Bitlis Eren Üniversitesi
Fen Bilimleri Dergisi

Bitlis Eren University
Journal of Science

ISSN : 2147-3129
e-ISSN : 2147-3188

Cilt / *Volume*: 13 Sayı / *Number*: 4 Yıl / *Year*: 2024

Yazışma Adresi

Bitlis Eren Üniversitesi
Lisansüstü Eğitim Enstitüsü
13000, Merkez, Bitlis/TÜRKİYE
Tel: 0 (434) 2220071

https://dergipark.org.tr/tr/pub/bitlisfen

BEU FBD

Bitlis Eren Üniversitesi Fen Bilimler Dergisi

Bitlis Eren University Journal of Science

Yıl/*Year:* 2024 - Cilt/*Volume:* 13 - Sayı/*Number:* 4

Editör (Yayın) Kurulu / *Editorial Board*

Danışma Kurulu / *Advisory Board*

Dergi Hakkında / *About Journal*

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, mühendislik ve temel bilimler alanlarındaki gelişmeleri ve yenilikleri takip etmek, meslek kuruluşlarının, araştırmacıların ve bireylerin ulusal ve uluslararası gelişimlerine katkıda bulunmak ve bu alanlarda elektronik bir kaynak oluşturmak amacıyla yayımlanmaktadır. Derginin yazım dili Türkçe veya İngilizcedir. Fen Bilimleri Dergisi, Bitlis Eren Üniversitesi Lisansüstü Eğitim Enstitüsü yayını olup, 2012 yılından bu yana ücretsiz ve açık erişimli olarak yayın hayatına devam etmektedir. Mühendislik ve temel bilimlerin bilgi tabanına ve teknolojik gelişmelere ışık tutması amacıyla bu alanlarda yapılmış deneysel ve teorik ilerlemeleri konu alan özgün araştırma makalelerine yer verilmektedir. Dergiye gönderilen çalışmaların benzerlik oranı %25'i geçmemelidir. Yazım kurallarına uymayan makaleler, hakemlere gönderilmeden önce düzeltilmek üzere yazara geri gönderilir. Bu nedenle, derginin yazım kuralları dikkate alınmalıdır. Ayrıca, editörlerden yazarlara iletilen düzeltmelere veya taleplere 30 gün içerisinde cevap verilmediği takdirde ilgili makaleler reddedilir. Makaleler şekiller ve tablolar dâhil 20 sayfayı geçmemelidir. Dergiye yayın için gönderilen makaleler en az iki hakem tarafından değerlendirilir. Yazarlardan hakem önerisi talep edilmemektedir. Makalelerin dergide yayımlanabilmesi için hakemler tarafından olumlu görüş bildirilmesi gerekmektedir. Dergi Editör Kurulu, hakem raporlarını (*en aziki hakemin değerlendirmeleri geldikten sonra*) dikkate alarak makalelerin yayımlanmak üzere kabul edilip edilmemesine karar verir. Fen Bilimleri Dergisi, yılda dört defa (*Mart, Haziran, Eylül, Aralık*) yayımlanmaktadır. **Dergimiz Tübitak-Ulakbim Mühendislik ve Temel Bilimler Veri Tabanı Dergi Listesinde taranmaktadır.**

B.E.Ü. Fen Bilimleri Dergisi 13(4), 2024 / *B.E.U. Journal of Science 13(4), 2024*
İÇİNDEKİLER / *CONTENTS*

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 885-895 YEAR: 2024
DOI: 10.17798/bitlisfen.1432965

A New Approach to Automatic Detection of Tactile Coating Surfaces with Deep Learning

Abdil KARAKAN[1*]

[1]*Afyon Kocatepe University, Dazkırı Vocational School, Department of Electrical, Afyonkarahisar, Türkiye*
(ORCID: 0000-0003-1651-7568)

Keywords: Tactile coating, Deep learning, Blind, Real-time detection, YOLO.

Abstract
In this study, tactile coating surfaces of visually impaired individuals were detected using the deep learning method. For this detection, 4 of the You Only Look Once (YOLO) architectures, one of the best deep learning methods, were used. No ready data set was used in the study. A unique and new data set was prepared for the study. For the data set, 6278 images were taken from tactile coating surfaces. Images for real-time applications were obtained from many different environments. The tactile coating surfaces in the pictures were labelled separately. A total of 9184 tags were made. The dataset was implemented in YOLOv5, YOLOv6, YOLOv7, and YOLOv8 architectures. The highest accuracy was achieved in the YOLOv8 architecture with an accuracy rate of 97%, F1-Score of 0.940, and mAP@.5 of 0.977. The model was applied with k-fold cross-validation to evaluate performance measurements. In order for the study to be used in real-time, the frame per second (FPS) was increased to 150.

1. Introduction

According to World Health Organization data, 1 billion people in the world have vision problems. Of this, 123.7 million are unidentified visual defects, 65.2 million are cataracts, 6.9 million are glaucoma, 4.2 million are corneal opacity, 3 million are diabetic retinopathy, 2 million are chronic conjunctivitis, and 826 million are unidentified presbyopia-related visual disorders [1].

People with different degrees of visual impairment are closely affected by physical environmental conditions as well as economic and socio-cultural conditions. Urban space arrangements that do not meet the needs of visually impaired people often result in these people being confined to their homes. Visually impaired individuals can play an active role in society by providing limited vision with modern technology, the use of a cane or a guide dog [2]-[5].

Sensible surfaces were first developed by Seiichi Miyake in Japan in 1965. They are called “Tenji blocks” [6]. They were first used in the city of Okayama in 1967 [7]. Then they gradually spread to other cities. The low cost of tactile surfaces has made them considered the most effective system for the visually impaired. Sensible surfaces are basically of two types: stimulating blocks and guiding blocks [8], [9]. The first use of tactile surfaces was by arranging dome-shaped points parallel to each other. Afterward, different tactile surfaces arranged in a zigzag shape, oval shape and bar/line shape were used. Sensible surfaces, usually yellow or red, are preferred for the circulation of the visually impaired [10-12]. They are used on sidewalks, subways, underpasses, bus and train stations, and other public areas. Although most people think that they are used for design purposes, they play a major role in the daily lives of visually impaired individuals [13]-[20].

Many studies have been done to aid the circulation of visually impaired individuals. With the development of technology, systems that can detect obstacles in advance and guide disabled individuals have been designed [20]-[30].

Einloft aimed to create a system that can detect indoor spaces such as shopping malls,

*Corresponding author: akarakan@aku.edu.tr
Received: 06.02.2024, Accepted: 03.10.2024

subways, and bus terminals in order to help disabled individuals. The study was carried out in two stages. In the first stage; the gray-level co-occurrence matrix method used in image classification was used. In the second stage; Support vector machines, one of the supervised learning methods, were used. Although the initial results were promising, no numerical values regarding study accuracy were given [31].

Ghilardi has detected tactile parquet surfaces in real time to help visually impaired individuals. In this study, edge detection algorithms such as Canny, Lap-lace, and Sobel were used. In this study, they obtained a new approach by combining decision trees. There are 521 photographs in the created data set. Photographs were obtained with a smartphone attached to the person's waist at a height of one meter and at a 45-degree angle. There are 320 photographs of tactile parquet surfaces. There is no tactile parquet surface in the 201 photographs in the data set. In this study, the system operates at 16 FPS. The accuracy rate of the developed system is 88.48% [32].

Tactile parquet surfaces were detected using Jie image processing algorithms. Multicolor image segmentation and Kirsch edge detection algorithm were used in the system. Using the Hough transformation orientation, the tactile parquet surface was obtained as a strip. No numerical value is mentioned regarding the operating accuracy of the system, which has a 2 FPS operating speed and can be used portablely [33].

Asami and Ohnishi created a system that helps visually impaired individuals find themselves at a pedestrian crossing. The system consists of four parts. These consist of a USB camera integrated into the glasses, a vibrating wristband attached to the wrist, a notification system that gives warnings in the visually impaired alphabet integrated into the visually impaired cane, and a computer used to process the data coming from the USB camera. In this study, pedestrian crossings are detected using the normalized cross-correlation method in photographs taken from a USB camera. When a pedestrian crossing is detected, the user's wristband vibrates, allowing the user to stop in front of the pedestrian crossing. The developed system detects pedestrian crossings with 86.7% accuracy [34], [35].

Aktaş and his colleagues detected tactile parquet surfaces in real-time in order to help visually impaired individuals. In their study, they used YOLOv2, YOLOv3, and YOLOv3 DenseNet architectures. They prepared the data set themselves. Photographs were obtained from places that are frequently used in daily life, such as libraries, subways, parks, walking paths, and streets in various districts and regions of Istanbul. The captured photographs were converted to 1024 × 768 in order to be used quickly and conveniently in the system. The larger the data set, the higher the accuracy rate of the study. For this purpose, a data augmentation process was applied. The data set initially consists of 1200 photographs. Then, the data set was replicated with the OpenCV Library. As a result of the data duplication process, 4580 data points were accessed in the system. During the data duplication process, operations such as mirroring, whitening, and rotation were performed. The dataset was used individually in three different YOLO architectures. The results were compared. The best results were achieved with the YOLOv3 DenseNet architecture. It has been understood that YOLOv3 DenseNet architecture is better than other models in detecting tactile parquet surfaces with 89% F1-score, 92% average sensitivity and, 81% IoU values [36], [37].

As a result of the literature review, it is understood that the use of previously made systems in daily life is quite difficult. These difficulties are systems with low FPS values. Real-time detection cannot be made with a low frame rate. In addition, previous studies have carried out the detection of tactile parquet surfaces both indoors and outdoors with only one system. Shopping mall, metro, etc. A separate system is created for interior spaces such as walkways and streets, and a separate system for exterior spaces such as walkways and streets. Therefore, the user must use extra equipment such as a computer, cane, and wristband in his backpack to use the system. In the studies conducted, there are problems such as the inability to complete the studies due to lack of data. Unlike the studies in the literature, the system used in this study performed real-time detection at 150 frames per second. With this system, tactile parquet surfaces were detected both indoors and outdoors without the need for extra equipment such as canes, wristbands, or computers. For this reason, a lot of photographs were used while preparing the data set. The photographs that make up the data set were collected from different places and at different times. Thus, a data set closer to real life was created. With the created data set, a system that works very well in real life has been realized.

In this study, a model that can detect tactile parquet surfaces in photographs, videos, and real-time operating systems that can be used in assistive technologies for individuals with visual impairments was developed. Considering the applicability and suitability of deep learning methods in the field of image processing, deep learning methods and image processing algorithms were mainly used in this study. When it comes to object detection based on deep learning, the YOLO method gives very good results

in real-time detection. A data set consisting of 6278 labeled photographs was used as the data set within the scope of this study. In the experiments conducted on the data set, tactile parquet surfaces were detected with an F1 score of 89%, an average sensitivity of 92%, and an IoU value of 81%. The study has three important contributions. In this study carried out, a new and large data set was created in this field, and tactile parquet surfaces were detected by using the YOLO method on this data set without depending on a specific camera angle. A new model with high-performance values has been developed. Deep learning methods have been used in the field of tactile parquet surface detection.

2. Material and Method

In this study, a model that can detect tactile parquet surfaces in photographs, videos, and real-time operating systems that can be used in assistive technologies for individuals with visual impairments was developed. The study was carried out in three stages. In the first stage, the data set was prepared. For this purpose, tactile coating surfaces found in real life were used. Thus, when the study was implemented, a higher rate of compliance was achieved in real life. In the second stage, the data set duplication process was applied. As the data set increased, the accuracy rate of the system increased. The highest accuracy was achieved by using the created data set in four different architectures. Since the work will be real-time, the FPS rate must also be very good. In the last stage it has been implemented in real life. Figure 1 shows the general structure of the study carried out for the detection of tactile coating surfaces.

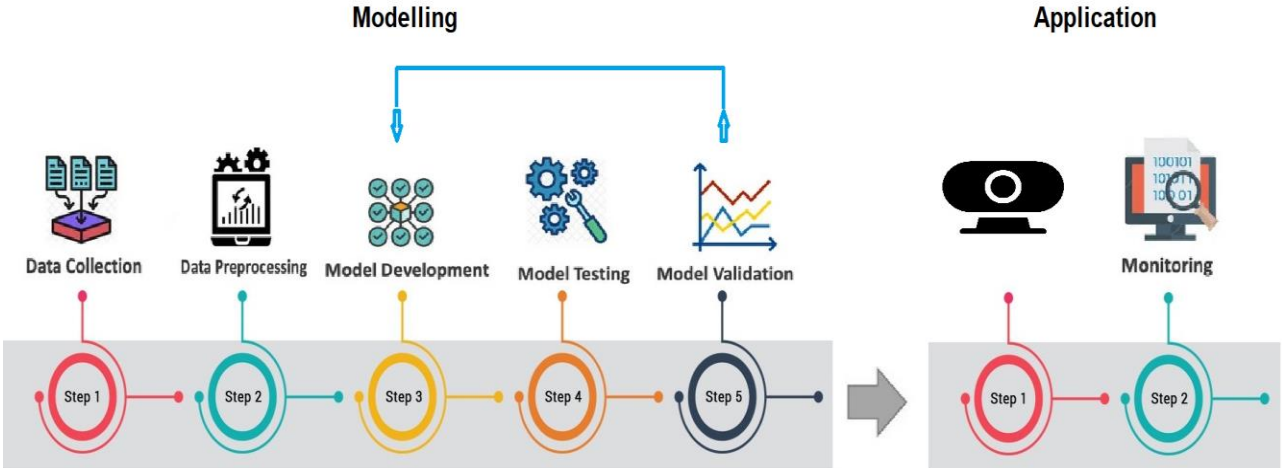


Figure 1. The general structure of the study was carried out for the detection of tactile coating surfaces.

2.1. Data Set

Tactile surface indicators should be arranged to indicate access to facilities along the route and warn of possible dangers. Among these indicators, the one in the form of a bar/line is called the guiding surface, and the one with dots is called the warning surface. Indicators in the form of bars/lines provide guidance in finding direction. Dotted indicators serve as warnings by placing them in potentially dangerous areas such as stairs, pedestrian crossings, and train/subway platforms. In short, bar/line indicators mean “go” and dotted indicators mean “stop” [38], [39].

While preparing the data set, tactile surface samples of different types and from different environments were taken. For this purpose, 6278 photographs were taken. In these photographs, 9814 tactile surfaces were labeled. Figure 3 shows 9 different tactile surface examples from the data set. Since real-time detection will be made in this study, photographs were taken from as real environments as possible. Figure 2 shows examples of tactile surface parquet.

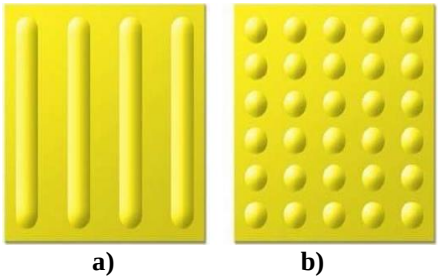


Figure 2 a) Guiding tactile surface **b)** Stimulating tactile surface.

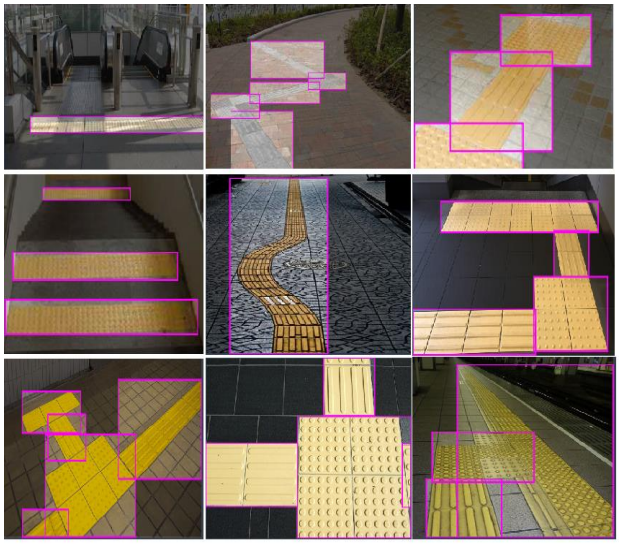


Figure 3. Labeling of different photographs used in the data set.

2.2. YOLO Architecture

Deep Learning is one of the sub-topics of the field of machine learning and is also the most current topic. Deep Learning methods are used as artificial neural network algorithms inspired by the structure of the human brain. Deep Learning is part of a broader family of machine learning methods based on learning data representations as opposed to task-specific algorithms. Deep Learning, along with studies in the field of machine learning, has also pioneered the expansion of the scope of artificial intelligence. Compared to superficial learning, the concept of deep learning has the advantage of building deep layers to reach more abstract information. Deep learning-based methods are used successfully in many fields. One of these is automatic object detection. In this study, tactile coating surfaces were detected with YOLOv4, YOLOv5, and YOLOv6, which is a version of the deep learning-based object detection method YOLO algorithm.

YOLO is a deep learning-based object recognition algorithm developed by Joseph Redmon. YOLO performs object detection using a single neural network to estimate class probabilities along with bounding boxes. YOLO's architecture is based on a convolutional neural network and has a fast structure [40].

