on Ubuntu for other applications; o mvn clean package install ▪ *CBARest* is compiled and run as a *Spring Boot* application in the console of the STS-bundle, or on *Terminal* as follows; o mvn clean package o java -jar target/cbarest-0.0.1.jar ▪ The *data bus* is tested
▪ *CBAManager* and *CBA* are compiled and run on the application server in the STS-bundle, or on the one that is installed on Ubuntu via *Terminal* as follows; o mvn clean install -Dmaven.test.skip=true o cp target/cbamanager-beta.war $TOMCAT_HOME/webapps ▪ The *data bus* is used via *CBAManager* and *CBA*

* (this folder includes the file "settings.xml" that is a reference for the *user-specific* configuration for *Maven*)

To use the *data bus* via the applications *CBAManager* and *CBA* the design given in Figure 2 was used.

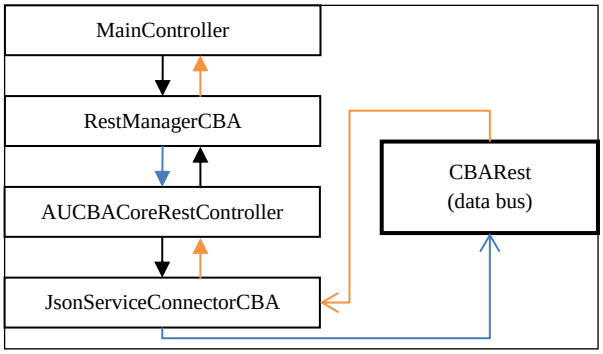


Figure 2. The design of using the *data bus* via the applications *CBAManager* and *CBA*

The design in Figure 2 works as follows. The *main-controller* is created based on the *singleton* design pattern and serves as a base controller of the application. The *rest-manager* of the *main-controller* sets the *configuration* in the *core-rest-controller* with a *static method* as given in Table 6 when initializing it. When a request is handled, the *core-rest-controller* returns the response to the *rest-manager* based on this *configuration* over the *JSON-service-connector* talking to the *data bus*, and the response is used via the *main-controller* to view or utilize in the UIs of the application. In this *dataflow*, the *core-rest-controller* passes the "CBA_REST_URL" given in Table 5 (*root*), the *enum* defined and the *parameters* ("MultiValueMap<String, String>") to the *JSON-service-connector*, and this *connector* returns the body of the response based on the *object model* "*class* of the *enum*" with the following: (i) the relevant URI obtained by the URL (*root* + "*URL* of the *enum*") and the *parameters*, (ii) HTTP method, (iii) HTTP request, and (iv) the *parameterized-type-reference* of the *class*.

Because of the "java.util.LinkedHashMap" cannot be cast to any *object model*, a few custom classes (*response wrapper actual types*) were created. The defined custom list for the base entity *module-base* is given in Table 13.

Table 13. Custom *module-base* list definition

CustomModuleBaseList.java

```
public class CustomModuleBaseList extends ArrayList<ModuleBase> {

 private static final long serialVersionUID = 1L;

 public CustomModuleBaseList() {
 }

 private final static ParameterizedTypeReference<CustomModuleBaseList> parameterizedTypeReference;
 static {
  parameterizedTypeReference = new ParameterizedTypeReference<CustomModuleBaseList>() {
  };
 }

 public static ParameterizedTypeReference<CustomModuleBaseList> getParameterizedTypeReference() {
  return parameterizedTypeReference;
 }
}
```

To provide data to the applications over the *data bus* based on a certain *object model* as seen in Table 13 two classes were created with *constant* and *enum* definitions.

The constant definition is given in Table 14.

Table 14. Constant definition in *CBACore*

AUCBACoreConstant.java

```
public class AUCBACoreConstant {

 private static Map<Class<?>, ParameterizedTypeReference> typeReferences;
 static {
  typeReferences = new HashMap<Class<?>, ParameterizedTypeReference>();
  setEntityTypeReferences();
  setCustomTypeReferences();
 }

 private final static void setEntityTypeReferences() {
  typeReferences.put(*.class, new ParameterizedTypeReference<*>() {
  });
  // …
 }

 private final static void setCustomTypeReferences() {
  putParameterizedTypeReference(*.class, *.getParameterizedTypeReference());
  // …
 }

 private static void putParameterizedTypeReference(Class<?> clazz, ParameterizedTypeReference<?> parameterizedTypeReference) {
  typeReferences.put(clazz, parameterizedTypeReference);
 }

 /* */

 public static ParameterizedTypeReference getParameterizedTypeReference(Class<?> clazz) {
  return typeReferences.get(clazz);
 }

 /* */

 public final static String *_SERVICE_URL = "/*";
 // other constant definitions…
}
```

* (*object models* and constants for the service URLs)

The enum definition is given in Table 15.

Table 15. Enum definition in *CBACore*

AUCBACoreEnum.java

```
public class AUCBACoreEnum {

 public enum URLCBACore {

  // …,
  *_URL(AUCBACoreConstant.*_SERVICE_URL, *.class);

  private final String url;
  private final Class<?> clazz;

  private URLCBACore(String url, Class<?> clazz) {
   this.url = url;
   this.clazz = clazz;
  }

  public String getUrl() {
   return url;
  }

  public Class<?> getClazz() {
   return clazz;
  }
 }

 /* */

 // other enum definitions…
}
```

* (enums for the service URLs with the corresponding *object models*)

With the help of the definition *URLCBA-core*, the *data bus* returns the response based on the *object model* specified in advance as seen in Table 15 by using the service URL of the related URL enum.

To demonstrate the success of the *data bus* for CBA we used (i) the *test management* web application *CBAManager* in managing question banks and tests as test managers, and (ii) the *test delivery* web application *CBA* in taking the tests as test takers, as follows. In each semester of *Akdeniz University*, any instructor may give a few courses in which each course is enrolled by a number of students with a set of exams. In our case-study, (i) the course exam to be applied on computer is associated with a *test format* within the entity *course-exam-test-format*, and exam date, duration and a unique password are specified for applying the test, (ii) the course students are assigned to at least one *test user group* belonging to the test as *test users*, and (iii) these students take the test on computer. University/semester, instructor, course, student and exam information are provided over another *RESTful web services* software program regarding the current *information system* at *Akdeniz University* that we developed before [19]. As a result, data is collected for the test over the *data bus* in this manner. In this scope, we conducted the following issues:

- using *CBAManager*;
  - creating a *question bank*
  - preparing an *exam booklet* with a set of questions including initial positions
  - preparing *test log* lists based on the *exam booklet* prepared with shuffled positions
- using *CBA*;
  - applying a *test* for the *exam booklet* by using the *test log* lists prepared

To prepare an *exam booklet* the algorithm in Table 16 was used.