YOLO treats object detection as just a regression problem, using a single neural network. In addition, YOLO also uses a single neural network framework to estimate bounding boxes and class probabilities. While the detection process is performed, the YOLO algorithm estimates the class and coordinates of all objects in the photo by dividing the relevant photo into grids. Thus, object detection is treated as a single regression problem. YOLO can directly calculate location coordinates and classification of objects without intermediaries. It processes photographs at 150 frames per second, thus providing users with real-time object detection. In this way, it has faster prediction power than other methods. This means that in the area where the object will be detected, it passes the object it needs to detect through the network only once and completes the detection process. Considering the literature, it is the fastest general-purpose object detection method [41].

YOLOv4 was a real-time object detection model released in April 2020 and achieved the highest performance on the COCO dataset. It works by dividing the object detection task into two parts; Bounding boxes are used to determine the class of the object. Secondly, the Regression method is used to describe the position of the object through classifiers. The Yolov4 application uses the Darknet framework. Using YOLOv4, many new contributions have been added to the YOLO family alongside previous research contributions, including new features specific to YOLOv4: CSP, WRC, SAT, CmBN, Mish activation, Mosaic data augmentation, DropBlock, CmBN orchestration, and Like CIoU loss. In summary, better object detection network architecture and new data development techniques are used with YOLOv4 [42].

YOLOv5 is a model in the YOLO family of computer vision models. YOLOv5 is widely used to detect objects. YOLOv5 comes in four main versions: small (s), medium (m), large (l), and extra-large (x), each offering increasingly higher accuracy rates. Each variant also takes a different amount of time to train. In the graph, the goal is to create a high-performance object detection model (Y-axis) based on extraction time (X-axis). Preliminary results show that YOLOv5 performs very well compared to other state-of-the-art techniques for this purpose. YOLOv5 variants run faster than Effi-cientDet. As the most accurate YOLOv5 model, YOLOv5x can process photographs much faster with the same level of accuracy as the Effi-cientDet D4 model. Although YOLOv5 derives most of its performance improvements from PyTorch's training routines, its modeling architecture is still close to YOLOv4 [43].

The YOLOv6 model was released by Meituan in June 2022. It gave good results in the COCO dataset comparison. The YOLOv6 model is built on the basis of the YOLO architecture and offers various improvements and new methods over other models of the YOLO family. YOLOv6 is written in PyTorch. YOLOv6 has three major updates: a hardware-friendly spine, neck design, and dedicated head for efficiency, and effective training strategies.

YOLOv6's object detection performance has been shown to be comparable to other CNN-based algorithms, with improvements in both speed and accuracy with incremental versions of the algorithm. In this study, the effect of different image sizes on the performance of YOLOv6 subarchitectures is shown. The data set used for the YOLOv6 model in this study was derived into two versions with dimensions of 416x416 and 640x640 [44].

YOLOv7 has a fast and powerful network architecture. It achieves this with four features it has. These features, integration method, increased label assignment, and model training efficiency. YOLOv7 requires much less computing hardware than other deep learning models. With YOLOv7, it can be trained quickly with small datasets even without pre-trained weights. These features make YOLOv7 a more effective model. Object tracking algorithms generally divide the image into regions, select possible regions containing objects, and classify each region separately, which increases the processing load [45].

YOLOv8 is equipped with more advanced post-processing techniques than its previous versions. These techniques are applied to estimated bounding boxes and object scores produced by the YOLOv8 neural network. With YOLOv8, it refines detection results, removes unnecessary detections, and increases the overall accuracy of predictions. These techniques include Soft-NMS, a variant of the non-maximum suppression (NMS) technique used in YOLO architectures. Instead of completely deleting overlapping bounding boxes, Soft-NMS applies a soft threshold to them. YOLOv8 passes the image through the convolutional ANN at once and performs group normalization. This provides a frame processing speed that can vary from 5 FPS to 160 FPS [46].

3. Results and Discussion

A computer with an AMD Ryzen 1500X 3.5 GHz processor was used in this study. NVIDIA GeForce GXT 1050Ti 4GB GDDR5 was preferred as the graphics card. Memory speed is 16 GB 3000 MHz. Firstly, the YOLOv8 dataset architecture was used. The training period for detecting tactile coating surfaces on the YOLOv8 architecture lasted 4 hours, 49 minutes, and 15 seconds. For tactile-coated surfaces, the highest accuracy rate was 97.2% on the stimulating surface and 98.9% on the guiding surface. The average accuracy rate was 97.9%. Table 1 shows the results occurring in the YOLOv8 architecture.

Table 1. Results of YOLOv8 architecture.

Class	Precision	Recall	F1-Score	mAP@.5	mAP@.5:.95
All	0.979	0.905	0.940	0.977	0.789
Stimulating	0.972	0.914	0.942	0.973	0.719
Guiding	0.989	0.896	0.940	0.980	0.859

The guiding and stimulating surfaces on tactile coating surfaces have different accuracy rates. This is due to differences between tactile coating surfaces. The average performance values of the YOLOv8 architecture were Precision 97.9%, Recall 90.5%, F1-Score 0.940, mAP@.5 0.977, mAP@.5:95: 0.789. Figure 4 shows the confusion matrix implemented in the system.

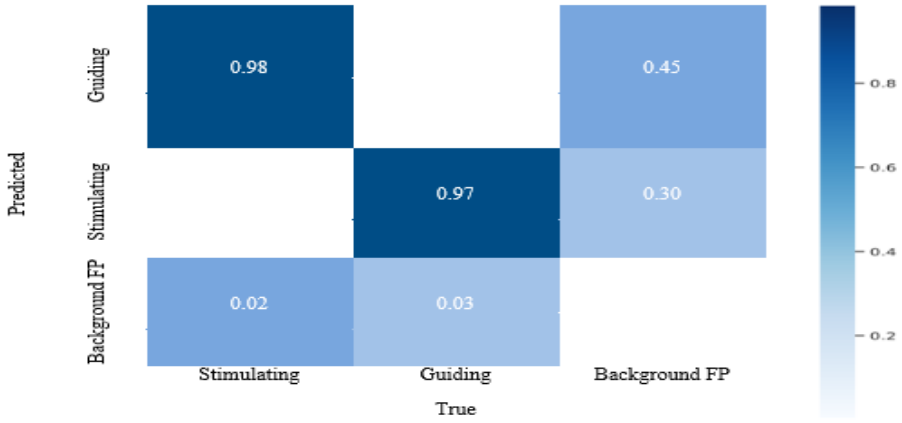


Figure 4. Confusion matrix resulting from YOLOv8 architecture.

A confusion matrix application was carried out to evaluate the performance of this study. The performance of the classification algorithm is visualized with the confusion matrix. For the confusion matrix, 406 images were used for detection. Figure 5 shows the confusion matrix results.

As a result of the confusion matrix, the average accuracy was determined as 95.50%. The confusion matrix results were very close to the simulation results. The confusion matrix results and simulation results show that the study is suitable for the real environment.

In this study, YOLOv5, YOLOv6, and YOLOv7 architectures were used as deep learning. The same data set was used in all studies. Thus, the error rate is minimized. The results for the YOLOv5, YOLOv6, and YOLOv7 architectures used in this study were shown in Figure 6.

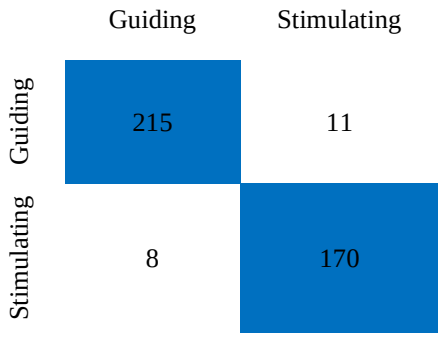


Figure 5. Confusion matrix result.

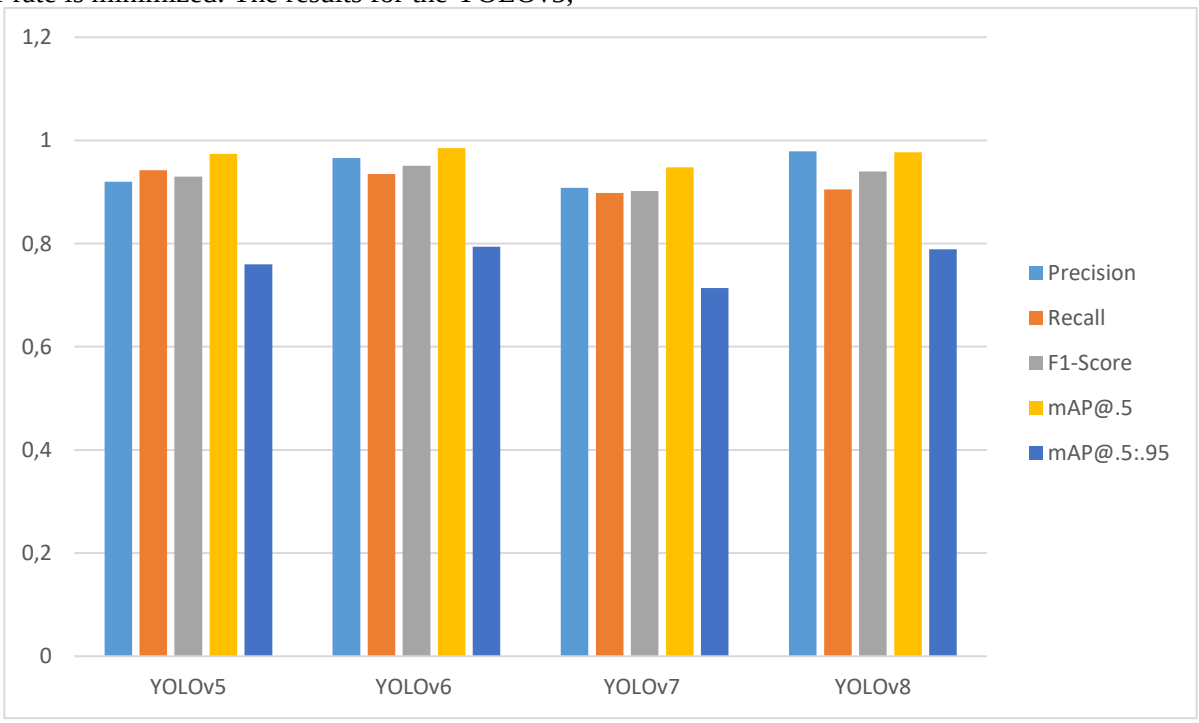


Figure 6. YOLOv5, YOLOv6, and YOLOv7 architectures results.

When the accuracy was examined in this study, the highest accuracy was obtained at 97.9% in the YOLOv8 architecture. The lowest accuracy was 90.8% in the YOLOv7 architecture. The second and third accuracy rates were 92% in the YOLOv5 architecture and 90.8% in the YOLOv7 architecture. When F1 scores are examined, the highest rate occurred in the YOLOv6 architecture. It was later implemented in YOLOv8, YOLOv5, and YOLOv7 architectures.

Since the work will be real-time, the FPS speed must be sufficient. In the YOLOv8 architecture, which gives the highest accuracy rate, the speed has increased up to 150 FPS. At this rate, it is sufficient for real-time detection. Real-time detection was carried out with photographs taken from the camera.

K-Fold Cross Validation method: It divides the data set into “k” equal parts and creates validation data for each part one by one. Thus, each data point is used as validation data at least once. In this way, the overall performance of the model is evaluated more accurately. There are 6278 samples in the data set of the study. Here, the number k is determined as 5. In other words, the data set is divided into 5. There are 1255 samples in each dataset. Their distribution is distributed in the same proportions as in the general data set. Figure shows the results after the K-Fold Cross Validation method.

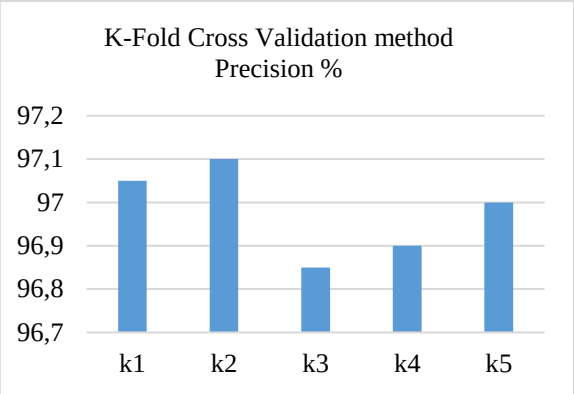


Figure 7. K-Fold Cross Validation method.

The average accuracy rate of the results obtained after the K-Fold Cross Validation method was 97.30. With this result, the overall result of the system is equal to the two results checked. Thus, it is understood that the overall performance of the model is good.

In Figure 8, tactile coating surfaces have been determined in 9 different ways. To ensure that the system adapts to real life, it has been implemented at different times and in different environments.

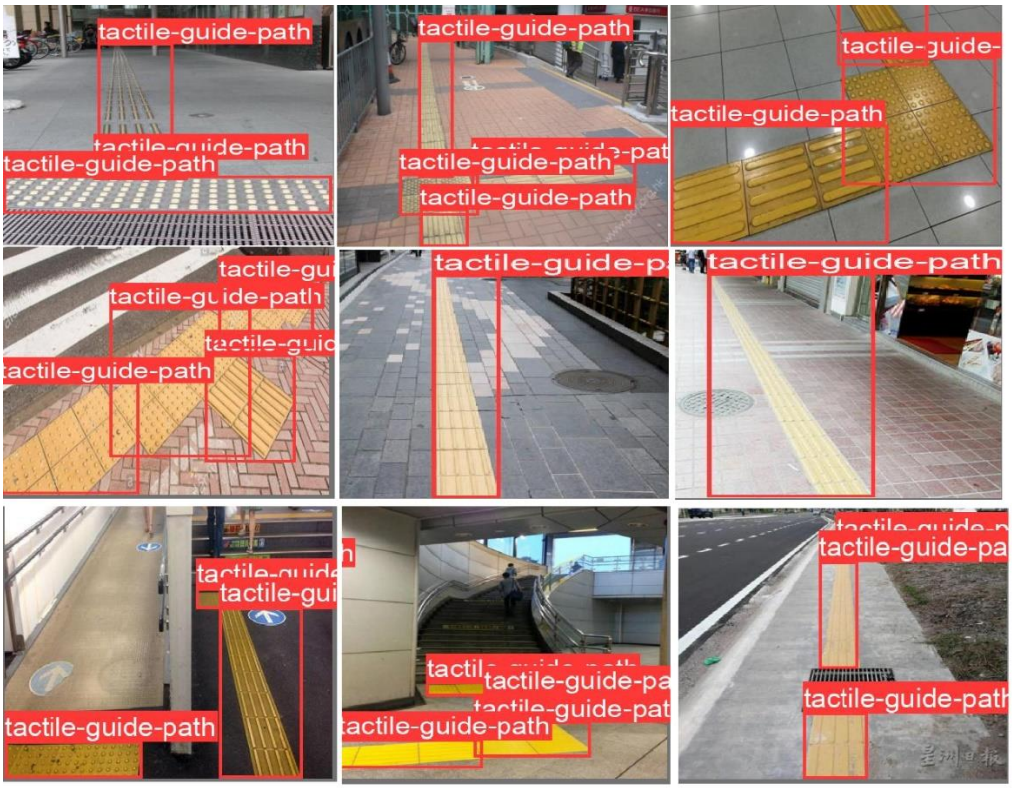


Figure 8. Detection of different tactile-coated surfaces.

When the literature studies are examined, many studies have been done. In these studies, ready-made data sets were not always used. In some studies, the most appropriate data set for the system to be prepared was prepared. Table 2 shows the comparison of literature studies.

Table 2. Comparison of literature studies.

Writer	FPS	Number of Image	Architectural	Architectural
Ghilardi et all [22]	16.27	7000	Faster R-CNN	%88.48
Einloft et all [21]	-	8766	Faster R-CNN	-
Jie et all [23]	2	1586	Faster R-CNN	-
Redmon et all [33]	70	6772	YOLOv3-Tiny	%71
Redmon et all [34]	60	350	Dronet	%80
Asami et all [24]	50	32	YOLOv5	%92
Wang et all [30]	60			%92

As a result of the literature review, it is understood that the usability of previous systems in daily life is quite difficult. These difficulties are that they have a low FPS value. For this reason, the systems made cannot be used in real-time. The systems made have focused only on the tactile coating in a certain place. In this case, different tactile coatings cause the surfaces to be detected at a very low rate or not detected at all.

In this study, four of the latest versions of the YOLO architecture were used. In this way, the highest accuracy rate and FPS speed have been determined by making a comparison. FPS speeds are very good because YOLO architectures take only one look at the image and detect it. The speed of 150 FPS has been reached in the YOLOv8 algorithm. In this way, the detection process can be performed very easily in real-time.

4. Conclusion and Suggestions

This study was not prepared for just one location. It is prepared according to the environments you will encounter in real life. Ready-made data sets were not used for this. A unique data set was prepared for the study. Images from many different environments were used for the data set. Each tactile coating surface in these images is individually labeled. Therefore, it was more easily applied in the real environment.

Additionally, YOLO architectures, the best deep learning method for real-time detection, were used. To achieve the best results in these architectures, YOLOv5, YOLOv6, YOLOv7, and YOLOv8 architectures were used. The highest result was 97% accuracy in the YOLOv8 architecture. A k-fold cross validation process was applied to measure the performance evaluation of the study. Ultimately, it proved the performance of the model. In order for this study to work in real time, the FPS speed must be good. In the study, the FPS rate was increased up to 150 with the YOLOv8 architecture.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] World Health Organization, Blindness and vision impairment, https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. (30. 01. 2024).

[2] J. Lu, K. Siu, and P. Xu, "A comparative study of tactile paving design standards in different countries," *9th International Conference on Computer-Aided Industrial Design and Conceptual Design*, Kumning, China, 753-758, 22-25 Nov. 2008. https://doi.org/10.1109/CAID/CD15004.2008

[3] A. Mancini, E. Frontoni, and P. Zingaretti, "Mechatronic System to Help Visually Impaired Users During Walking and Running," *IEEE Transactions on Intelligent Transportation Systems*, 2018, 19(2), 649-660. https://doi.org/10.1109/TITS.2017.2780621

[4] J. Redmon, and A. Farhadi, "Yolov3: An incremental improvement," *Computer Vision and Pattern Recognition*, 2018, 1-15. https://doi.org/10.48550/arXiv.1804.02767

[5] G. Huang, Z. Liu, and L. Maaten, "Weinberger, K. Q. Densely Connected Convolutional Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii-USA, 2261-2269, 21-26 July 2017. https://doi.org/10.5555/2149960

[6] A. Garcia-Garcia, S. Orts, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Rodríguez, "A Survey on Deep Learning Techniques for Image and Video Semantic Segmentation," *Applied Soft Computing*, 2018, 70, 41-65. https://doi.org/10.1016/j.asoc.2018.05.018

[7] A. Krizhevsky, I. E. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems*, 2012, 25(2). https://doi.org/10.1145/3065386

[8] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and F. Li, "Ieee. ImageNet: A Large-Scale Hierarchical Image Database," *IEEE-Computer-Society Conference on Computer Vision and Pattern Recognition Workshops*, Miami Beach, FL, USA, 248-255, 20-25 Jun. 2009. https://doi.org/10.1109/CVPRWorkshops15504.2009

[9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, and F. F. Li, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, 2015, 115, 211-252. https://doi.org/10.48550/arXiv.1409.0575

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *Proceedings of the IEEE international conference on computer vision*, Massachusetts, USA, 1026-1034, 7-13 Oct. 2015. https://doi.org/10.1109/ICCV33071.2015

[11] M. Tan, and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *2019 International Conference on Machine Learning*, California-USA, 6105-6114, 14-16 February 2019. https://doi.org/10.1109/COMITCon45641.2019

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, Columbus, OH, USA, 580-587, 25 Sept. 2014. https://doi.org/10.1109/CVPRW34339.2014

[13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A.C. Berg, "Ssd: Single shot multibox detector," *European conference on computer vision(ECCV2016*), Amsterdam,Netherlands, 21-37, 11-14 Oct. 2016. https://doi.org/10.1007/978-3-319-46466-4

[14] A. Kaya, A.S. Keçeli, and A.B Can, "Examination of various classification strategies in classification of lung nodule characteristics," *Journal of the Faculty of Engineering and Architecture of Gazi University*, 2019, 34(2), 709-725. https://doi.org/10.17341/gazimmfd.416530

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, 779-788, 27-30 June 2016. https://doi.org/10.1109/CVPR33180.2016

[16] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks", *International Conference on Learning Representations*, Scottsdale, Arizona, USA, 2-4 May 2013. https://doi.org/10.48550/arXiv.1312.6229

[17] Y. Du, N. Pan, Z. Xu. F. Deng, Y. Shen, and H. Kang, "Pavement distress detection and classification based on YOLO network," *International Journal of Pavement Engineering*, 2020, 1659-1672. https://doi.org/10.1080/10298436.2020.1714047

[18] J. Shen, N. Liu, H. Sun, X. Tao, and Q. Li, "Vehicle Detection in Aerial Photographs Based on Hyper Feature Map in Deep Convolutional Network," *KSII Transactions on Internet & Information Systems*, 2019, 13(4).No:4. https://doi.org/10.3837/tiis.2019.04.014

[19] K. Arai, Y. Ando, and M. Mizukawa, "Development of signal recognition improvement system in crosswalk", *Jsme Robomec*, 2008, pp.2A1- E12(1)-(2). https://doi.org/10.1299/jsmermd.2008. 2A1-I03_1

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 39,No:6. https://doi.org/10.1109/TPAMI.2012.205

[21] D. C. Einloft, M.C. Ghilardi, and I. H. Manssour, "Automatic Detection of Tactile Paving Surfaces in Indoor Environments", *Workshop of Undergraduate Works (WUW) in the 29th Conference on Graphics, Patterns and Photographs*, Sau Paul, Brazil, 4-7 Oct. 2016. https://doi.org/10.1109/SIBGRAPI.2016.001

[22] M. C. Ghilardi, R.C.O Macedo, and I. H. Manssour, "A New Approach for Automatic Detection of Tactile Paving Surfaces in Sidewalks", *Procedia Computer Science*, 2016, 80, 662-672. https://doi.org/10.1016/j.procs.2016.05.356

[23] X. Jie, W. Xiaochi, and F. Zhigang, "Research and implementation of blind sidewalk detection in portable eta system", *International Forum on Information Technology and Applications*, Kunming, China, 431-434, 16-18 July 2010. https://doi.org/10.1109/IFITA.2010.360

[24] T. Asami, and K. Ohnishi, "Crosswalk location, direction and pedestrian signal state extraction system for assisting the expedition of person with impaired vision", *10th France-Japan/8th Europe-Asia Congress on Mecatronics*, Tokyo, Japan, 285-290, 27- 29 Nov. 2014. https://doi.org/10.1109/MECATRONICS33801.2014

[25] A. Kassim, T. Yasuno, M. S. Mohd, A. Hjshukor, H.I. Jaafar, F. Baharom, and F. Jafar, "Vision based of tactile paving detection method in navigation system for blind person", *Jurnal Teknologi*, 2015, 77, 20. https://doi.org/10.5772/intechopen.79886

[26] O. Yildiz, "Melanoma detection from dermoscopy photographs with deep learning methods: A comprehensive study", *Journal of the Faculty of Engineering and Architecture of Gazi University*, 2019, 34(4), 2241-2260. https://doi.org/10.17341/gazimmfd.435217

[27] E. Dandil, and R. Polattimur, "Dog Behavior Recognition and Tracking based on Faster R-CNN", *Journal of the Faculty of Engineering and Architecture of Gazi University*, 2019, 35(2), 819-834. https://doi.org/10.17341/gazimmfd.541677

[28] J.J. Lv, X.H. Shao, J.S. Huang, X.D. Zhou, and X. Zhou, "Data augmentation for face recognition". *Neurocomputing*, 2017, 230, 184-196. https://doi.org/10.1016/j.neucom.2016.12.025

[29] S. Longfei, Y. Zheng, Z. Zimu, Z. Xiaolong, W. Chenshu, and L. Yunhao, "Crossnavi: Enabling real-time crossroad navigation for the blind with commodity phones", *In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 14*, pages 787–798, New York, USA, 1-4 September 2014. https://doi.org/10.1145/2632048.2632083

[30] J. Wang, N. Wang, L. Li, and Z. Ren, "Real-time behavior detection and judgment of egg breeders based on YOLOv3", *Neural Computing and Applications*, 2019, 32, 2. https://doi.org/10.1007/s00521-019-04645-4

[31] W. Byung, Y. Sung, and J. Kang, "Brick path detection from shape pattern and texture feature", *In System Integration (SII), 2011 IEEE/SICE International Symposium on*, Kyoto, Japan, 78–83, 1 Dec 2011. https://doi.org/10.1109/SII.2011.6147423

[32] S. Shoval, J. Borenstein, and Y. Koren, "The NavBelt-a computerized travel aid for the blind based on mobile robotics technology", *IEEE Transactions on Biomedical Engineering*, 1998, 45(11), 1376-1386. https://doi.org/10.1109/10.725334

[33] J. Redmon, and A. Farhadi, "YOLO9000: Better, Faster, Stronger", *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii-USA, 6517- 6525, 21-26 July 2017. https://doi.org/10.1109/CVPR35066.2017

[34] C. Goutte, and E. Gaussier, "A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation", *Proceedings of the 27th European conference on Advances in Information Retrieval Research, Spain*, 345-359, 21-23 March 2005. https://doi.org/10.5555/2149960

[35] R. Redmon, and A. Farhadi, "Yolov3: An incremental improvement", *Computer Vision and Pattern Recognition*, 2018, 1-6. https://doi.org/10.1007/s11263-009-0275-4

[36] R. Pyun, Y. Kim, P. Wespe, and R. Gassert, "Advanced augmented white cane with obstacle height and distance feedback", *IEEE Int. Conf. Rehabil. Robotics*,Seattle, WA, USA, pp.1-6. 24-26 June 2013. https://doi.org/10.1109/ICORR21102.2013

[37] A. M. Kassim, M. H. Jamaluddin, M. R. Yaacob, N. S. N. Anwar, Z. M. Sani, and A. Noordin, "Design and development of MY 2nd EYE for person with impaired vision", *IEEE Symposium on Computers & Informatics*, Kuala Lumpur, Malaysia, pp.700-703. 20-23 March 2011. https://doi.org/10.1109/ISCI17961.2011

[38] D. Ahmetovic, C. Bernareggi, A. Gerino, and S. Mascetti, "Zebrarecognizer: Efficient and precise localization of pedestrian crossings", *In Pattern Recognition (ICPR), International Conference on Pattern Recognition*, Stockholm, Sweden, 2566–2571, 24-28 Aug 2014. https://doi.org/10.1109/ICPR33632.2014

[39] A. M. Kassim, H .İ. Jaafar, M. A. Azam, N. Abas, and T. Yasuno, "Design and development of navigation system by using rfid technology", *In System Engineering and Technology (ICSET), IEEE International Conference on System Engineering and Technology*, Shah Alam Malaysia, pages 258–262, 19-20 Aug. 2013. https://doi.org/10.1109/ICSEngT.2013.6650181

[40] K. W. M. Jiangyan, and X. Ping, "A comparative study of tactile paving design standards in different countries", *International Conference on Computer-Aided Industrial Design and Conceptual Design*, Kunming, China, pages 753–758, 22-25 Nov. 2008. https://doi.org/10.1109/CAID/CD15004.2008

[41] V. N. Murali, and J. M. Coughlan, "Smartphone-based crosswalk detection and localization for visually impaired pedestrians", I*n Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, San Jose, California, USA, pages 1–7, 15-19 July 2013. https://doi.org 10.1109/ICMEW32311.2013

[42] H. Shuihua, P. Hangrong, Z. Chenyang, and T. Yingli, "Rgb-d image-based detection of stairs, pedestrian crosswalks and traffic signs", *Journal of Visual Communication and Image Representation*, 2014, 25(2): 263-272. https://doi.org/10.1016/j.jvcir.2013.11.005

[43] W. Shuihua, and T. Yingli, "Detecting stairs and pedestrian crosswalks for the blind by rgbd camera", *In Bioinformatics and Biomedicine Workshops (BIBMW), 2012 IEEE International Conference on*, Philidelphia, USA, 732–739, 4-7 Oct. 2012. https://doi.org/10.1109/BIBMW.2012.6470197

[44] C. Shorten, and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning", *Journal of Big Data*, 2019, 6(1), No: 60. https://doi.org/10.1186/s40537-019-0197-0

[45] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new stateof-the-art for real-time object detectors", *Computer Vision and Pattern Recognition*, 2022, 1-15. https://doi.org/10.48550/arXiv.2207.02696

[46] B. Hu, M. Zhu, L. Chen, L. Huang, P. Chen, and M. He, "Tree species identification method based on improved YOLOv7", *IEEE 8th International Conference on Cloud Computing and Intelligent Systems*, 622-627, Chengdu, China, 26-28 November 2022.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 896-904 YEAR: 2024
DOI: 10.17798/bitlisfen.1458956

Design, Fabrication, and Piezoelectric Performance Evaluation of a Nanogenerator for Vibrational Energy Harvesters

Levent PARALI[1*]

[1]*Manisa Celal Bayar University, Electronics & Automation Department, Turgutlu Vocational School, Manisa / Türkiye*
(ORCID: 0000-0002-4462-7628)

Keywords:
PVDF, PZT, Graphene, Electrospinning, Piezoelectric Nanogenerator Manufacturing, Energy Harvesting.

Abstract
In recent years, to provide power for wearable electronics, the mechanical energy obtained from environmental conditions through the piezoelectric nanogenerator into electricity has attracted interest. In this study, polyvinylidene fluoride (PVDF), lead zirconium titanate (PZT), and graphene nanoplatelets (GNP) based piezoelectric nanogenerators (PENs) were fabricated using electrospinning method. The experimental results evaluated using Thevenin's, Norton's, and the maximum power transfer theorems exhibited that the PVDF/PZT/GNP-based PEN had a 2.76 times greater electrical power efficiency (0.24 μW) at the resonance frequency of 20 Hz compared to that of the PEN based on the pure PVDF (0.09 μW) at the vibrational frequency of 25 Hz. The piezoelectric energy harvesters are highly suitable as self-powered wearable motion sensors because of the direct relationship between the vibration frequency and the generated output power.

1. Introduction

Although batteries are among the traditional power sources, they are limited in terms of power efficiency, energy storage capacity, power efficiency, and lifetime and require recurrent recharging. Therefore, recent advances in various small, mobility portable, remote-sensing, and low-power devices have increased the need for unconventional power sources [1]- [3]. As a result, in recent years there has been a significant increase in studies investigating energy harvesting devices as a preferred primary power source to bring down battery recharging and replacement costs and extend device life [4], [5]. For electronic devices used in sensor points [6], biomedical devices, health monitoring systems [7], and large-scale sensor networks [8], [9] that detect interactions at a distance, the use of energy harvesting devices can extend the lifetime of electronic devices and provide alternative specialized applications. Energy harvesting can be explained as the direct conversion of environmental energy sources such as mechanical, thermal, solar, bioenergy, hydropower, and wind into electrical energy using a specific material or transmission mechanism.

In order to convert human movement [10], mechanical vibration [11], water flow [12], etc. into electrical energy, piezoelectric nanogenerators commonly used in energy harvesting systems offer a simple and innovative approach. Flexible polymer piezoelectric nanogenerators (PENs) have been widely built as structured in different forms using polyvinylidene difluoride (PVDF) and its copolymers. The PVDF has five crystalline types (α, β, γ, δ, and Ɛ) depending on its chain conformation, the β and γ phases among crystalline types provide an advantage for the fabrications of actuators, sensors, batteries, and filters due to the fact that they have the greatest electrically active phases [13]-[14]. In order to enhance the β-phase transformation of PVDF, several approaches such as increasing the phase transition by applying a strong electric field [15], mechanical stretching [16], and the addition of nucleating fillers and particles [17], [18], etc. have been used in the literature.

Since the PZT ($Pb\ (Zr_xTi_{(1-x)})\ O_3$) ceramic particles have higher piezoelectric values and

*Corresponding author: levent.parali@cbu.edu.tr
Received: 26.03.2024, Accepted: 10.10.2024

mechanical properties, they are widely used in the fabrication of the PVDF/PZT-based piezoelectric nanogenerator [19], [20]. Recently, in order to further strengthen the connection between randomly dispersed ceramic particles in the polymer matrix (three-phase composite) and to improve the PVDF's β-phase, there has been a focus on the incorporation of some amount of filler (third conductive phase) such as carbon nanotubes (CNTs) [21], graphene particles (GNPs) [22], hydrated ionic salts [23], halides [24] into the polymer matrix.

This study aims to introduce the production of electrospun mats obtained through the electrospinning process, to show how they can be designed as a piezoelectric nanogenerator, and to determine their electrical performances using the vibrational energy harvesting system.

2. Material and Method

2.1. Materials

Polyvinylidene difluoride (PVDF) (MW: 534.000 g/mol, Alfa Aesar) was chosen as a polymer substrate while a PZT-5A (Navy Type II / APC 850, a particle size of < 1 μm) and Graphene Nanoplatelets (Nova Scientific Resources-M Sdn. Bhd.) were decided on doping materials respectively. Furthermore, N-dimethylformamide (DMF, Merck) (99.5%) and acetone (C3H6O, Merck) were used as chemical solvents.

2.2. Production of Electrospun Mat Through Electrospinning

The production stages of all electrospun mats are depicted in detail in Figure 1. The electrospinning method is a voltage-controlled producing process controlled by a specific electro-hydrodynamic phenomenon, to produce thin fiber mats from a polymer containing solutions. In this study, the initial sample consisted of pure PVDF, while the content of the second composite was built of the PVDF incorporated with the PZT and the pristine GNPs.

For the first specimen, the pure PVDF solution was constructed by dissolving PVDF powder (10% w/v) in a solvent consisting of a mixture of DMF and Acetone (DMF: 4.5 ml, Acetone: 10.5 ml) at 50 °C for 2 hours with magnetic stirring. On the other side, in order to dissolve GNPs in a weight ratio of 1.5wt.% in acetone, an ultrasonic mixing process was utilized for 3 hours, and then the PZT particles (15vol.%) were added to the GNP solution. All mixed solutions were stirred as additional by an ultrasonic and a magnetic stirrer for 5 hours again. Finally, the second sample solution was obtained by transferring the homogeneous PZT/GNP suspension to the PVDF solution and additional stirring for 5 h with a magnetic stirrer.