Table 16. The algorithm of preparing an *exam booklet*

Algorithm: exam-booklet-preparation(course-exam-test-format-id)
1. EB = create an *exam-booklet*
2. get *course-exam-test-format* by *course-exam-test-format-id*
3. set the properties of the EB via *course-exam-test-format*
4. get *test-format* by *course-exam-test-format*
5. get *test-question-format-list* by *test-format*
6. list = []
7. **for** *test-question-format* **in** *test-question-format-list* **do**
8. list += { the procedure in Table 17 via *test-question-format* }
9. **end for**
10. EB.exam-booklet-question-list = list
11. **return** EB

To populate the *exam-booklet-question* list (step 10 in Table 16) the algorithm in Table 17 was used.

Table 17. The algorithm of populating an *exam-booklet-question* list

Algorithm: exam-booklet-question-list-population(test-question-format)
1. get *question-group-id-list* by *test-question-format*
2. get *question-list* by *question-group-id-list* and *test-question-format*
3. list = []
4. **for** *question* **in** *question-list* **do**
5. EBQ = create an *exam-booklet-question*
6. EBQ.question = question
7. c-list = { the procedure in Table 18 via *question* }
8. EBQ.exam-booklet-question-answer-list = c-list
9. list += EBQ
10. **end for**
11. **return** list

To populate the *exam-booklet-question-answer* list (step 8 in Table 17) the algorithm in Table 18 was used.

Table 18. The algorithm of populating an *exam-booklet-question-answer* list

Algorithm: exam-booklet-question-answer-list-population(question)
1. get *question-answer-list* by *question*
2. list = []
3. **for** *question-answer* **in** *question-answer-list* **do**
4. EBQA = create an *exam-booklet-question-answer*
5. EBQA.question-answer = question-answer
6. list += EBQA
7. **end for**
8. **return** list

Using the algorithm in Table 16, *exam booklets* are (i) firstly prepared by test managers and saved in database, and (ii) then retrieved from database to prepare *test log* lists in applying tests to test takers. For each *test user* in a test to be applied, a *test log* list with a set of answers per each *test log* is created based on the *exam booklet* prepared. The entities *exam-booklet-question* and *test-log* include the base entity *question-base* without any answers. The entities *exam-booklet-question-answer* and *test-log-answer* include the base entity *question-answer-base* without the information about the correct answers. Over the *data bus*, any *test log* is manipulated on the *manager side* while being displayed on the delivery side, and any *test log answer* is created on the *manager side* while being updated on the *delivery side*.

The applications *CBAManager* and *CBA* could work separately in integrity over the *data bus* since one single *configuration* was defined regarding the design in Figure 2 that behaves independently for both of the applications. The *data bus* is not known directly by any other application to which data is provided. Therefore, it becomes possible to use any programming language and/or web technology for the design and implementation of *front-end* issues with specific *back-end* issues.

The approach of implementing such a generic *data bus* system is fairly appropriate in long-term and huge software projects that are to be able to include many applications with/without large teams. For the developers especially wishing to completely control how an application will behave, MVC is a good development approach even if more implementation needs to be done.

6. CONCLUSIONS

Responding to all the *information retrieval* requests from the UIs of the applications via one single *front controller* (e.g. *main-controller*), the following benefits become the case: (i) preventing any application accesses to the database server directly, (ii) maintaining data integrity with a consistent flow (*redirections*) in the system, and (iii) avoiding potential dependency problems with flexibility and scalability.

When any overall change is needed throughout a software system, it becomes complicated to work both within *object models* and on database queries. Therefore, it is a good approach to implement a system for *information services* separately for the applications with UIs that do not include the related procedures. While testing the *test management* and *test delivery* web applications, we observed that *information retrieval* over the *data bus* is fairly flexible in CBA. Using the implemented *data bus*, we plan to extend our CBA system to *assessment* and *feedback* with a *test report system* for data analysis issues while collecting data by applying tests.

REFERENCES

[1] M. Maqableh, R. M. T. Masa'deh, A. B. Mohammed, "The Acceptance and Use of Computer Based Assessment in Higher Education", *Journal of Software Engineering and Applications*, 8, 557–574, 2015.

[2] F. Scheuermann, A. G. Pereira, **Towards a Research Agenda on Computer-Based Assessment - Challenges and Needs for European Educational Measurement**, OPOCE, 2008.

[3] K. Scalise, B. Gifford, "Computer-Based Assessment in E-Learning: A Framework for Constructing "Intermediate Constraint" Questions and Tasks for Technology Platforms", *Journal of Technology, Learning, and Assessment*, 4(6), 2006.

[4] A. A. Prisacari, J. Danielson, "Rethinking testing mode: Should I offer my next chemistry test on paper or computer?", *Computers & Education*, 106, 1–12, 2017.

[5] J. Pawasauskas, K. L. Matson, R. Youssef, "Transitioning to computer-based testing", *Currents in Pharmacy Teaching and Learning*, 6, 289–297, 2014.

[6] Y. P. Chua, "Effects of computer-based testing on test performance and testing motivation", *Computers in Human Behavior*, 28, 1580–1586, 2012.

[7] S. A. Nikou, A. A. Economides, "The impact of paper-based, computer-based and mobile-based self-assessment on students' science motivation and achievement", *Computers in Human Behavior*, 55, 1241–1248, 2016.