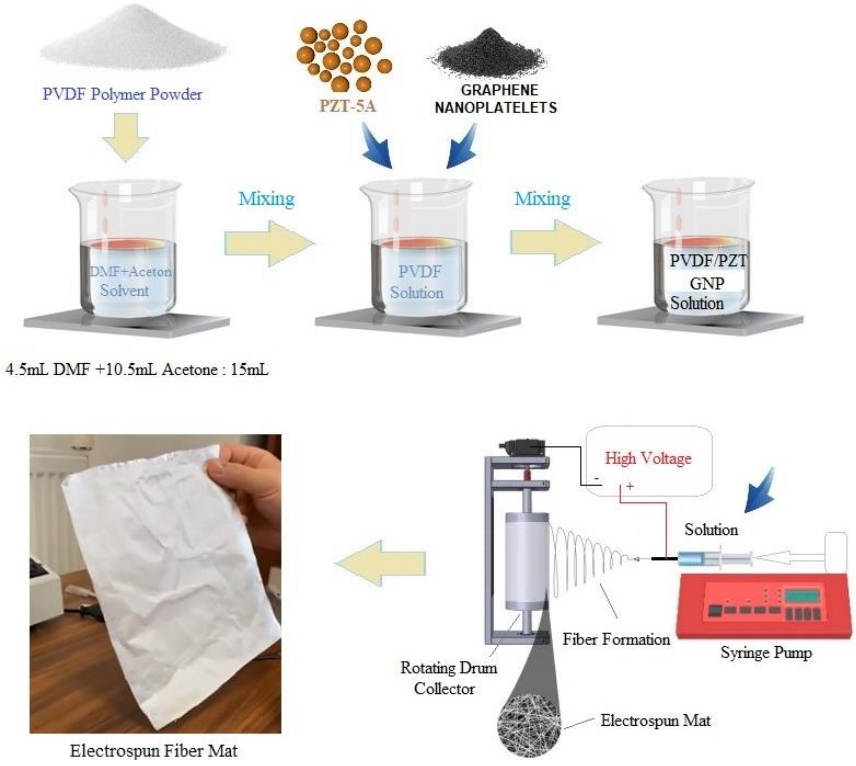


Figure 1. Production of the electrospun fiber mats using Electrospinning.

Consequently, the pure PVDF and the PVDF/PZT/GNP solutions were filled into the plastic syringe (diameter of syringe: 10 mL, a metallic needle with a 0.4 mm inner diameter), separately. The syringe was placed vertically on the pump, and the positive tip of the high-voltage supply was connected to the metal needle tip while the negative lead of the power supply was grounded. The rotation speed of the drum collector surrounded with aluminum sheet was 2400 rpm. When the solution was injected via a pump delivery rate of 4.0 mL/h into the rotating drum collector, the interval between the needle tip and the rotating drum collector was 15 cm. During solution injection, when the high-voltage supply value reached 15kV, the electrostatic force acted in the opposite direction, and the polymer solution at the Taylor cone was adhered as fiber form on the collector by overcoming the surface tension.

2.3. Fabrication of the PEN

In order to design a PEN device, an electrospun film cut to a size of 4 x 4 cm^2 was positioned between two conductive aluminum electrodes as shown in Figure 2. a.

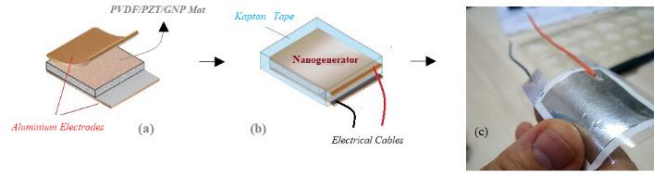


Figure 2. The fabrication stages of the piezoelectric nanogenerator.

Two copper cables were then attached to the surface of the aluminum electrodes on both sides with silver glue to ensure electrical conductivity. Afterward, as seen in Figure 2. b, the fabricated sample was fully protected by Kapton tape for a uniform packaging structure and finally obtained the flexible PEN device (Figure 2. c).

3. Results and Discussion

3.1. Structural Characterizations

The crystalline phase (X-Ray Diffraction, PANalytical Empyrean), Fourier-transform infrared spectroscopy (Bruker FTIR-ATR system), and morphology analyses (SEM: scanning electron microscopy, SUPRA 40VP) of the pure PVDF and the PVDF/PZT/GNP based fiber mats were shown Figures 3, 4, and 5 respectively as explained in detail in our previous study [25].

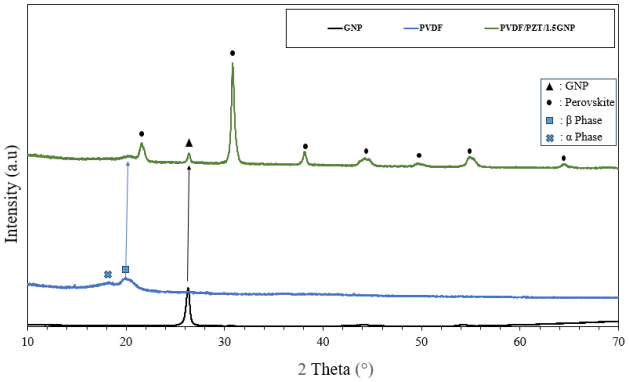


Figure 3. XRD Patterns of the neat PVDF and the PVDF/PZT/GNP with GNP loading concentration of 1.5wt.% [25].

According to Figure 3, the diffraction peak at 20.37° associated with the β phase, the characteristic diffraction peaks related to both the PZT (21.4°, 30.7°, 38°, 44.1°, 49.4°, 54.8°, and 64.4°) and the GNP (26.27°) exhibited the implementation of the electrospinning method in the formation of all samples remarkably reinforces the formation of the piezoelectric nanogenerator [25].

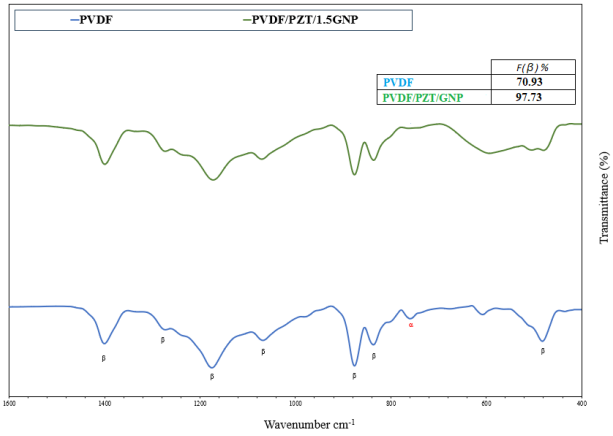


Figure 4. FTIR spectra of the pristine PVDF and the PVDF/PZT/1.5GNP [25].

When the FTIR spectra of all samples in Figure 4 are examined, it is deduced that the PEN based on the PVDF/PZT/1.5GNP reached the highest fraction of the β phase (97.73%) compared to the pristine PVDF (70.93%) [25]. Furthermore, the peaks (482°, 837°, 876°, 1068°, 1175°, 1272°, and 1400°) corresponding with the β phase shown in Figure 4 demonstrated an increase in the β phase formation of the composite structure [25]. As seen in SEM micrographs in Figure 5, it is observed that the PZT particles collected on fibers associated with the PVDF/PZT/GNP composite (Figure 5. b) due to the revolting force of the electric charges in contrast to the neat PVDF (Figure 5. a) [25].

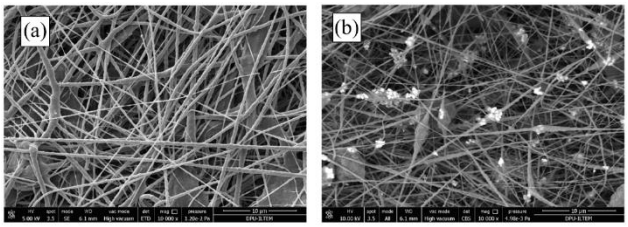


Figure 5. SEM images for the neat PVDF (a), the PVDF/PZT/1.5GNP electrospun mats (b) [25].

3.2. Electrical Performances of the PEN

In order to determine the electrical performances (voltage, current, and power) of the PEN device, Thevenin's, Norton's, and the maximum power transfer theorems were utilized through the two-terminal network system shown in Figure 6.

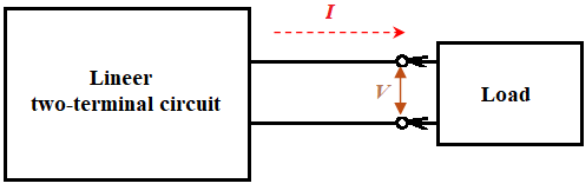


Figure 6. A linear two-terminal circuit.

Figure 7 indicates Thevenin's, Norton's, and the maximum power transfer equivalent circuits associated with the PEN.

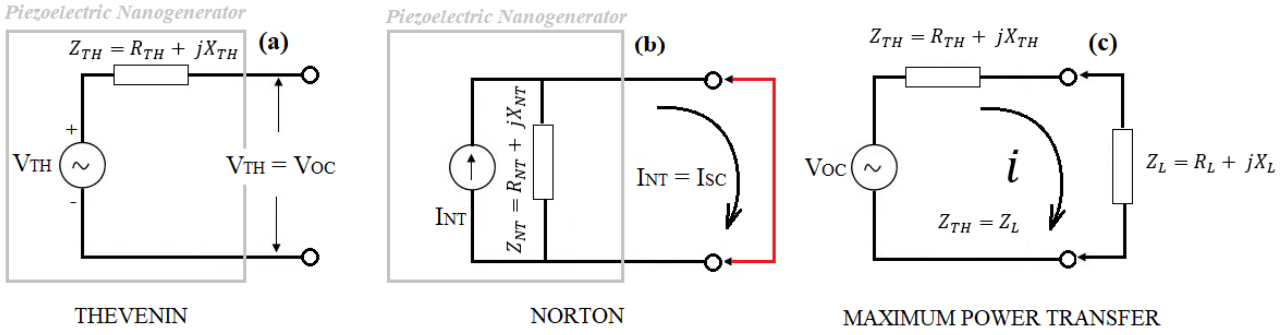


Figure 7. The Thevenin's (a), the Norton's (b), and the maximum power transfer (c) equivalent circuits of a PEN.

For AC circuit analysis, Thevenin's equivalent circuit consists of Thevenin's voltage-V_{TH} connected in series with Thevenin's impedance-Z_{TH} (Figure 7. a) while Norton's equivalent circuit contains Norton's current source-I_{NT} connected in parallel with Norton's impedance-Z_{NT} (Figure 7. b) [26]- [27].

As seen in Figure 7. a, if the terminals of the PEN device are open-circuited by removing the load impedance (Z_L), the open-circuit voltage between the terminals-V_{OC} will be equal to the voltage source V_{TH}, since no current is flowing in the circuit (V_{TH}=V_{OC}). When all independent sources are turned off, the input impedance-Z_i between the PEN tips will be equal to the Thevenin's impedance-Z_{TH} ($Z_i = Z_{TH}$).

According to Figure 7. b, when the PEN's tips are shorted, Norton's impedance-Z_{NT} will be shorted ($Z_{NT} = 0$) as well, hence Norton's current-I_{NT} equals a short circuit current-I_{SC} (I_{NT}=I_{SC}). If the independent energy sources are turned off, the input impedance-Z_i between the PEN terminals will be equal to Norton's impedance-Z_{NT} ($Z_i = Z_{NT}$).

As a result of all these explanations, considering the relationship between Thevenin's and Norton's theorems, we can write the following equation 1:

$$Z_{TH} = Z_{NT} = \frac{V_{OC}}{I_{SC}} \quad (1)$$

According to Figure 7. c, the electrical power on the load impedance (Z_L) can be expressed by the following equation 2:

$$P = i^2.Z_L = \left(\frac{V_{TH}}{Z_{TH}+Z_L}\right)^2.(Z_L) =$$

$$\left(\frac{V_{TH}}{R_{TH}+jX_{TH}+R_L+jX_L}\right)^2.(R_L + jX_L) \quad (2)$$

Due to the fact that the load resistance is a variable parameter, in order to find the maximum power, we differentiate the electrical power according to the load resistance, and this derivative must equal zero. After simplification, we get the below expressions.

$$X_L + X_{TH} = 0$$

$$X_L = -X_{TH}$$

It can be obtained from equation 3 by using the above relations [28].

$$\frac{dP}{dR_L} = V_{TH}^2 \left\{ \frac{(R_{TH}+R_L)^2 \frac{d}{dR_L}(R_L) - R_L \frac{d}{dR_L}(R_{TH}+R_L)^2}{(R_{TH}+R_L)^4} \right\} \quad (3)$$

$$\frac{dP}{dR_L} = V_{TH}^2 \left\{ \frac{(R_{TH}+R_L)^2 - 2R_L(R_{TH}+R_L)}{(R_{TH}+R_L)^4} \right\} = 0 \quad (3.1)$$

$$(R_{TH}+R_L)^2 - 2R_L(R_{TH}+R_L) = 0 \quad (3.2)$$

$$R_{TH} = R_L \quad (3.3)$$

From Eq.3.3, when Thevenin's resistance-R_{TH} is equal to the load resistance-R_L, it can be concluded that the maximum power will be transferred from the source to the load. Hence, we can achieve the equation. 4.1 via the equation. 4 [28].

$$P_{Max} = \left(\frac{V_{TH}}{R_{TH}+R_{TH}}\right)^2 . R_{TH} \quad (4)$$

$$P_{Max} = \frac{(V_{TH})^2}{4R_{TH}} = \frac{(V_{TH})^2}{4R_L} \quad (4.1)$$

We can also rewrite equation 1 as equation 5.

$$R_{TH} = \frac{V_{OC}}{I_{SC}} \quad (5)$$

Figure 8 depicts the experimental setup of the vibration energy harvesting system used to determine the piezoelectric performance of the PEN, which varies in proportion to the mechanical sensing performance, as well as several photographs of the manufactured PEN (Figure 8. a).

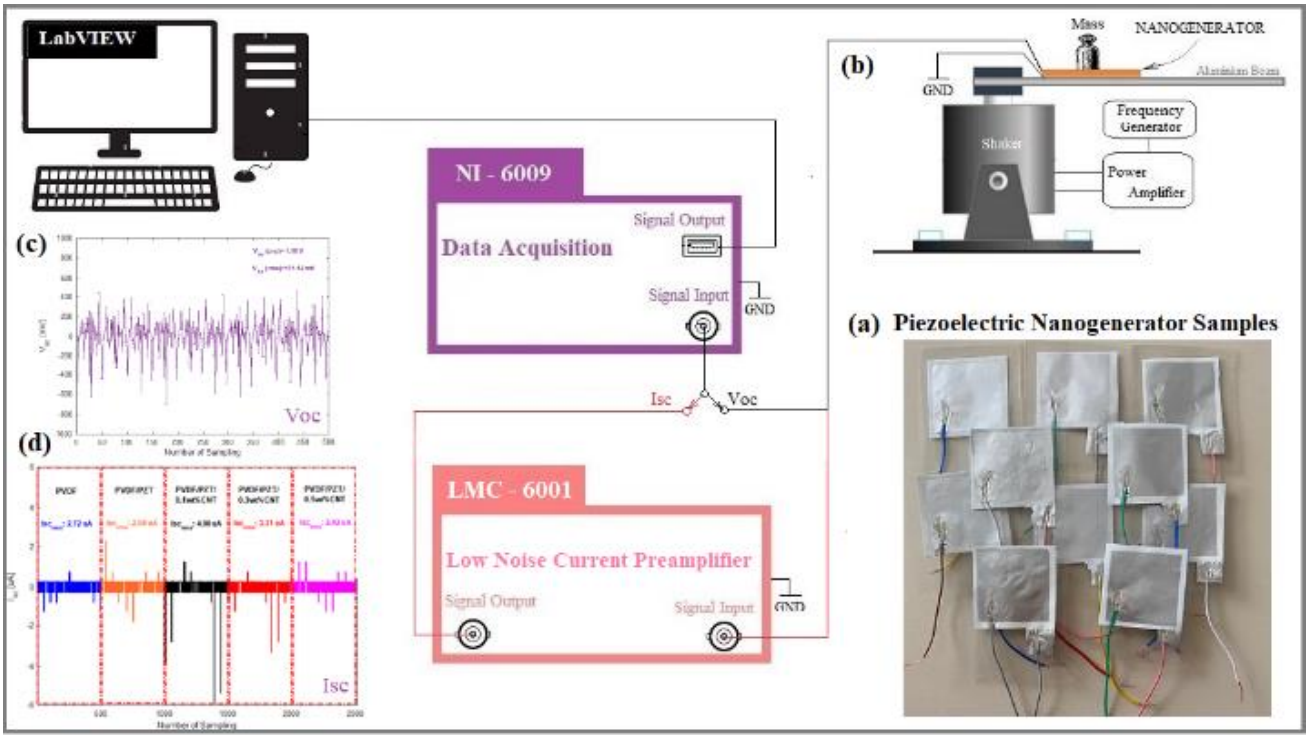


Figure 8. Vibrational energy harvesting system, and a photograph associated with fabricated PENs.

The vibrational energy harvesting system (Figure 8. b) is built of a magnetic shaker that vibrates an aluminum cantilever through a frequency signal generator at vibrational frequencies (5, 10, 15, 20, and 25 Hz). A power amplifier in the system provides a suitable amplitude for the efficient vibration of the magnetic shaker, whereas the NI-6009 and LMC-6001 data loggers are employed to measure the open-circuit voltage and the short-circuit current, respectively. Figures 8. c and d display the symbolic open circuit voltage-V_{OC} changes, and the symbolic short circuit current-I_{SC} changes.

The open circuit voltages (V_{OC}) all PEN samples were defined as V_{RMS} (RMS: the root-mean-square) by $V_{RMS} = \sqrt{\frac{1}{T}.\int_0^T Voc(t)^2 dt}$, while their

short-circuit currents (I_{SC}) were determined as I_{RMS} through $I_{RMS} = \sqrt{\frac{1}{T} \cdot \int_0^T Isc(t)^2 dt}$, where V_{RMS} = the effective voltage, the V_{OC} = the open-circuit voltage (V_{PP}: peak-to-peak), I_{RMS} = the effective short-circuit current, the I_{SC} = the short-circuit current (Isc_{PP}: peak-to-peak) and T = the periodic time.

The resonance (resonant) frequency is the characteristic frequency of a body or system that reaches the maximum degree of oscillation [29]. In the PEN devices, the resonant frequency is described by the vibration frequency at which the transfer function reaches its maximum voltage values. In this study, PEN devices were vibrated at specific frequencies (5, 10, 15, 20, and 25 Hz), and the maximum output voltage of the PEN was investigated at which frequency. Determining the resonant frequency of the PEN device is important in discovering the most efficient application range.

Figures 9 and 10 exhibit the V_{OC} and V_{RMS} changes of the PEN samples according to the vibrational frequencies (5, 10, 15, 20, and 25 Hz).

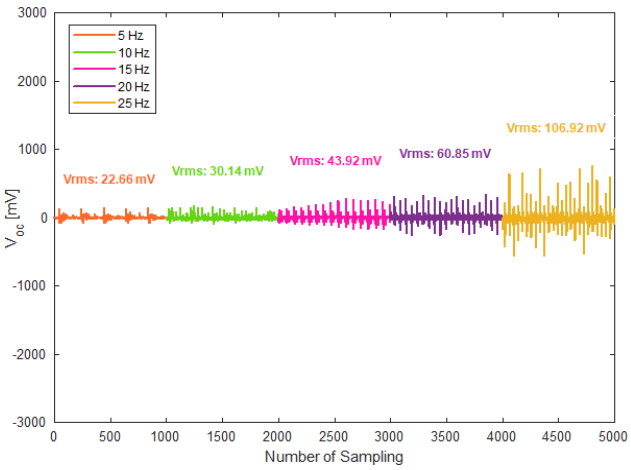


Figure 9. According to vibrational frequencies, the V_{OC} and V_{RMS} changes of the PEN based on the pure PVDF.

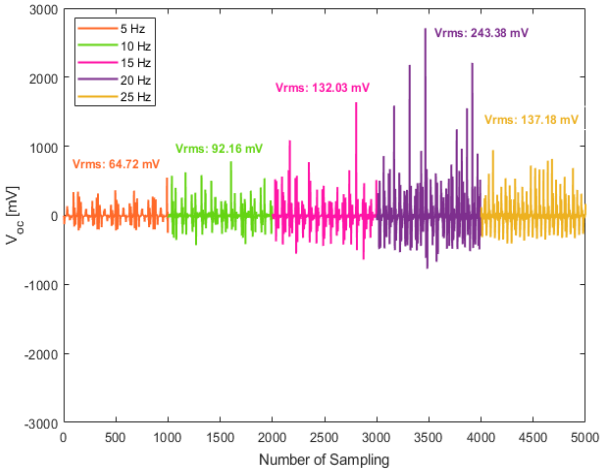


Figure 10. According to vibrational frequencies, the V_{OC} and V_{RMS} changes of the PEN based on the PVDF/PZT/GNP.

Figures 11 and 12 reveal the I_{SC} and the I_{RMS} alternates of all samples in compliance with the vibrational frequencies (5, 10, 15, 20, and 25 Hz).

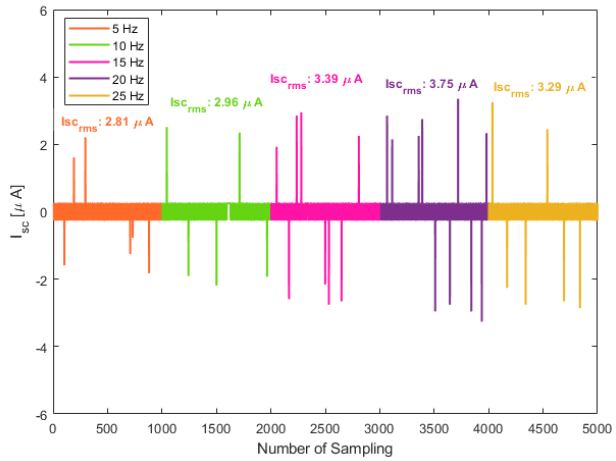


Figure 11. The I_{SC} and the I_{RMS} values depend on the vibrational frequency of the PEN based on the pure PVDF.

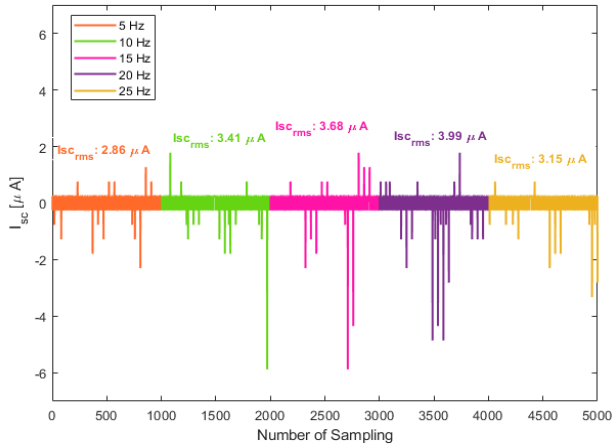


Figure 12. The I_{SC} and the I_{RMS} values depend on the vibrational frequency of the PEN based on the PVDF/PZT/GNP.

The maximum V_{RMS} value of 243.38 mV and I_{SC} value of 3.99 µA at 20 Hz for the PVDF/PZT/GNP electrospun mat-based PEN were measured as compared to the PEN based on the PVDF mat (V_{RMS}: 106.92 mV and I_{SC}: 3.29 µA at 25 Hz).

Figures 13 and 14 show the V_{RMS}, the I_{RMS}, and the P_{MAX} variations as a function of frequency for all samples.

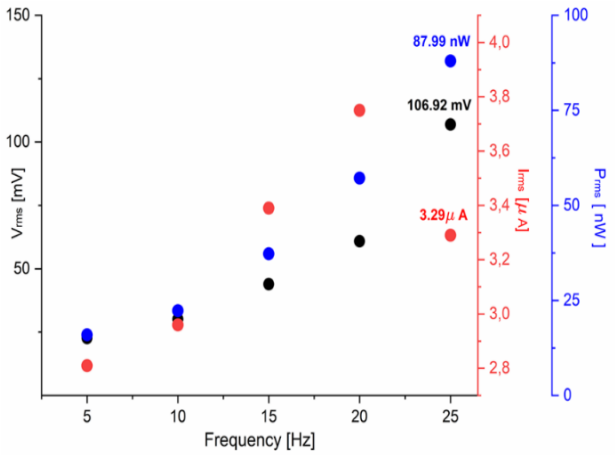


Figure 13. The V_{RMS}, the I_{RMS}, and the P_{RMS} changes of the PEN based on the pure PVDF.

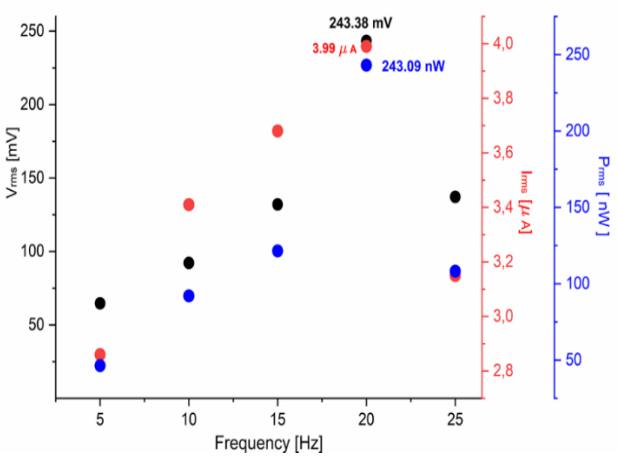


Figure 14. The V_{RMS}, the I_{RMS}, and the P_{RMS} changes of the PEN based on the PVDF/PZT/GNP.

According to Figure 15, in order to attitudes a general evaluation of samples, we can say that the fabricated PEN based on the PVDF/PZT/GNP reached an open circuit voltage of 0.24 V, and the highest electrical power of 0.24 µW by drawing a current of 1.99 µA at the vibrational frequency values of 20 Hz under a resistive load of 60.91 KΩ whereas the pure PVDF based PEN had 0.1 V, 1.64 µA, and 0.09 µW at the vibrational frequency values of 25 Hz under the resistive load of 32.48 KΩ, respectively.

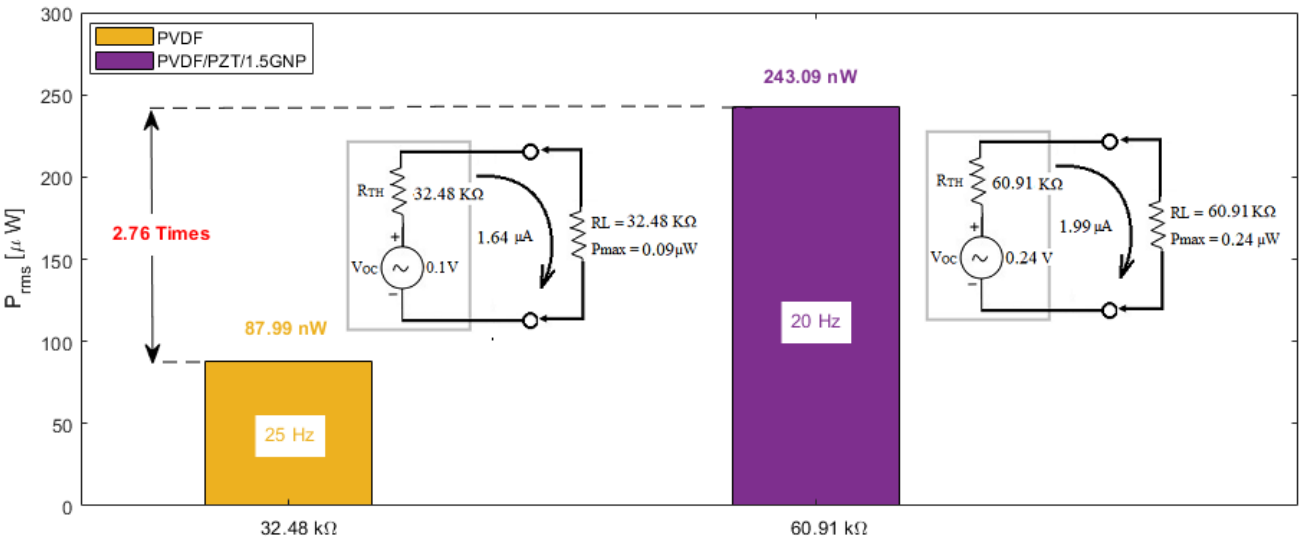


Figure 15. Comparing the $P_{MAX\text{-}RMS}$ alternations between the pure PVDF and the PVDF/PZT/1.5GNP.

In order to assess the stability and durability of the PVDF/PZT/1.5GNP-based piezoelectric energy harvester, a mechanical vibration test was performed under a vertical application force of 1 N at a frequency of 20 Hz.

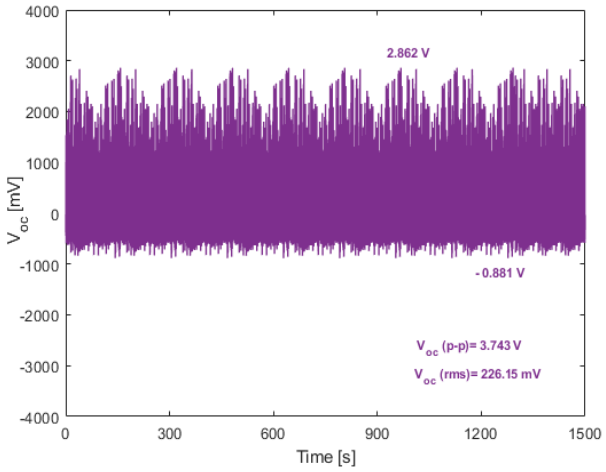


Figure 16. Stability test of the piezoelectric energy harvester for the 1500s.

It can be seen from Figure 16 that the output voltage alternations persisted almost stable for the 1500s, indicating the highest quality mechanical stability and durability.

4. Conclusion and Suggestions

The pure PVDF and the PVDF/PZT/GNP-based PENs were successfully fabricated through electrospinning. According to Thevenin's, Norton's, and the maximum power theorems, the PEN device based on the PVDF/PZT/1.5GNP which has the maximum open circuit voltage of 0.24 V (V_{RMS}) compared to the neat PVDF-based PEN, reached the highest power capacity of 243 nW (P_{RMS}) by drawing a current (I_{RMS}) of 1.99 μA at a vibrational frequency of 20 Hz (impact force of 1N) under a resistive load of 60.91 KΩ. In other words, the PEN based on the PVDF/PZT/1.5GNP had a better electrical power efficiency rate of 175% than that of the pure PVDF. Considering a microwatt-based energy harvesting system, the direct use of the PEN device is not enough for the energy harvesting system. Therefore, to boost the power transfer and minimize signal reflection (for better Signal/Noise ratio), it is mandatory to use a power management system that includes impedance matching, low & band-pass filters, a pre-amplifier, etc.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] R. Fazio, D. Cafagna, G. Marcuccio and P. Visconti, "Limitations and characterization of energy storage devices for harvesting applications", *Energies*, vol.13-4, 783, February 2020.

[2] A.Z.A. Shaqsi, K. Sopian and A. Al-Hinai, "Review of energy storage services, applications, limitations, and benefits", *Energy Rep*. vol. 6, pp. 288–306, 2020.

[3] A. Olabi, M.A. Abdelkareem, T. Wilberforce and E.T. Sayed, "Application of graphene in energy storage device–A review", *Renewable Sustainable Energy Reviews*, vol.135, 2021, 110026.

[4] R.A. Surmenev, R.V. Chernozem, I.O. Pariy and M.A. Surmeneva, "A review on piezo and pyroelectric responses of flexible nano-and micropatterned polymer surfaces for biomedical sensing and energy harvesting applications", *Nano Energy*, vol. 79, 2021, 105442.

[5] S.A. Graham, S.C. Chandrarathna, H. Panama, P. Manchi, J.-W. Lee and J.S. Yu, "Harsh environment–tolerant and robust triboelectric nanogenerators for mechanical-energy harvesting, sensing, and energy storage in a smart home", *Nano Energy*, vol.80, 2021, 105547.

[6] U. Yaqoob, S. M. I. Uddin and G.-S. Chung, "A Novel Tri-Layer Flexible Piezoelectric Nanogenerator Based on Surface- Modified Graphene and PVDF-$BaTiO_3$ Nanocomposites", *Applied Surface Science*, vol. 405, pp. 420-426, 2017.

[7] Y. Wu, et al., "A multi-mode triboelectric nanogenerator for energy harvesting and biomedical monitoring", *Nano Energy*, vol. 92, 2022, 106715.

[8] N. Piovesan, A.F. Gambin, M. Miozzo, M. Rossi and P. Dini, "Energy sustainable paradigms and methods for future mobile networks: a survey", *Computer Communications,* vol. 119, pp. 101–117, 2018.

[9] H. Li, K. Ota and M. Dong, "Energy cooperation in battery-free wireless communications with radio frequency energy harvesting", *ACM Transactions Embedded Computing Systems*, vol. 17, pp. 1-17, 2018.

[10] A. Ali, S. Iqbal and X. Chen, "Recent advances in piezoelectric wearable energy harvesting based on human motion: Materials, design, and applications", *Energy Strategy Reviews*, vol. 53, pp. 1-16, 2024.

[11] S. Bairagi, S. Islam, M. Shahadat, D.M. Mulvihill and W. Ali, "Mechanical energy harvesting and self-powered electronic applications of textile-based piezoelectric nanogenerators: A systematic review", *Nano Energy*, vol. 111, pp. 1-29, 2023.

[12] Z. Li, J. Roscow, H. Khanbareh, G. Haswell and C. Bowen, "Energy Harvesting from Water Flow by using Piezoelectric materials ", *Advanced Energy & Sustainability Research*, vol. 5, pp. 1-33, 2024.

[13] A. Salimi and A.A. Yousefi, "Analysis Method: FTIR studies of beta-phase crystal formation stretched PVDF films ", *Polymer Testing*, vol. 22, pp. 699-704, 2003.

[14] H. Pan, B. Na, R. Lv, C. Li, J. Zhu and Z. Yu, "Polar phase formation in poly (vinylidene fluoride) induced by melt annealing", *Journal of Polymer Science, Part B: Polymer Physics*, vol.50, pp. 1433–1437, 2012.

[15] S.F. Mendes, C.M. Costa, C. Caparros, V. Sencadas and S. Lanceros-Mendez, "Effect of filler size and concentration on the structure and properties of poly (vinylidene fluoride)/$BaTiO_3$ nanocomposites", *Journal of Materials Science*, vol. 47, pp. 1378–1388, 2012.

[16] Y. Wang, D. Lei, L. Wu, R. Ma, H. Ning, N. Hu and A. Lee, "Effects of stretching on phase transformation of PVDF and its copolymers: A review", *Open Physics*, vol. 21, 2023, 2022-0255.