[8] C. Ricketts, S. J. Wilks, "Improving Student Performance Through Computer-based Assessment: Insights from recent research", *Assessment & Evaluation in Higher Education*, 27(5), 475–479, 2002.
[9] S. K. Singh, A. K. Tiwari, "Design and Implementation of Secure Computer Based Examination System Based On B/S Structure", *International Journal of Applied Engineering Research*, 11(1), 312–318, 2016.
[10] R. Clariana, P. Wallace, "Paper-based versus computer-based assessment: key factors associated with the test mode effect", *British Journal of Educational Technology*, 33(5), 593–602, 2002.
[11] M. Thelwall, "Computer-based assessment: a versatile educational tool", *Computers & Education*, 34, 37–49, 2000.
[12] B. Csapó, J. Ainley, R. E. Bennett, T. Latour, N. Law, "Technological Issues for Computer-Based Assessment", **Assessment and Teaching of 21st Century Skills**, Editors: P. Griffin, B. McGaw, E. Care, New York: Springer, 143–230, 2012.
[13] A. Aksoy, M. Gunay, M. Karakoc, "A Flexible Question Storage Hierarchy for Computer Based Assessment Systems", **International Conference on Engineering Technologies (ICENTE'17)**, Selcuk University, Konya, 17, December 07–09, 2017.
[14] M. Karakoc, A. Aksoy, M. A. E. Sen, M. Gunay, "Outcome Focused Computer Based Assessment System Using Intelligent Shuffling Algorithm", **26th International Conference on Educational Sciences (ICES-UEBK 2017)**, Side Starlight Resort Hotel, Antalya, 1283–1284, April 20–23, 2017.
[15] M. Garriga, "Towards a Taxonomy of Microservices Architectures", **International Conference on Software Engineering and Formal Methods (SEFM 2017)**, Editors: A. Cerone, M. Roveri, Springer, Cham, 203–218, 2018.
[16] P. Di Francesco, I. Malavolta, P. Lago, "Research on Architecting Microservices: Trends, Focus, and Potential for Industrial Adoption", **IEEE International Conference on Software Architecture (ICSA)**, Gothenburg, Sweden, April 3–7, 2017.
[17] L. Baresi, M. Garriga, A. De Renzis, "Microservices Identification Through Interface Analysis", **Service-Oriented and Cloud Computing (ESOCC 2017): 6th IFIP WG 2.14 European Conference**, Editors: F. De Paoli, S. Schulte, E. B. Johnsen, Springer International Publishing, Oslo, Norway, 19–33, September 27–29, 2017.
[18] Internet: Spring Framework: @RestController vs @Controller - Genuitec, https://www.genuitec.com/spring-frameworkrestcontroller-vs-controller/, 19.06.2018.
[19] M. Karakoc, M. Gunay, "Akdeniz Üniversitesi BYS için RESTful Web Servisleri", **Akademik Bilişim 2018 Konferansı**, Karabük Üniversitesi, Karabük, 8, 31 Ocak – 02 Şubat, 2018.

DESIGN AND IMPLEMENTATION OF AN AGENT BASED AUTOMATED AUCTION ENVIRONMENT

Uğur GÜLERYÜZ[1] and Prof. Dr. Nadia ERDOĞAN[1]

[1] Istanbul Technical University, Maslak, 34467 Sarıyer/İstanbul
{guleryuzu,nerdogan}@itu.edu.tr

Abstract. In this study, an execution environment is designed and implemented in which users can purchase and sell products using different types of auction methods manually or automatically according to the parameters and strategies to be entered beforehand. The environment has an entirely agent-based design where the main components of the environment are represented by agents. There are four types of agents: server, gui, bidder, and auctioneer. Different types of agents communicate among each other and constitute the main architecture of the system. The system supports English and Dutch Auctions and offers different strategies to the buyer and seller side for each auction type. In the experiments, it was observed that the system meets the requests of the users within two types of auctions with the options offered by the system, and it can participate in auctions with the strategies and parameters chosen by buyers and sellers according to their budgets, their desire to buy or sell the product. The presented strategies have been compared and it has been observed that different strategies come to the fore in different situations and have a large impact on the outcome of the auction.

Keywords: Agent Based Programming, Jade, Multi Agent Systems, English Auction, Dutch Auction.

1 Introduction

Today, people have started to meet their shopping needs over the internet besides many other needs. Thanks to e-commerce sites or applications, people can do the job of both buying and selling products with little effort. Products can be sold at a fixed price or can be sold online via various auction methods. More than 2500 online auction houses serve the users on the internet. The best known of these are auction houses such as eBay, Amazon, Yahoo! Auction, Ubid and these auction houses support auction types such as English, Dutch, first-price sealed-bid auction, second-price sealed-bid auction [1].

While buying and selling through auction methods, sellers aim to sell the product at the maximum price they can sell and buyers aim to buy the product with the minimum amount they can buy. In addition, it is critical for both sides to sell the product in the optimum time possible. While the sale of any product, many buyers compete with each

other and the seller checks the auction by examining the offers of each buyer while selling. In line with their budgets, price increases with the right strategy ensures that they are the winners of the auction. In addition, in the auction methods where the price change is made by the seller right strategy enables the product to sell at the price seller wants. For this reason, both buyers and sellers must follow the auction properly during the auction.

The aim of this work is to design and implement an agent based environment where agents can perform both product purchase and sale operations on behalf of their owners. In addition, while a product is being bought or sold, making transactions such as making an offer, increasing the price, comparing the offers with the previous offers and deciding the highest offer, informing the buyers about the current price of the product, declaring the winner and selling the product are completely automated by the software and minimum effort of the users. It is aimed to realize the buying and selling operations with maximum efficiency. Various strategy options are offered to both buyers and sellers according to their budget and their desire to buy the product. Buyers and sellers can automatically buy and sell products according to these options. In the implementation phase agent based programming is used and main components of the system are represented as agents.

2 Related Work

Auction is a method of sale where movable or immovable property is sold to the person offering the highest amount. Auctions have been used in many different areas until today. During the seventeenth century and shortly after the French revolution, artistic products were sold using auctions in tavernas and cafes, and catalogs were printed to announce products at daily auctions [2].

There are many different auction methods such as English, Dutch, first-price sealed-bid auction, second-price sealed-bid. English auction is the most used auction type, the price of the product is increased until a single buyer remains, and the buyer who made the last bid and made the highest bid has purchased the product [3]. In the Dutch auction, the seller starts the auction with a high price that the buyers cannot afford and regularly informs the buyers by decreasing the price. At each price decrease, buyers can accept or reject that price by examining the price information, when any buyer accepts the current price, that buyer wins the auction [4]. The strategy that the buyer and seller will make during the auction affects the sales price and sales time of the product critically.

KASBAH strategies are price reduction and increase strategies are presented in order to effectively ensure the agreements between buyers and sellers. KASBAH strategies consist of three separate strategies for the buyer and the seller, and according to these strategies, price decreases for the seller and price increases for the buyer. Strategies are shaped according to the parameters entered by the user. These parameters are parameters such as end date, desired sales price, lowest price to be accepted and pricing strategy. Using these values and mathematical formulas that use these values, a price curve is formed according to time, which always corresponds to a selling or buying price.

Thus, the amounts to be offered are determined according to the current time. The shape of the curve is determined by the pricing strategy that the user chooses. KASBAH consists of three different strategies shown in Figure 1 [5]:

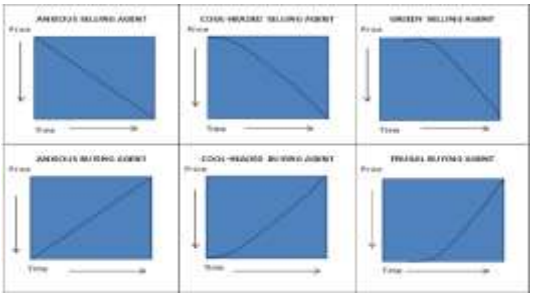


Fig. 1. KASBAH Strategies [6].

- Anxious Buyer / Seller: In this method, the seller side reduces the price linearly, while the buyer side increases the price linearly. This method refers to the user who is willing to buy or sell in the pricing changes earlier. This method is recommended to the user who wants to buy or sell the fastest way.
- Cool-Headed Buyer / Seller: In this method, prices are the second degree function of time and the increase and decrease operations are done quadratic. Since it is a method between the willing and calm, it is balanced.
- Greedy / Frugal Buyer / Seller: In this method, compared to other methods, there is a minimum change in prices at the earliest times and maximum changes towards the end date. It is recommended for the seller who wants to sell at the highest price he can sell, or for the buyer who wants to receive the lowest price [5].

As price changes are the most critical parameter in English and Dutch auctions, KASBAH strategies are very suitable for both buyers and sellers. With these strategies, buyers and sellers can automatically create bids at any time and these strategies enable them to conclude their purchase and sale as they wish.

Similar to this study, an application using KASBAH strategies was implemented using JADE and JAVA. KASBAH strategies have been applied to the seller side of the English auction, unlike this study, and the English type auction is implemented by the seller's price changes differently from this work [7].

3 System Description

In this work, an environment has been developed in which users can act both as buyers and sellers, put products they want on sale and bid on products they want to buy. JAVA programming language and JADE[9] framework are used in the implementation stage. English and Dutch auction methods are supported by the environment. Many options and strategies are provided for users to purchase and sell easily and au-

tomatically. KASBAH buyer and seller strategies are used while performing automatic price updates. When looking at the general system, users can create two types of auctions, English and Dutch types, and for each auction type, there are different strategies and options that allow users to buy and sell products as they wish, on both the buyer and seller side. Users can also bid manually while making their purchase and sales transactions.

In English Auction the buyer side can program his bids in two ways while participating in an English Auction. The user who wants to bid periodically enters the frequency, maximum bid, starting bid and price increasing strategy as input. According to the parameters entered, price increases automatically with time. The user who wants to make a single offer at a certain time before the end, enters the time and price information. The seller side enters the auction end date, minimum accepted price and the lowest price to be sold. If the product reaches the lowest accepted price that can be sold until the auction end date but the product which cannot reach this amount until the auction end date is sold to the amount that the seller reaches on the end date.

In Dutch Auction the seller side enters the start price of the product, minimum price he can drop, auction end date and the price reduction strategy. The buyer side enters the price check frequency and maximum bid parameters when programming his bids. So that seller side purchase the product when it falls to the desired price.

When the system first runs which consists of the database and the server agent. This structure refers to the state of the system before any user logs in. When users log into the system, new agents for each user are connected to the system. After users open the application and login via the interface, the gui and bidder agent belonging to each user are constructed and connected to the system. For example, the agent model of the system, in which four users log in and no auction is created, is showed in Figure 2.

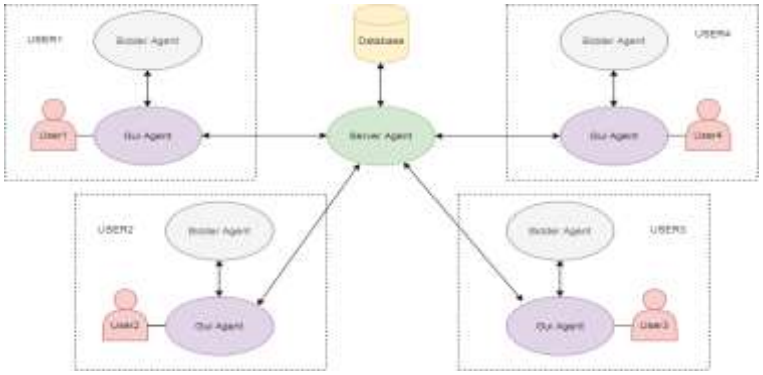


Fig. 2. Agent Model After Users Login.

Users who participate in the system can create a new auction using the auction method they choose for the products they want to sell, an auction agent is connected to the system for each new auction created and it performs all the operations related to the auction. Users who want to participate the auction can do it through their representitive

bidder agent through manual or automatic bids. Once the agents are created and a number of users participate in the auctions, the architecture will be populated by the agents depicted in Figure 3. In the figure, as an example two auctions have been created by the users and two users have participated in each auction.

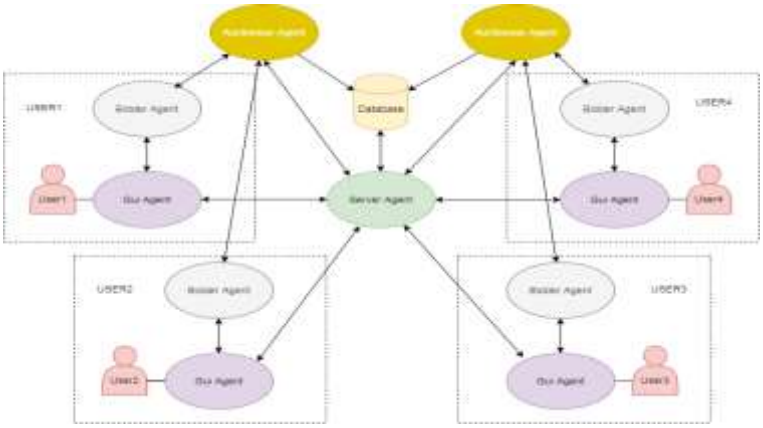


Fig. 3. Agent Model After The Auctions Are Created.