[17] R.S. Sabry and A.D. Hussein, "PVDF: ZnO/$BaTiO_3$ as high out-put piezoelectric nanogenerator", *Polymer Testing*, vol. 79, 2019, 106001.

[18] H.G. Yeo, X. Ma, C. Rahn and S. Trolier-McKinstry, "Efficient piezoelectric energy harvesters utilizing (001) textured bimorph PZT films on flexible metal foils", *Advanced Functional Materials* vol.26, pp. 5940–5946, 2016.

[19] X. Chen, S. Xu, N. Yao and Y. Shi, "1.6 V nanogenerator for mechanical energy harvesting using PZT nanofibers", *Nano Letter*, vol. 10, pp. 2133–2137, 2010.

[20] M. Sobocinski, M. Leinonen, J. Juuti, N. Mantyniemi and H. Jantunen, "A co-fired LTCC–PZT monomorph bridge type acceleration sensor", *Sensors and Actuator A: Physical*, vol. 216, pp. 370–375, 2014.

[21] X. Lin, F. Yu, X. Zhang, W. Li, Y. Zhao, X. Fei, Q. Li, C. Yang, and S. Huang, "Wearable Piezoelectric Films Based on MWCNT-$BaTiO_3$/PVDF Composites for Energy Harvesting, Sensing, and Localization", *ACS Applied Nano Materials*, vol. 6, pp. 11955−11965, 2023.

[22] R. Sridar, V. Amith, S. Aditva, A. Gangadhar and K.A. Vishnumurthy, "Electrospun PVDF/Cloisite-30B and PVDF/$BaTiO_3$/graphene nanofiber mats for development of nanogenerator", Journal of the Indian Chemical Society, vol. 99, 2022, 100501.

[23] S. K. Ghosh, A. Biswas, S. Sen, C. Das, K. Henkel, D. Schmeisser and D. Mandal, "Yb3+ assisted self-polarized PVDF based ferroelectric nanogenerator: A facile strategy of highly efficient mechanical energy harvester fabrication", Nano Energy, vol. 30, pp. 621-629, 2016.

[24] L. H. Meng, C. Yang, J. Meng, Y. Wang, Y. Ge, Z. Shao, G. Zhang, A. L. Rogach and H. Zhong, "In-situ fabricated anisotropic halide perovskite nanocrystals in polyvinyl alcohol nanofibers: Shape tuning and polarized emission", *Nano Research*, vol. 12, pp. 1411-1416, 2019.

[25] L. Paralı, F. Tatardar, M. Koç, A. Sarı and R. Moradi, "The piezoelectric response of electrospun PVDF/PZT incorporated with pristine graphene nanoplatelets for mechanical energy harvesting", *Journal of Materials Science: Materials in Electronics*, vol. 35, 41, 2024.

[26] John Hiley, Keith Brown and Ian McKenzie Smith, "Electrical and Electronic Technology", 10th edition, Pearson Education Limited, 2008.

[27] Alexander, Charles K., and Sadiku, Matthew N. O., "Fundamentals of Electric Circuits", 5th Ed, McGraw Hill, Indian Edition, 2013.

[28] Md. Abdus Salam and Quazi Mehbubar Rahman, "Fundamentals of Electrical Circuit Analysis", ISBN 978-981-10-8624-3 (eBook), Springer Nature Singapore Pte Ltd. 2018.

[29] David Halliday, Robert Resnick and Jearl Walker, "Fundamentals of Physics", 7th Ed, ISBN 978-0-471-71716-4, John Wiley & Sons Limited, 2005.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 905-915 YEAR: 2024
DOI: 10.17798/bitlisfen.1459003

Analysis of Current-Voltage Properties of Schottky Diode with TiO_2 Interlayer Prepared by RF Magnetron Sputtering

Barış POLAT[1], Elanur DİKİCİOĞLU[2*]

[1]*Ankara Medipol University, Faculty of Engineering and Natural Science, Department of Industrial Engineering, Ankara, Türkiye*
[2]*Yüksek İhtisas University, Vocational School of Health Services, Ankara, Türkiye*

(ORCID: 0000-0003-3314-2091) (ORCID: 0000-0002-8984-1054)

Keywords: Schottky Diode, Thermionic Emission Method, Current-Voltage, RF Magnetron Sputtering.

Abstract

This study focuses on investigating the electrical behaviour of Metal-Insulator-Semiconductor (MIS) type Schottky barrier diodes based on titanium oxide (TiO_2). An MIS-type Al/TiO_2/p-Si Schottky diode structure was fabricated by depositing a TiO_2 metal oxide thin film as an interlayer on p-type silicon using the technique of Radio Frequency Magnetron Sputtering at room temperature. The electrical performance of this fabricated structure was evaluated through the measurements of current-voltage (I-V) conducted in a dark environment at ±5 V and room temperature. These measurements enabled the determination of key Schottky diode parameters, including barrier height (Φ_b), saturation current (I_o), and ideality factor (n), using both the Thermionic Emission (TE) method and the Cheung method. Utilizing the TE method, approximate values for Φ_b, n, and Io parameters were calculated as 0.59 eV, 4.07, and 2.78E-06 A, respectively. Meanwhile, employing Cheung's method yielded approximate values of Φ_b and n parameters as 0.39 eV (H(I) vs I) and 4.39 (dV/dln(I) vs I), respectively. The analysis indicates that the developed Schottky diode functions as a rectifier diode, demonstrating typical diode characteristics. Furthermore, a comparison of numerous devices reported in the literature was conducted based on TiO_2 preparation methods against the parameters of the TiO_2/p-Si host device.

1. Introduction

Conventional diodes use p-type and n-type semiconductors to form a p-n structure. In Schottky diodes, a metal and a metal-semiconductor are combined to form a Schottky barrier at the junction [1], as shown in Figure 1. Looking at the characteristics of Schottky diodes, it can be seen that they have many good properties, such as fast switching and low voltage drop. Fast switching is described as the transition between the on-state and the off-state, where the conduction is stopped when the voltage is applied forward. In fact, this speed is related to the fact that the memory load of Schottky diodes is much less than that of other diodes. When compared to conventional diodes, Schottky diodes are more resistant to the low voltage drops known as forward voltage drops, thus reducing power losses and increasing energy efficiency. Schottky diodes have both advantages and disadvantages; examples of disadvantages include the fact that they produce a lot of heat in an operating system and are resistant to low voltage when it's fed backwards. The fact that they produce a lot of heat in the operating system is a situation that should be taken into consideration in systems that need to operate at high power. Examples of its advantages include being preferred in situations where efficiency, power, high switching speed, and high-frequency operation are important [2]. Given these characteristics, they are used in power supplies,

*Corresponding author: elanurdikicioglu@yiu.edu.tr
Received: 26.03.2024, Accepted: 25.10.2024

power rectifier circuits [3], radio frequency sensitive detectors [4], and circuits requiring high switching speeds [5].

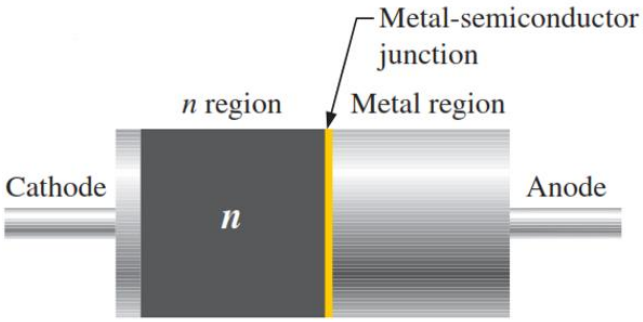


Figure 1. Schematic of a typical Schottky diode.

Depending on the thickness of the interlayer between the n-region and the metal region, Schottky diode structures are referred to as Metal/Semiconductor (MS), Metal/Insulator/Semiconductor (MIS) or Metal/Oxide/Semiconductor (MOS) structures. They consist of metals corresponding to the chosen p- or n-semiconductors. Literature suggests that metals such as gold (Au), silver (Ag), and aluminium (Al) are used for the metallic area because of their high purity. The preferred choice of p-type or n-type semiconductor material is silicon (Si). This is because it is more economical and stable [6].

When semiconductor components are manufactured, charge transfer between structural metals and semiconductors plays a major role. The control of this transfer is the key to the performance of the solid-state device. Therefore, insulating layers such as silicon dioxide (SiO_2), tin dioxide (SnO_2), silicon nitride (Si_3N_4) and TiO_2 are commonly used as interlayers. For this reason, a TiO_2 interlayer was used in the present study. The interlayers in question do not have to be exclusively inorganic. Organic interfacial layers such as polyindole, polyaniline, and polyvinyl alcohol (PVA) are known to form. The interfacial layers between the metal and the semiconductor are chosen according to a certain logic. They are chosen to provide surface passivation, minimum leakage current, a controllable current transfer mechanism, and high dielectric stability close to rectifier properties [7]-[9].

In terms of their electrical characteristics, MIS and MOS structures are very similar. These structures can be electrically distinguished by whether their interface structures are insulating or polymeric and have variable metal-semiconductor interface properties, and the series resistance (R_s) and barrier height of the resultant Schottky diode can be uniform. While the MS structure means that there is no insulator or polymer interface layer, the interface layers in MIS and MOS structures can be insulator and polymer. This situation causes changes in the interface states. This change serves to regulate the interface transitions by isolating the metal-semiconductor structure from each other, as well as altering the electrical properties of the metal-semiconductor combination that forms the Schottky diode structure. Consequently, current transfer in MIS and MOS structures differs from that in MS structures [10].

When the thickness of the insulating interface layer between the metal and the semiconductor is less than 100Å, Schottky diodes are called MIS, and if it is more than 100Å, they are called MOS. For a fabricated MIS structure to be ideal, two situations must occur. The first is the absence of any intermediate charge and interface state in the insulator between the metal and the semiconductor. The second is that this situation does not occur at the insulator-semiconductor interface. In reality, MIS structures differ from the ideal [11], [12] because of the interactions between the insulating layer within the MIS structure and the charges at the interface between the insulator and the semiconductor. Another deviation from the ideal is that the dielectric properties of the isolation layer at the interface are similar to those of a parallel plate capacitor [12]-[14].

The insulating semiconductor metal, titanium dioxide (TiO_2), is used as the interface in this paper and has high transmittance and permittivity over a wide range of wavelengths in the visible region of the electromagnetic spectrum. TiO_2 has three different natural polymorphs: anatase (tetragon), rutile (tetragon), and brookite (orthorhombic). The most thermodynamically stable phase is rutile TiO_2. However, metastable anatase and brookite TiO_2 phases can be converted to rutile phases after calcination at a certain temperature. Rutile and anatase TiO_2 are generally tetragonal in structure [15], whereas brookite TiO_2 is orthorhombic. The phases of TiO_2 are shown in Figure 2. TiO_2's dielectric constant differs between these phase conditions, and its transparency and electrical conductivity vary as its bandgap alters. For example, widening the bandgap improves their transparency but reduces their electrical conductance. If the electrical conductivity is to be increased in parallel with the increase in transparency, metals are added to the TiO_2 structure to increase the defect density of the structure, and it is observed that the overall conductivity increases. The ability to use TiO_2 as a coating material is another important aspect. Its high refractive index and high melting temperature are related to this

coating ability. As a result, it has been widely used as a coating material in optical circuits and optical applications. For the semiconductor metal TiO_2 grown as an insulating interfacial layer in the present study, there are many growth processes. They are spraying, evaporation, chemical deposition, and sol-gel methods.

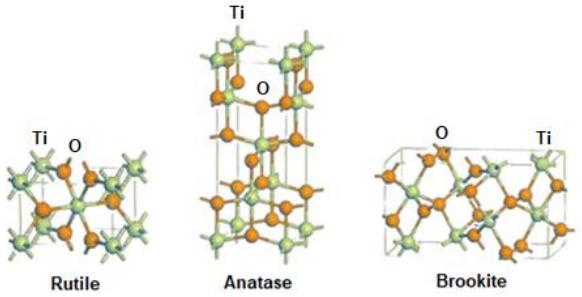


Figure 2. Crystal configurations of titanium dioxide in different phases [15].

In the present study, RF magnetron sputtering was used to grow rutile TiO_2 as an interfacial insulating layer. This method relies on creating a plasma on the surface of the object, and the atoms or molecules in the plasma form a thin layer on the surface for deposition [16]. It has been widely used in the literature for the fabrication of Schottky diode and photodiode structures. Using the RF magnetron sputtering method, Ali fabricated Ag/ZnO/Al Schottky photodiodes with rectifying properties [17]. Using RF magnetron sputtering, Ferhati et al. fabricated metal/TCO/p-Si structures. They mentioned that the Al /ZnO /p-Si Schottky structure can be a reliable and cost-effective solar cell [18]. Using the same method, Raj et al. have developed a p-type CdTe/V_2O_5/Ag Schottky diode structure. Purpose the V_2O_5 interface layer was a contributor to the electrical behaviour of the fabricated Schottky diode. It was shown to be useful in the fabrication of multi-structure solar cells such as TCO/CdS/CdTe/V_2O_5/Ag [19], with positive results from this contribution.

In order to understand the current response of the structure to voltage, many studies have been carried out to investigate the interface states at the insulator/semiconductor junctions that form Schottky structures [20]. As can be seen from these studies, many theories have been developed for the determination of the I-V of the Schottky structure [21]. Examples of these theories are the TE method and the Cheung method. The most common of these methods is the traditional I-V method [22], and in this method, the diode parameters are found using the slope of the linear region of the ln I versus V curve. Another method, the Cheung method, is applied to the data in the downward curved region of the forward bias Ln. I-V curve or to the linear region, and the diode parameters can be found with the Cheung functions [23].

How close to ideal the I-V characteristics of a Schottky structure determines whether it can be described as an ideal diode. It is therefore necessary to examine the I-V characteristics of the structure to determine how close it is to the ideal. This is done by examining the junction states of the structure. By understanding and controlling the interfacial behaviour, it is possible to make diodes which come close to the ideal Schottky diodes. It extends the life of the construction and makes it possible to change the construction more quickly.

It is the release of electrons and holes from the surface due to the heating of any surface and the generation of a TE. In Schottky diode structures, the metal part of the Schottky structure is allowed to pass through the junction between the semiconductor and the bulk, or vice versa, if the charge carriers do not have enough energy to overcome the potential barrier due to the temperature rise of the surface caused by heat, and this situation is called TE. In the case of n-type semiconductors, electrons are emitted. In the case of p-type semiconductors, holes are used to create Schottky structures [24].

In this study, the RF magnetron sputtering technique was used to fabricate the Al/TiO_2/p-Si Schottky diode structure. From the I-V measurements in the dark and at 300 K of the fabricated diode, the electrical parameters of the structure, such as series resistance (Rs), barrier height (Φ_b), and ideality factor (n) were examined using TE and Cheung methods.

2. Material and Method

The present study used a p-Si crystal as a substrate with two different structures (one matte, one shiny). The surface orientation of the crystal is 100, its thickness is 380 μm, and its resistivity is in the range 1-10 Ω-cm. The fabrication of the Schottky structure started with the chemical cleaning of the p-type Si substrate. It was first submerged in acetone for 10 minutes at 50 °C, followed by a deionized water wash and a two-minute release into methanol. Once more, the wafer was immersed in a NH_4OH:H_2O_2:H_2O solution for 15 minutes at 70 °C after being cleaned with deionized water. To get rid of the solution on the wafer surface, it was dipped into deionized water. The wafer was washed in a 2% Hydrogen Fluoride (HF) solution for two minutes to eliminate any free oxygen from the surface. Lastly, the cleaning process was finished with deionized

water [25]. A high-purity Al contact with 124 nm thickness was deposited on the matte surface of the silicon after the cleaning process. The purpose of this contact is to provide an ohmic contact. On the whole of the bright side of the p-type Si crystal, a 17 nm thick TiO_2 insulating interface layer was deposited and grown. RF magnetron sputtering was used for deposition, and the deposition parameters are given in Table 1. The final technique used to create the Schottky diode structure was physical vapour deposition (PVD). This technique is used for the fabrication of the rectifier contacts. Approximately 128 nm of Al metal is deposited on TiO_2 as the rectifier contacts. Figure 3 shows the resulting Schottky diode structure (TiO_2/p-Si Schottky barrier diode). Clearly, the TiO_2/p-Si diode has a rectifying function and was constructed using mathematical approaches.

Table 1. Parameters used in RF magnetron sputtering process for TiO2 thin-film fabrication.

Process parameters	**Values**
Power	150 Watt
Coating pressure	4.2×10^{-3} mbar
Base pressure inside the chamber	3.3×10^{-7}mbar
Target-to-substrate distance	50 mm
Sputtering gas	Argon
O_2 flow rate	1.3 sccm
Argon flow rate	12 sccm
Ar/O_2 ratio	9/1
Deposition rate	0.1 Å/s

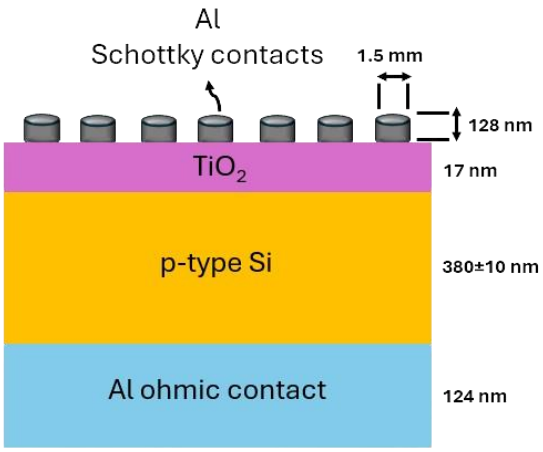


Figure 3. Schematic layout of Schottky structure.

3. Results and Discussion

The electrical parameters of the Al/TiO_2/p-Si Schottky structure, including barrier height (Φ_b), ideality factor (n), and series resistance (R_s), were examined in the present study by utilizing current-voltage (I-V) characteristics. All electrical characterization measurements were taken in a dark room and under room temperature, and the I-V graph obtained from the data is shown in Figure 4.

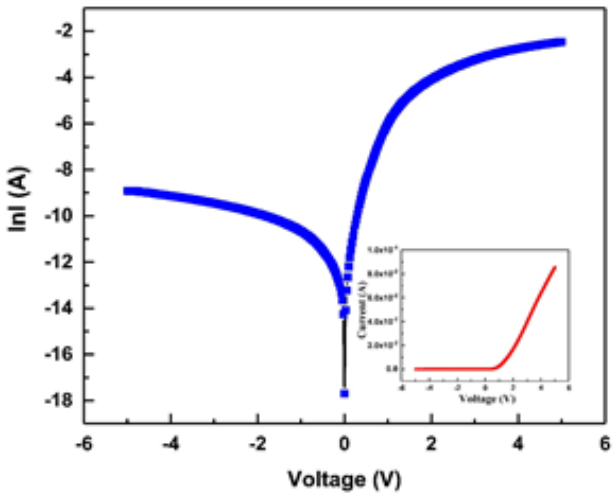


Figure 4. ln (I)-V characteristics of Al (124 nm) / TiO_2 (17 nm) / p-Si (380±10 nm) Schottky diode structure.

The presence of TiO_2 in the Schottky structure as an interlayer shows the behaviour of a high rectifying, low leakage diode, as shown in Figure 4. As higher currents are achieved, the current side of the ln(I)-V curve becomes progressively curved. This is the forward polarised region, and the bend is related to the mass resistance of the R_s or MIS structures of the contact wires used to measure the I-V curve. In the forward bias region, the current increases exponentially with the voltage, while in the reverse bias region, the current is seen to have a weak dependence on the voltage. At higher bias voltages, the ln I-V plot has a significant deviation from the linear case due to the effects of series resistance. The R_s effect and the existence of interfacial states cause the I-V curves to bend downward at sufficiently high supply voltages.

It is also possible to insert an artificial insulating layer between the metal and the semiconductor. The presence of this insulating layer transforms the structure into a metal-insulator-semiconductor shape and completely modifies the I-V characteristics of the layer. Therefore, a single current conduction mechanism may not be sufficient to study the electrical characterisation of the

Schottky diode structure at a desired temperature and voltage.

In the presented study, the TE method was used to make this characterization. If MIS structures have an insulating interface layer and V>3kT/q, there is a relationship between forward bias voltage and current as in equation (1) [1].

$$I = I_0 \, exp\left(\frac{qV}{nkT}\right)\left[1 - exp\left(-\frac{qV}{kT}\right)\right] \quad (1)$$

To perform the electrical characterization of an ideal Schottky diode structure using the TE method, it is expected that the applied voltage should not be too high. In this way, the current transmission model will be suitable for TE. Considering the case V>3kT/q, equation (1) can be arranged as equation (2);

$$I = I_0 \, exp\left(\frac{qV}{nkT}\right) \quad (2)$$

where n represents the ideality factor of the Schottky diode structure, q is the amount of charge of the electron, T represents temperature in K, kB is the Boltzmann constant, V is the voltage applied to the structure, and Io is reverse saturation current (equations (1) and (2)). I_o is obtained from the straight-line intersection of the ln(I)-V graph shown in Figure 4. The mathematical approach is expressed by equation (3);

$$I_0 = AA^*T^2 \, exp\left(-\frac{q\varphi_b}{kT}\right) \quad (3)$$

where A*, A, and Φ_b in equation (3) denote the Richardson constant of p-type Si taken as 32 A/cm^2K^2, the surface area of the Schottky diode, and the effective barrier height of the structure, respectively.

The graph of ln(I) versus V should be linear for the ideal Schottky diode structure. It is the high value of n that causes the curve to deviate from linearity. This will depend on how far the diode is from being in its ideal state. Ideally, the Schottky diode is expected to work in harmony with the TE technique in low-current transmissions. However, there can be various deviations in practical applications. This far-from-ideal behaviour is explained by the presence of the parameter n. This parameter is explained by the equation (2). In fact, to determine the compatibility of Schottky diodes with the TE technique, this parameter is often used. This can be readily reached using the slope of the linear part of the graph in Figure 4. Equation (2) is rearranged, and the mathematical approach for the parameter n is rearranged as equation (4);

$$n = \frac{q}{kT}\frac{dV}{d(ln\, I)} \quad (4)$$

The parameter n can be found by the slope of the ln(I)-V graph. The slope in question is that it can be determined by finding the nonlinear region in the graph and extrapolating that region. The barrier height Φ_b of the Schottky diode structure is expressed by equation (5);

$$\varphi_b = \frac{kT}{q} ln\left(\frac{AA^*T^2}{I_0}\right) \quad (5)$$

The values of Φ_b and n can be calculated from the intersections of the ln(I)-V curve, which shows a forward trend due to the room temperature experimental study. In an ideal diode, n equals one. However, n usually has a value larger than unity. The reasons why this parameter moves away from 1 are the presence of an insulating layer between the metal and the semiconductor, Rs, and the interface density of states (Nss) in the forbidden energy band [26, 27]. The voltage applied to diode structures during electrical characterisation affects the barrier height within the structure, and this effect is related to the naturally occurring insulating oxide interfacial layer (~5-20Å thick) even in ideal diode structures.

The effect of the TiO_2 layer on the series resistance (Rs) of the device was also observed using Cheung functions [23]. The current flowing through a component is not linear at high forward bias. It is the use of this non-linear region by the R_s and Cheung functions that causes the potential drop in this region. The value of R_s can be attained using the following equations for the Al/TiO_2/ p-Si heterojunction [23]:

$$\frac{dV}{d(\ln I)} = n\frac{kT}{q} + IR_s \quad (6)$$

$$H(I) = V - \frac{nkT}{q}\ln\left(\frac{I}{AA^*T^2}\right) \quad (7)$$

$$H(I) = n\phi_b + IR_s \quad (8)$$

where Rs in equations (6) and (8) denote the series resistance.

When the parameters of Cheng's functions are examined, the barrier height is influenced by the n. A higher n (greater than 1) often indicates

increased recombination in the depletion region, which can lead to a lower effective Φ_b. The Φ_b can change with temperature due to the intrinsic properties of the materials involved. Higher temperatures can increase the thermal energy available, potentially reducing Φ_b [28]. If there is significant Rs, it can distort the I-V characteristics, leading to an increased n. A lower Rs generally indicates better diode performance and can result in an n closer to 1 [29]. The n can also vary with temperature, as changes in thermal energy affect charge carrier dynamics and recombination rates. The Io is exponentially related to the Φ_b. A higher Φ_b typically results in a lower Io, as fewer carriers can overcome the barrier. Io increases with temperature, as higher thermal energy helps more charge carriers to overcome the barrier, which can also affect the Φ_b determination. The Rs can obscure the true behaviour of the diode. High Rs can cause a nonlinear I-V response, leading to inaccurate determinations of Φ_b and n. Ideally, a Schottky diode should have a minimal Rs. Increased Rs leads to higher voltage drops at higher currents, making the diode appear less ideal and affecting the extracted parameters [30]. By considering these interactions, Cheung's functions provide a more comprehensive understanding of the diode's performance characteristics.

The plots of dV/d(lnI) and H(I) vs. I will both be linear, as per the last two equations, and they will be utilized to find the experimental n, Φ_b, and R_s. On the other hand, two R_s values are determined, one from Equation (6) and the other from Equation (8). The Φ_b was attained from the linear part in the forward bias for I–V characteristics. Figure 5 displays the graphs of dV/d(lnI) vs. I and H(I) vs. I for the Al/TiO_2/p-Si diode at 300 K, respectively.

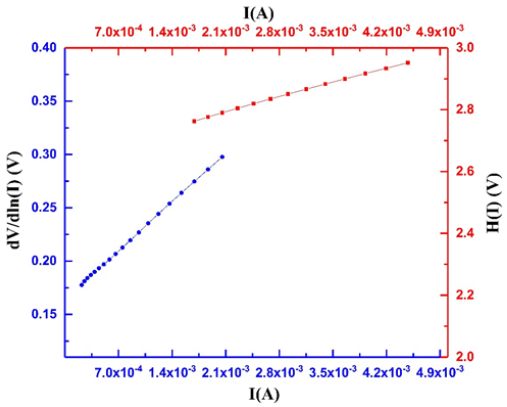


Figure 5. Cheung plots of Al (124 nm) / TiO_2 (17 nm) / p-Si (380±10 nm) Schottky diode structure.

Rs and n values are calculated from the slope and intersection of the straight line on the curve of dV/dln(I) vs. I. The values of Φ_b are acquired from the intersection of the curve in Eq. (8), and R_s is determined from the slope of the H(I) vs. I curve using the n value in Eq. (6). Table 2 displays the computed values of n, Φ_b, and R_s for the Schottky diode. Table 2 provides the diode characteristics for the Al/*RF magnetron method-grown* TiO_2/p-type Si.

The n, Φ_b and I_o values attained from the I-V measurements of the Al/TiO_2/p-Si Schottky diode structure created specifically for the presented study are given in Table 2.

Table 2. The crucial parameters attained from I-V measurements.

Method		**n**	**Φ_b (eV)**	**R_s (kΩ)**
Standard (TE)		4.07	0.59	-
Cheung	dV/d(lnI) vs I	4.39	-	0.366
	H(I) vs I	-	0.39	0.125

Table 2 illustrates the near-ideal diode characteristics of the TiO_2/p-Si heterojunction diode utilized in our research, featuring a Φ_b value of approximately 0.59 eV. To facilitate comparison with our experimental findings, we also present heterojunction parameters acquired through various techniques documented in the literature in Table 3. The values recorded in Table 2 are within the range of values previously reported for TiO_2/p-Si heterojunctions produced using various deposition techniques [31]-[38]. The main advantage of our study is that the technique used is suitable for the deposited film, based on the comparisons in Table 3. It was found that using the TE technique and Cheung functions, the n values for the diode were 4.07 and 4.39, respectively. Using the TE technique and Cheung functions, the Φ_b values were found to be 0.59 eV and 0.39 eV respectively. The non-ideal behaviour of the diode (n>1) confirmed the existence of interface states. The Φ_b values for the diode obtained by both methods are in agreement. The diode properties of various Schottky diodes with TiO_2 interlayers of different thicknesses, fabricated at 300 K using different techniques in the literature [31]-[38], are presented in Table 3 to comparison with our research.

Table 3. Comparing diode characteristics across different types of Schottky diodes incorporating a TiO_2 interlayer.

Diodes	Interlayer Coating Methods	n	Φ_b (eV)	References
Al (124 nm) / TiO_2 (17 nm) / p-Si (380±10 nm)	RF magnetron sputtering method	4.07 (TE) 4.39 (Cheung)	0.59 0.39	Present work
Al/TiO_2 (4 nm) /p-Si	ALD	12.71 (TE) 12.73 (Cheung)	0.64 0.68	[2]
Al/n-TiO_2/*p*-Si	Spin coating	5.4	0.58	[3]
Al/TiO_2 (760 nm) /p-Si	Drop-casting method	2.39 (first linear region) 8.29 (second linear region)	0.74 eV 0.65 eV	[33]
Au/TiO_2/p-Si	Drop-casting method	17.42	0.654	[34]
Al/TiO_2(55 nm)/p-Si	Pulsed laser deposition	3.73	0.69	[35]
Al/TiO_2/p-Si	Spin coating	8.97 (TE) 9.53 (Cheung)	0.74 0.55	[36]
Al/TiO_2/p-Si	Chemical	1.51	0.72	[37]
Au/TiO_2/p-Si	Electrochemical	2.07	0.84	[38]

Table 3 illustrates how the n and Φ_b values were impacted by the variations in the techniques used to construct the interface layer in the measurements. In this study, when basic diode parameters such as n and Φ_b found using the RF magnetron sputtering method were compared with the values obtained with other methods, lower n and Φ_b values were found than the n and Φ_b values obtained using ALD [31], Spin coating [22], [36] and Drop-casting [33], [34] methods. The basic diode parameters of the Al (124 nm) / TiO_2 (17 nm) / p-Si (380±10 nm) Schottky diode in our study were found to be higher than the values found by chemical [37] and electrochemical [38] methods.

In the presented study, electrical characterization of Schottky diode structures was carried out under forward bias voltage at room temperature. Characterization parameters n and Φ_b were calculated using the TE technique, which allows the determination of forward I-V properties. The Schottky diode structure was fed with flat power, and a structure-specific I-V graph was obtained. With the help of this graph, the n value was calculated as 4.07 and the Φ_b value as 0.59 eV. As illustrated in Figure 6, the diagram of the energy band of the TiO_2/p-Si heterostructure is fabricated under equilibrium at zero bias in the dark.

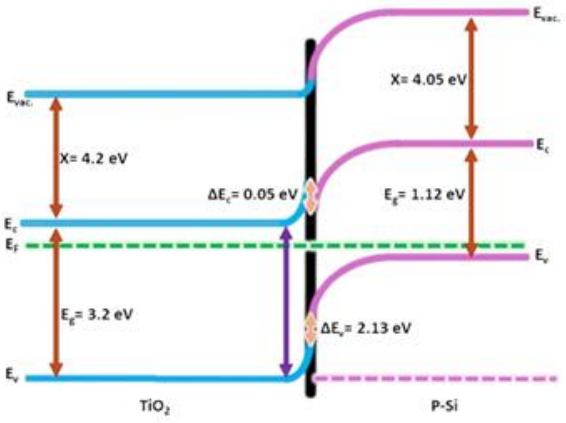


Figure 6. Schematic the diagram of the energy band of TiO_2/p-type Si Schottky structure under dark [32].

4. Conclusion and Suggestions

The electrical characterisation of the Al/TiO_2/p-Si Schottky diode structure has been achieved by keeping it in two different environments, dark and room temperature. Its I-V characteristics have been discussed. Electrical parameters such as n and Φ_b,

which change due to the effect of TiO_2, which acts as an insulating interfacial layer in the structure, on the interfacial states, have been calculated using the TE technique. The electrical properties of the Al/TiO_2/p-Si Schottky diode structure are crucial for the evaluation of the interface states. Using the I-V curve generated under flat feeding, the n and Φ_b values were calculated to be 4.07 and 0.59 eV, respectively. In an ideal diode, n equals one. However, in our study, it is seen that the value of n is greater than 1. The reason for this is thought to be the presence of the insulator TiO_2 between the metal and the semiconductor and the distribution of the interface states between the metal and the semiconductor. It can be said that a thin oxide insulator layer of almost 5-20 Å thickness is naturally formed even in the most ideal Schottky diode structures. The deviation from the ideal situation caused by the dependence of the barrier height on the applied voltage is expressed as $1/n=1-(\delta\Phi_b/\delta V)$. Part of the voltage applied to the Schottky diode structure in the electrical characterisation process will fall on Φ_b and the rest on the interface insulating oxide layer, as can be seen from the expression. This makes Φ_b a function of the applied voltage. As part of the applied voltage falls on the oxide layer, the R_s value increases and the n parameter deviates from the ideal value of one, as the amount of voltage falling on the Schottky diode is reduced.

It has been investigated how the values of n and Φ_b are affected by changes in the interfacial layer formation techniques. A remarkable improvement was observed when comparing the experimental results found in our study with the existing literature. In this study, n and Φ_b were found to be lower than those obtained by ALD, spin coating and drop casting when comparing the basic diode parameters such as n and Φ_b obtained by RF magnetron sputtering with those obtained by other methods. In this study, the fundamental diode parameters of the Al/TiO_2/p-Si Schottky diode are higher than those obtained by chemical and electrochemical methods.

Metal-semiconductor interface insulators play a crucial role in various electronic and optoelectronic applications. Studies and models for these layers are developed to understand and optimize several key properties, such as electrical properties, thermal stability, interface states, optical properties, mechanical properties, chemical properties, dielectric properties, and integration with other materials. In our study, it is only interested in the electrical properties of the Schottky diode, and it is preferred TE and Cheung methods for investigating its electrical properties. Both methods have been validated in various studies, and they are generally reliable for assessing Schottky diode performance, provided that the measurements are taken under suitable conditions and the models are applied correctly. Using TE alongside the Cheung method provides a robust framework for analysing Schottky diodes, enabling accurate parameter extraction and deeper insights into device behaviour. For this reason, it's considered that the methods we followed are reliable for researching diode's electrical properties.

In summary, a MIS-type Al/TiO_2/p-Si Schottky diode structure was fabricated. Compared to SiO_2 and SnO_2, which are widely used in the microelectronics industry, TiO_2 has many advantages. Some of these advantages are the long service life due to its high adhesion capacity, the achievement of Schottky diode structures capable of fast switching, and the possibility of easy development of transistors and integrated circuits. Considering all these important points, it is predicted that original Schottky diode structures will be obtained far beyond the results presented in the study, which are close to ideal, by studying the changes to be made in the preparation of TiO_2 thin films using the RF magnetron sputtering technique. The results show that the fabricated diode can be used in photodiode, photosensor, and optoelectronic applications.