3.1 Server Agent

Server agent is one of the two agents that have direct access to the database along with the auction agent. The server agent is single in the system and it is created at the during the initialization of the system. As long as the system is running, the server continues to actively perform its duties. The tasks of the server agent can be listed as follows.

- When a new user is created, it takes the information of the user to be created in the gui and saves it to the database.
- When a user want to login, it takes the user information from the gui agent and checks the accuracy of information from the database after which it allows for user login.
- When a new auction is to be added to the system, it adds the new auction to the database using the information it receives from the gui agent.
- When an auction is to be deleted, it deletes the auction from the database.
- It transmits the information of auctions that are currently active to the gui agent by retrieving them from the database so that users can view the active auctions through the gui.

3.2 Gui Agent

A gui is created for each user, and continues execution as long as the user actively uses the application. It is used to implement the actions that the user performs on the gui and to communicate with other agents. The tasks of the gui agent can be listed as follows.

- The gui agent takes the actions of the user from the gui form and transmits the messages required for these operations to other agents.
- When users want to login or create a new account, it sends the user information to the server agent.
- When the users want to add or delete the auction, it sends the auction information to the server agent and enables the auction to be added or deleted.
- When the users want to see the auctions that are already active, the gui agent sends a request to the server agent to receive the active auction list.
- When the users want to participate in one of the active auctions, it sends a request to auction agent with that id, allowing the user to participate to that auction.
- When the users want to send an manual bid to one of the auctions that user participates in, it receives the bid information from the gui form and sends it to the auction agent with that id.
- When the users want to program their bids for one of the auctions that user participated in, it sends the configuration information from the gui form to the bidder agent.

3.3 Bidder Agent

The bidder agent created for each user after the user logs into the application and performs operations such as joining the auction, bidding the auction automatically, checking the auction price, and increasing the bid price according to the specified strategies. The bidder agent acts according to the auction method and user's preferences. Every bidder agent has a behavior table which keeps the behaviors that correspond to every auction so that a single bidder agent can serve accordingly for each auction that is participated by the user. Tasks of the bidder agent can be listed as follows.

- When the user makes automatic bidding configuration from the gui form, it takes the configuration information and adds the behavior information to the behavior table.
- When the Dutch type is programmed to automatically bid at the auction, it checks the auction price according to the frequency value entered by the user and purchases the product by paying the price if the current price less than maximum price that user can pay.
- In the English Auction, when the user periodically configures the bid by entering the bid start date, frequency, maximum bid, start bid, and price increase strategy ("anxious", "cool-headed", "frugal") periodically increasing the bid makes an offer to auction.
- When the user wants to program their bids with a single bid prior to the English Auction end date it bid the given time automatically.

3.4 Auctioneer Agent

Auctioneer agent is created when a user adds an auction to the system and remains alive until the end of the auction. There is one auctioneer agent for each auction created. Since two types of auctions can be created, English and Dutch types, the duties and behaviors of the auction agent vary according to the auction type. Behaviors that are

appropriate for the auction type are connected to the auctioneer agent so that it acts according to the type of auction it represents. The auctioneer agent does the works such as adding the participants to the auction, informing the participants about the status of the auction, accepting the bids, ending the auction by declaring the winner of the auction. The tasks of the auction agent can be listed as follows.

- It enables the user to participate to the auction by taking the request of the user coming from the gui agent.
- Responds to requesting information about the auction from the bidder by sending up-to-date information about the auction.
- If the auction type is a English Auction, it accepts and processes the bids that come from the bidder and gui agents.
- If the auction type is a Dutch Auction, it makes price changes according to certain periods and the strategy choosed by the user.

4 Experiments and Evaluation

In order to examine how the auctions end according to different strategies, auction types and user preferences, we have conducted a number of experiments and we compare and evaluate the results. We describe the experiments in two groups according to their type, namely English and Dutch auction. In addition, we have also performed some experiments to assess the performance of the system. “Test runs were executed on a single computer, however the system is designed such that it is capable to run on different computers of each buyer gui and the server.

4.1 Dutch Auction Experiments

Effects of the Price Reduction Strategy. In the first 3 experiments, all under the same conditions, it was aimed to observe the effects of the sellers price reduction strategy on the auction. For this reason, an auction consisting of three buyers and a seller, where the seller put the product on sale as a Dutch Auction, was created. At the Dutch Auction, the seller side updates the price of the product every 2 minutes, and chooses one of the "anxious, cool-headed, greedy" strategies.

Table 1. Parameters of Seller in Experiment 1-2-3.

	Auction Time	**Starting Price**	**Min Price**	**Strategy**
Experiment 1	20 minutes	150 units	20 units	Anxious
Experiment 2	20 minutes	150 units	20 units	Cool-Headed
Experiment 3	20 minutes	150 units	20 units	Greddy

Table 2. Parameters of Buyers in Experiment 1-2-3.

Parameter	**Buyer-1**	**Buyer-2**	**Buyer-3**
Check Frequency	3 minutes	3 minutes	3 minutes
Maximum Bid	15 units	25 units	55 units

When we look at the auction participants, the maximum bid they can make is different, but the control frequencies are the same. It is clear that the buyer with the maximum bid can win the auction, but it depends on the seller's strategies how long is takes the auction to conclude and at what price. In the first three experiments, price reductions made by the sellers depending on the time occurred as shown in the Figure 4. The auction in experiment-1 was sold to the buyer-3 at a price of 46 units and took 16 minutes. The auction in experiment-2 was sold to the buyer-3 at a price of 44 units and took 18 minutes. The auction in experiment-3 was sold to the buyer-3 at a price of 20 units and took 20 minutes. In the first three experiments, by keeping all other parameters constant in the Dutch auction, only the seller's price reduction strategy has been changed and the effect of price reduction strategies on the sales price and duration of the product has been observed. The maximum prices that can be given by the buyers are kept close to the minimum price that the seller can sell, and when the anxious strategy is used as a strategy, it is observed that the product is sold as soon as possible due to the linear decrease depending on the time. In addition, the anxious strategy has always kept the same amount changes, and the other two strategies have lower amounts changes in early time and higher amounts changes in the late time. The Anxious strategy allowed the buyer to sell his product at a faster rate and higher price than the other two strategies. The fact that the maximum prices that the buyers can give is close to the minimum price accepted by the seller was effective in this. For this reason, in the next three experiments, the maximum prices that can be given by the buyers will be kept closer to the price that the seller has started the auction with and the results produced by these parameters will be examined.

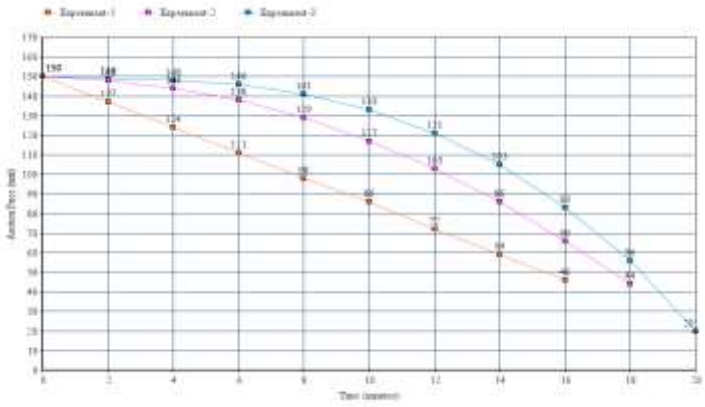


Fig. 4. Price Changes According to Time in Experiment 1-2-3.

In the fourth, fifth and sixth experiment, the maximum prices that the buyers can pay were kept close to the seller's starting price. Buyer and seller parameters in the fourth, fifth and sixth experiment are shown in Table 3-4.

Table 3. Parameters of Seller in Experiment 4-5-6.

	Auction Time	**Starting Price**	**Min Price**	**Strategy**
Experiment 4	20 minutes	150 units	20 units	Anxious
Experiment 5	20 minutes	150 units	20 units	Cool-Headed
Experiment 6	20 minutes	150 units	20 units	Greddy

Table 4. Parameters of Buyers in Experiment 4-5-6.

Parameter	**Buyer-1**	**Buyer-2**	**Buyer-3**
Check Frequency	3 minutes	3 minutes	3 minutes
Maximum Bid	50 units	75 units	105 units

The auction in experiment-4 was sold to the buyer-3 at a price of 98 units and took 8 minutes. The auction in experiment-2 was sold to the buyer-3 at a price of 103 units and took 12 minutes. The auction in experiment-3 was sold to the buyer-3 at a price of 105 units and took 14 minutes. In the fourth, fifth and sixth experiments, price reductions made by the sellers depending on the time occurred as shown in the Figure 5.

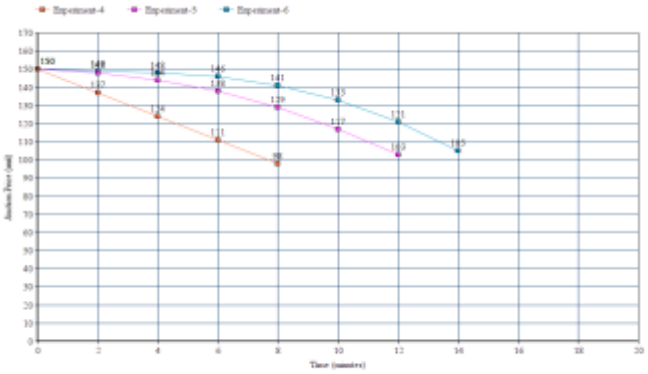


Fig. 5. Price Changes According to Time in Experiment 4-5-6.

As the price reduction strategy became linear in the first three experiments, it was concluded that the product sales were faster and this result was supported by the fourth, fifth and sixth experiments. In addition, if the maximum price parameter that buyers can give is close to the auction start price, the anxious strategy enables the seller to sell the product for a higher price since it reduces the price linearly. However, if the maximum price parameter that the buyers can give is close to the minimum price value of the seller, the greedy strategy causes the seller to sell a higher amount of the product, since the greedy strategy causes lower amounts of price decrease in the early time's of auction and the higher the end of the auction. Cool-headed strategy is a balance strategy that has some features in the two strategies. The seller can sell his product faster with anxious strategy regardless of the parameters of each buyer.

Effects of the Price Update Frequency. In the system, the default value of the price change frequency of the Dutch auction is 2 minutes. In this section, the effects of this value change on the auction are examined. Experiments were made by keeping the other parameters constant and changing the frequency and the results were formed as follows.

Table 5. Parameters of Seller in Experiment 7.

Auction Time	**Starting Price**	**Min Price**	**Strategy**	**Frequency**
20 minutes	150 units	20 units	Anxious	1-2-3-4-5-6

Table 6. Parameters of Buyers in Experiment 7.

Parameter	**Buyer-1**	**Buyer-2**	**Buyer-3**
Check Frequency	3 minutes	3 minutes	3 minutes
Maximum Bid	50 units	75 units	105 units

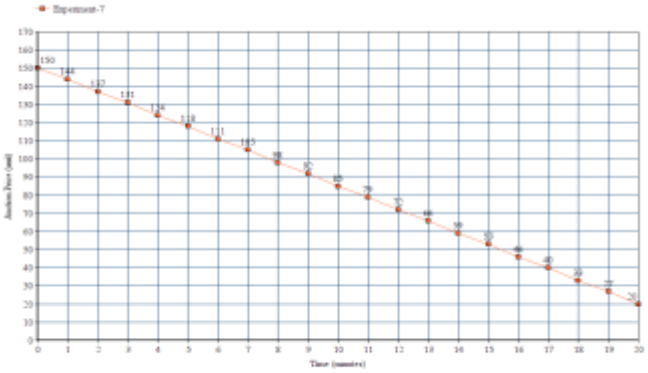


Fig. 6. Price Changes According to Time in Experiment 7.

Price update frequency does not affect the winner at Dutch auction, but it affects the sales price of the product. If the price update frequency is low, the more the seller is more likely to sell the product at a higher price

Table 7. Experiment 7 Results

Price Update Frequency	1	2	3	4	5	6
Sale Price	105	98	92	98	85	72

Effects of the Number of Buyers. The number of buyers does not affect the result of the Dutch auction. The buyer with the highest maximum bid value wins the Dutch type auction, but other parameters affect the auction time and the sale price. The increase in

the number of buyers affects the system performance and the experiments for this will be explained in the next sections.

Effects of the Auction End Time. In this experiment, the other parameters were kept constant and the auction end time was changed and its effect on the auction was examined. Three experiments were carried out and buyer and seller parameters are as in Table 8-9.

Table 8. Parameters of Seller in Experiment 8-9-10.

	Auction Time	**Starting Price**	**Min Price**	**Strategy**
Experiment 1	20 minutes	150 units	20 units	Anxious
Experiment 2	36 minutes	150 units	20 units	Anxious
Experiment 3	80 minutes	150 units	20 units	Anxious

Table 9. Parameters of Buyers in Experiment 8-9-10.

Parameter	**Buyer-1**	**Buyer-2**	**Buyer-3**
Check Frequency	3 minutes	3 minutes	3 minutes
Maximum Bid	50 units	75 units	105 units

The auction in experiment-8 was sold to the buyer-3 at a price of 98 units and took 8 minutes. The auction in experiment-9 was sold to the buyer-3 at a price of 100 units and took 10 minutes. The auction in experiment-8 was sold to the buyer-3 at a price of 105 units and took 28 minutes. In the eighth, ninth and tenth experiments, price reductions made by the sellers according to shown in the Figure 7.

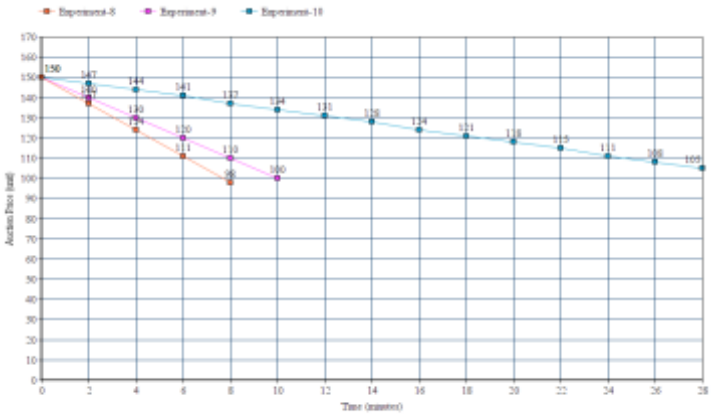


Fig. 7. Price Changes According to Time in Experiment 8-9-10.

Since the increase of the auction period spreads the seller's price decreases over a longer period, the price decreases are realized with smaller values. For this reason, it takes

longer for the product price to decrease to the amounts that the buyers can pay. However, smaller price decreases increase the possibility of the seller to sell the product at a higher price.

4.2 English Auction Experiments

Effects of Price Increase Strategy. In the following experiments, the parameters entered by the buyers in the English auction and the effects of the price increasing strategies they have chosen were examined. For this reason, auctions consisting of one seller and three buyers were compared with different parameters.

Table 10. Seller Parameters of Experiment 11.

Starting Price	**Auction Time**	**Minimum Accepted Price**
5 units	21 minutes	75 units

Table 11. Buyer Parameters of Experiment 11.

Parameter	**Buyer-1**	**Buyer-2**	**Buyer-3**
Bid Frequency	3 minutes	3 minutes	3 minutes
Starting Bid	20	15	40
Maximum Bid	100	50	150
Increase Strategy	Anxious	Anxious	Anxious

In the English type auction, when the buyers are programmed to bid periodically, they make their bids according to their chosen strategy and bidding frequency. The periodic bids of buyers to be given periodically are showed in Figure 8, their bids are accepted or rejected according to the auction status. The auction ends when the auction reaches the minimum price that the seller will accept. Auctioneer Agent only accepts bids that are better than the best offer. The auction in the eleventh experiment showed in Table 18 and the buyer-3 won the auction with a price of 97 units. Also, the auction ended in 11 minutes.

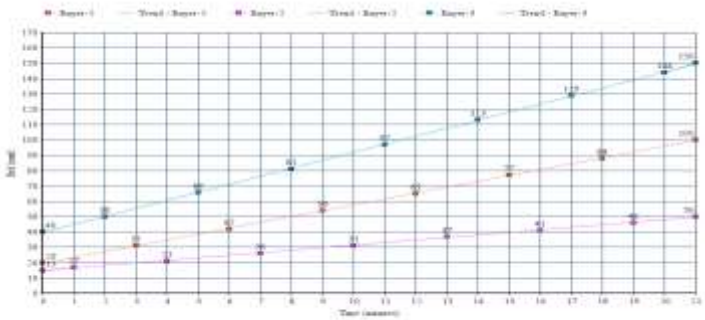


Fig. 8. Bids of Buyers in Experiment 11.

Table 12. Experiment 11 Bids.

Bid	Price	Bidder
1	17 units	Buyer-2
2	50 units	Buyer-3
3	54 units	Buyer-1
4	97 units	Buyer-3

In the twelfth experiment, in order to examine the effects of price increase strategies on the auction sales price, the auction winner and the duration of the auction, all other parameters were kept constant and the price-raising strategy of the buyer-3 was determined as frugal. The parameters of buyers and sellers are as in Table 13 and Table 14.

Table 13. Seller Parameters of Experiment 12.

Starting Price	Auction Time	Minimum Accepted Price
5 units	21 minutes	75 units

Table 14. Buyer Parameters of Experiment 12.

Parameter	Buyer-1	Buyer-2	Buyer-3
Bid Frequency	3 minutes	3 minutes	3 minutes
Starting Bid	20	15	40
Maximum Bid	100	50	150
Increase Strategy	Anxious	Anxious	Frugal

Due to the fact that only buyer-3 made a change of strategy, the bids of buyer-3 were formed according to time as in Figure 12. Since the buyer-3 chose frugal instead of anxious as the auction method of auction, he increased his bids less early in the auction and made higher increases towards the end. The auction lasted 15 minutes and the buyer-1 won the auction with a price of 77 units. The bids accepted by the auctioneer at the auction are as in Table 15.

Table 15. Experiment 12 Bids.

Bid	Price	Bidder
1	17 units	Buyer-2
2	41 units	Buyer-3
3	42 units	Buyer-1
4	46 units	Buyer-3
5	54 units	Buyer-1
6	55 units	Buyer-3
7	65 units	Buyer-1
8	72 units	Buyer-3
9	77 units	Buyer-1

In this experiment, although the maximum price was high, it was observed that the buyer-3 lost the auction because he chose the frugal and also the auction time was extended. As a result of the experiment, it was seen that both the maximum price to be paid by the buyer side and the strategy significantly affect the the auction.

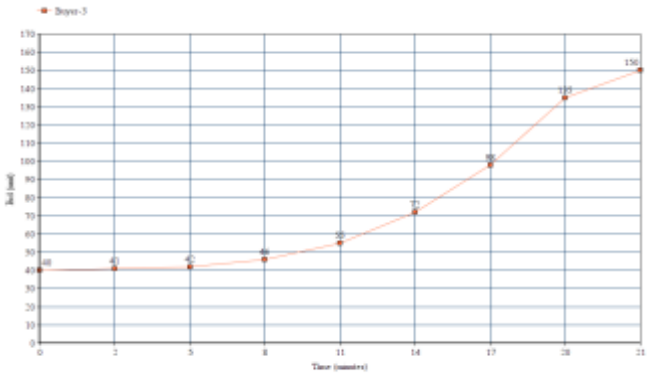


Fig. 9. Bids of Buyer-3 in Experiment 12.

In the thirteenth experiment, the parameters were kept the same as the previous experiment and an extra buyer was added to the auction. The added buyer was programmed to bid 80 units 10 minutes before the end of the auction. The accepted proposals are showen in Table 16.

Table 16. Experiment 13 Bids.

Bid	**Price**	**Bidder**
1	17 units	Buyer-2
2	41 units	Buyer-3
3	42 units	Buyer-1
4	46 units	Buyer-3
5	54 units	Buyer-1
6	55 units	Buyer-3
7	65 units	Buyer-1
8	72 units	Buyer-3
9	80 units	Buyer-4

Effects of the Number of Buyers. The number of buyers does not affect the result of English auction because buyers produce the bids independently from other buyers. However, the added buyers have better strategy and budget than the other buyers, as in the previous experiment, which enables them to win the auction.

Effects of the Auction End Time. In English auction, auction end date does not affect the auction result as the auctions are terminated when the minimum price desired by the seller. However, this time is insufficient when the buyers do not reach the minimum price demanded by the seller until auction end time.

4.3 Performance Experiments

Various performance tests have been carried out to measure the performance and scalability of the system. When measuring response times, the time before the sent ACL message and the time after the return of ACL message was taken into account. In the first stage, the response time of the server was measured by increasing the number of users. When the number of users was increased up to 2000, it was observed that there were no serious delays as shown in Table 17.

Table 17. Response Time of Server Agent.

Number of Users	500	1000	1500	2000
Response Time of Server	20 ms	23 ms	25 ms	29 ms

Next, the number of auction participants were increased and the response time of the bid was measured by increasing the number of participants.

Table 18. Response Time of Auction Agent for Bid.

Number of Users	500	1000	1500	2000
Response Time of Auction	12 ms	15 ms	21 ms	25 ms

Finally, by increasing the number of auctions, the response of the auctions for users who request to participate was measured.

Table 19. Response Time of Auction Agent for Participation.

Number of Auction	500	1000	1500	2000
Response Time of Auction	8 ms	13 ms	17 ms	21 ms

Thanks to JADE's fast ACL-messaging service, the system can tolerate a high number of users, auctions and auction participants.

5 Result

In this study, a agent-based environment has been implemented in which users can automatically sell their products with the parameters of the English and Dutch Auction methods only with the parameters they enter. Many options and strategies have been presented and a flexible structure has been created in order to enable users to direct their auctions as they wish. While designing the system, the system components were represented by agents and four different types of agents were created: server, auctioneer, bidder, and gui. Each agent constitutes the main system by performing its own specialized task and communicating with each other via messaging. As a result of the experiments, it was observed that the system successfully performed the purchase and sale transactions according to the user preferences, and the auction could be completed in different ways according to the parameters and strategies entered. A distributed structure was created thanks to agent-based programming. By adding the necessary behaviors to the agents, it was ensured that they perform the necessary operations without user intervention. For future work, the number of servers can be increased as it may

become a bottleneck and communication with each server can be used. In addition, different product sales methods and different strategies can be added for sales methods to enrich user ‘s preference options.

References

1. Anthony, Patricia & Jennings, Nicholas. (2003). Developing a bidding agent for multiple heterogeneous auctions. ACM Trans. Internet Techn.. 3. 185-217. Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016).
2. Auction, <https://www.newworldencyclopedia.org/entry/Auction>, viewed 24.03.2020.
3. Klemperer, P., (1999). Auction Theory: A Guide to the Literature, Forthcoming, Journal of Economics Surveys, 1999.
4. McAfee R. P. ve diğ., (1987). Auctions and Bidding, Forthcoming, Journal of Economic Litarature, Vol. 25, No. 2 (Jun. 1987), 699-738.
5. Chaves A. S. (1997). KASBAH: An Agent-Based Marketplace For Buying and Selling Goods, Massachusetts Institute of Technology (Feb, 1997)
6. Srivastava Durgesh & Jain, Saket. (2016). Evaluating Negotiation Protocols and Negotiation Strategies for Automated E-Commerce:. International Journal of Engineering Research and. V5.
7. Srivastava, Durgesh. (2016). Auction System for Automated E-Commerce: JADE based Multi-agent Application. 5.
8. JADE Home Page <http://jade.tilab.com>, viewed: 24.03.2020.

ICATCES 2020

INTERNATIONAL CONFERENCE ON ADVANCED TECHNOLOGIES, COMPUTER ENGINEERING AND SCIENCE

WWW.ICATCES.ORG

ISBN: 978-605-9554-18-3