Contributions of the authors

Barış Polat: Supervision, performed the computations, and verified the analytical methods.
Elanur Dikicioğlu: Supervision, conducted the experiments, performed the computations, and verified the analytical methods.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] M. A. Mastro, "Power MOSFETs and diodes," *Gallium Oxide, Elsevier*, pp. 401-418, 2019.

[2] E. F. Krimmel, R. Hezel, U. Nohl, and R. Bohrer, "Schottky Diodes" *Si Silicon: Silicon Nitride in Microelectronics and Solar Cells, Heidelberg: Springer Berlin Heidelberg*, pp. 204-205,1991.

[3] D. Sadek, A. Zekry, H. Shawkey, and S. Kayed, "10 GHz Compact Shunt-Diode Rectifier Circuit Using Thin Film Ag/AZO Schottky Barrier Diode for Energy Harvesting Applications," In *International Conference on Human-Computer Interaction, Cham: Springer International Publishing*, pp. 436-445, 2022.

[4] C. Xu, P. Yu, and Y. Jiang, "A macro model of RF Schottky diode in 22-nm CMOS and its application," *Solid-State Electronics*, vol. 154, pp. 7-11, 2019.

[5] N. Daghestani *et al.*, "Room temperature ultrafast InGaAs Schottky diode based detectors for terahertz spectroscopy," *Infrared Physics & Technology*, vol. 99, pp. 240-247, 2019.

[6] M. Tilli and A. Haapalinna, "Chapter 1 - Properties of Silicon," in *Handbook of Silicon Based MEMS Materials and Technologies (Second Edition)*, M. Tilli, T. Motooka, V.-M. Airaksinen, S. Franssila, M. Paulasto-Kröckel, and V. Lindroos Eds. Boston: William Andrew Publishing, pp. 3-17, 2015.

[7] Ö. Güllü, T. Kilicoglu, and A. Türüt, "Electronic properties of the metal/organic interlayer/inorganic semiconductor sandwich device," *Journal of Physics and Chemistry of Solids*, vol. 71, no. 3, pp. 351-356, 2010.

[8] Ö. Güllü and A. Turut, "Electrical analysis of organic interlayer based metal/interlayer/semiconductor diode structures," *Journal of Applied Physics*, vol. 106, pp. 103717-103717, 2009.

[9] İ. Dökme, Ş. Altındal, T. Tunç, and İ. Uslu, "Temperature dependent electrical and dielectric properties of Au/polyvinyl alcohol (Ni, Zn-doped)/n-Si Schottky diodes," *Microelectronics Reliability*, vol. 50, no. 1, pp. 39-44, 2010.

[10] S. Sonmezoglu, S. Şenkul, R. Taş, G. Çankaya, and M. Can, "Electrical and interface state density properties of polyaniline–poly-3-methyl thiophene blend/p-Si Schottky barrier diode," *Solid State Sciences*, vol. 12, pp. 706-711, 2010.

[11] H. K. Henisch, "Metal-semiconductor Schottky barrier junctions and their applications," *Proceedings of the IEEE*, vol. 74, no. 6, pp. 894-894, 1986.

[12] S. M. Sze and K. K. Ng, *Physics Of Semiconductor Devices, 3RD ED*. Wiley India Pvt. Limited, 2008.

[13] E. H. Nicollian and J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology*. Wiley, 2002.

[14] H. Bentarzi, *Transport in Metal-Oxide-Semiconductor Structures : Mobile Ions Effects on the Oxide Properties*. 2011.

[15] M. Noman, M. A. Ashraf, and A. Ali, "Synthesis and applications of nano-TiO_2: a review," *Environmental Science and Pollution Research*, vol. 26, 2019.

[16] F. Wang and J. Wu, "Chapter 8 - Magnetron sputtering," in *Modern Ion Plating Technology*, F. Wang and J. Wu Eds.: Elsevier, 2023, pp. 189-228.

[17] G. M. Ali, "Performance analysis of planar Schottky photodiode based on nanostructured ZnO thin film grown by three different techniques," *Journal of Alloys and Compounds,* vol. 831, pp. 154859, 2020.

[18] H. Ferhati, F. Djeffal, K. Kacha, A. Bendjerad, and A. Benhaya, "Influence of TCO intermediate thin-layers on the electrical and thermal properties of metal/TCO/p-Si Schottky structure fabricated via RF magnetron sputtering," *Physica E: Low-dimensional Systems and Nanostructures*, vol. 106, pp. 25-30, 2019.

[19] R. Raj, H. Gupta, and L. P. Purohit, "Performance of RF sputtered V_2O_5 interface layer in p-type CdTe/Ag Schottky diode." *Optical Materials,* vol. 126, pp. 112176, 2022.

[20] N. AlAhmadi, "Metal oxide semiconductor-based Schottky diodes: a review of recent advances," *Materials Research Express*, vol. 7, p. 032001, 2020.

[21] D. Aldemir, A. Kökce, and A. Özdemir, "The comparison of the methods used for determining of Schottky diode parameters in a wide temperature range," pp. 1-1, 2017.

[22] E. H. Rhoderick, and R. H. Williams, "Metal-Semiconductor Contacts.", Second Edition, Clarendon Press, Oxford, 1988.

[23] S. K. Cheung and N.W. Cheung, "Extraction of Schottky diode parameters from forward current-voltage characteristics", *Appl. Phys. Lett.*, 1986.

[24] H. A. Bethe and M. I. o. T. R. Laboratory, *Theory of the Boundary Layer of Crystal Rectifiers*. Radiation Laboratory, Massachusetts Institute of Technology, 1942.

[25] S. B. Ocak, A. B. Selçuk, G. Aras and E. Orhan, "Electrical analysis of Al/ZnO/p-Si, Al/PMMA/p-Si and Al/PMMA/ZnO/p-Si structures: comparison study.", *Materials Science in Semiconductor Processing*, vol. 38, pp. 249-256, 2015.

[26] E. Orhan *et al.*, "Effect of Gadolinium on Electrical Properties of Polyethyleneimine Functionalized and Nitrogen-Doped Graphene Quantum Dot Nanocomposite Based Diode," *Advanced Electronic Materials,* vol. 9, no. 8, pp. 2300261, 2023.

[27] A. K. Bilgili, T. Güzel, and M. Özer, "Current-voltage characteristics of Ag/TiO_2/n-InP/Au Schottky barrier diodes," *Journal of Applied Physics*, vol. 125, no. 3, 2019.

[28] M. Ravinandan, P. K. Rao, and V. R. Reddy, "Analysis of the current–voltage characteristics of the Pd/Au Schottky structure on n-type GaN in a wide temperature range." *Semiconductor Science and Technology*, vol. 24, no.3, 2009.

[29] E. Elgazzar, A. D. E. M. Tataroğlu, A. A. Al-Ghamdi, Y. Al-Turki, W. A. Farooq, F. El-Tantawy, and F. Yakuphanoglu, "Thermal sensors based on delafossite film/p-silicon diode for low-temperature measurements." *Applied Physics A*, vol. 122, pp. 1-9, 2016.

[30] A. Ahaitouf, H. Srour, S. O. S. Hamady, N. Fressengeas, A. Ougazzaden, and J. P. Salvestrini, "Interface state effects in GaN Schottky diodes." *Thin Solid Films*, vol. 522, pp. 345-351, 2012.

[31] S. Asmontas, M. Anbinderis, J. Gradauskas, R. Juskenas, K. Leinartas, A. Lucun, A. Selskis, L. Staisiunas, S. Stanionyte, A. Suziedelis, A. Silenas, E. Sirmulis, "Low resistance TiO_2/p-Si heterojunction for tandem solar cells", *Materials*, vol. 13, p. 2857, 2020.

[32] S. Aksoy and Y. Caglar, "Structural transformations of TiO_2 films with deposition temperature and electrical properties of nanostructure n-TiO_2/p-Si heterojunction diode", *Journal of Alloys and Compounds*, vol. 613, pp. 330-337, 2014.

[33] İ. H. Taşdemir, Ö. Vural and İ. Dökme, "Electrical characteristics of p-Si/TiO_2/Al and p-Si/TiO_2-Zr/Al Schottky devices", *Philosophical Magazine*, vol. *96*(16), pp. 1684-1693, 2016.

[34] A.A. Hendi and F. Yakuphanoglu, "Graphene doped TiO_2/p-silicon heterojunction photodiode", *J. Alloys Compd.*, vol. 665, pp. 418–427, 2016.

[35] A. Dewasi and A. Mitra, "Effect of variation of thickness of TiO_2 on the photovoltaic performance of n-TiO_2/p-Si heterostructure", *J. Mater. Sci. Mater. Electron.*, 28, 18075–18084, 2017.

[36] S. Ruzgar, "Enhancement of the electrical performance of TiO_2/p-Si heterojunction diode by Gadolinium doping", *Appl. Phys. A*, vol. 126, pp. 770, 2020.

[37] O. Pakma, N. Serin, T. Serin, S. Altindal, "The influence of series resistance and interface states on intersecting behavior of I-V characteristics of Al/TiO2/p-Si (MIS) structures at low temperatures.", *Semiconductor science and technology*, vol. 23(10), p. 105014, 2008.

[38] M. Koca, M. Yilmaz, D. Ekinci and S. Aydogan, "Light sensitive properties and temperature-dependent electrical performance of n-TiO_2/p-Si anisotype heterojunction electrochemically formed TiO_2 on p-Si", *J. Electron. Mater.*, vol. *50*(9), pp. 5184-5195, 2021.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 916-922 YEAR: 2024
DOI: 10.17798/bitlisfen.1462590

Chromatographic Determination of Triticonazole in Oat Samples by High Performance Liquid Chromatography-Ultraviolet Detection (HPLC-UV)

Elif Seda KOÇOĞLU[1*]

[1] *Yıldız Technical University, Central Research Laboratory, İstanbul, Türkiye*
(ORCID: 0000-0001-6175-2335)

Keywords: Triticonazole, Fungicide, Pesticide, High pressure liquid chromatography, Oat samples.

Abstract

The chemical burden on ecosystems has been increased day by day because of the growing activities of agricultural sector. The chemicals that are used to protect plants and humans from different diseases in agricultural lands are known as pesticides. On the other hand, these chemicals can be caused a wide range of side effects on health and environment. The proposed study aimed to determine triticonazole belonging to the fungicide class by high pressure liquid chromatography equipped with ultraviolet detection system (HPLC-UV). Hence, suitable chromatographic conditions were applied to detect triticonazole at a certain retention time. 5 different validation parameters including limit of detection (LOD), limit of quantification (LOQ), percent relative standard deviation (RSD%), linear range (LR) and coefficient of linear regression (R^2) were investigated. A comprehensive validation of the developed method was achieved by all these values. The selectivity of the method for the analyte was ensured by selecting the wavelength of maximum absorption (263 nm), which is peculiar to the analyte. Very low LOD/LOQ values of presented method were recorded as 0.87 and 2.90 µg kg^{-1} under the proposed conditions, respectively. Triticonazole was extracted from oat samples by using acetonitrile and the presence of the triticonazole was investigated by analytical measurements. No detectable analytical signal was obtained for triticonazole in the samples at the retention time interval. The accuracy and applicability of the developed method to the real samples was verified with recovery experiments and this yielded satisfactory percent recoveries. Matrix matching calibration strategy was used to improve quantification accuracy for triticonazole. The results were obtained in the range of 92.9 - 101.4 % for different spiked concentrations of samples within linear range.

1. Introduction

Pesticide is described as different kind of chemicals such as insecticides, herbicides, fungicides, and rodenticides that are used to wipe out targeted pests [1], [2] and they are also useful to avoid or control the occurence of detrimental organisms [3]. They play a significant role in different cultivation steps to protect agricultural areas against unwanted organisms to avoid their negative effects on crops [4]-[6]. Pesticides are used for not only growing crops, but also the protection of crops after harvesting [4], [7]. In addition, the usage of pesticides has also been considerable increased with the increased demand for food materials obtained by agricultural practices [8]. Although pesticides are highly effective against improving the crop yield and quality [9], they can cause side effects on human and environmental health [10]. Pesticide ruins can relocate to several environmental media by wind currents or leaching and be the reason for the pollution of aqua, air and soil [9]. Moreover, the challenges of storage and disposal of pesticides safely have also hazardous effects on environment [11]. Excess usage of pesticides causes pollution of soil and water and prevent the absorption of essential nutrients by plants. The bioaccumulative and high toxic characteristics of agricultural pesticides have also caused to damage the human

*Corresponding author: ekocoglu@yildiz.edu.tr
Received: 01.04.2024, Accepted: 17.11.2024

being as they affect the functionality of endocrine and reproductive systems of living organisms [12]. However, the necessity of pesticide usage is known to fight various kind of pests in agricultural activities against some negations such as destroying crops and poor quality of the produced food [13], [14]. In summary, the usage of pesticides in modern agriculture is an important factor, having both benefits and drawbacks for human health and environmental quality [15].

Triazole fungicides are between azole pesticides that are the most widely preferred fungicides for agricultural activities such as growing various kind of fruits, vegetables and grain crops. They are called as endocrine disruptors that are considerable hazardous to the environment and the mankind because of their some chemical and physical properties [13] and they have been considered as the critical monitoring targets of environmental endocrine disruptors in many countries [16]. Exposed to the triazoles is easily occurred by nutrition, breathing and skin contact in daily life as they are used for many times in a growing season that catalyze their accumulation in products and transfer between environmental medias such as air, soil and water [17]. [(RS)-5-(4-chlorobenzylidene)-2,2-di- methyl-1-(1H-1,2,4- triazol-1-ylmethyl) cyclopentanol] is named as triticonazole and this broad-spectrum molecule has been widely preferred for the treatments of foliar and seed. In addition, the functionality of this molecule is based on interfering with biosynthesis and inhibiting steroid demethylation against phytopathogen [18]. Triticonazole hinders the progress of fungi mycelium both of over and inside of seeds and soil in seed applications. According to the literature data, triticonazole have various toxicological impacts on mammals [19]. Triticonazole could create toxicity effects for human health because this compound could also pass through blood brain barrier to deactivate a few cytochrome P450 enzymes (CYP) and cause skin irritation, carcinogenicity, and gentle distruption on endocrine system. Thus, developing a rapid, sensitive and a reliable method is necessary not only to determine triticonazole residues in real matrices, but also to determine this chemical for effective assessment of the risks to living organisms and environments [19], [20].

Selection of the most appropriate analytical determination method is based on physical and chemical characteristics of the related analytes because each of them has different properties of solubility and volatility owing to chemical structure of them [21]. In the literature, several traditional chromatographic separation methods have been used for determination of trace triazole pesticides including gas chromatography (GC) [22], supercritical fluid chromatography (SFC) [23], high pressure liquid chromatography coupled with ultraviolet detection (HPLC- UV) [24] and ultra performance liquid chromatography-tandem mass spectrometry (LC-MS/MS) [25]. In addition, many studies for monitoring pesticide residues in different mediums were also performed by various instrumental systems in Türkiye including GC-MS [26] and GC-MS/MS [27], LC-MS/MS [28]. In order for a compound to be determined by HPLC, it should be low-volatile and thermally unstable [29]. Thus, HPLC is applied to polar, thermolabile and nonvolatile compounds easily on the contrary to GC and is highly preferred for pesticide residue analysis. Low resolution obtained on HPLC can be enhanced by coupling selective detection devices to the system for reliable results. Ultraviolet (UV) spectroscopy is commonly opted for the analysis of pesticide residues. Although this technique has low selectivity, it is preferred for monitoring targets owing to low cost, simplicity and wide range of application [30].

Herein, the determination of triticonazole by HPLC-UV in oat samples was planned. High accuracy and precision for triticonazole in related matrix was aimed by the established method presented. The novelty of this work relies in the optimization and validation of an efficient, interference-free, rapid and simple method for the determination of triticonazole that is one of the more heavily used triazole fungicides in agriculture and is predicted to increase usage in the environment. Moreover, low detection limits have been achieved by using a single sample preparation step followed by high pressure liquid chromatography equipped with ultraviolet determination.

2. Material and Method

2.1. Instrumentation and Chromatographic Conditions

HPLC-UV is one of the analytical techniques for detecting the residues of triazole fungicides. Various analytical approaches [31], [32] were presented for the elution of triazoles in the literature. In this study, chromatographic separation and quantification of triticonazole was carried out by Shimadzu LC-20AT HPLC system coupled with UV detector. A Phenomenex-Aqua C18 (250 mm, 4.6 mm, 5 µm) analytical column was used to elute analyte by passing through with the help of a mobile phase that was composed of acetonitrile and ammonium formate buffer (50.0 mM, pH 4.0). Elution was achieved by the isocratic elution mode with a ratio of 70:30 (v/v). The flow rate of mobile phase was 1.2 mL min^{-1}, the injection volume of sample was 20 µL and the

operation wavelength of the UV detection system was 263 nm. The retention time of triticonazole and the total run time were 4.6 and 6.5 min, respectively. pH adjustments were carried out by a Hanna Instruments Edge® Multiparameter– HI2020 pH meter. An OHAUS Pioneer PA214C precision scale (0.10 mg) was used for weighing processes. Sample agitation processes were performed by a HAPA M−100 model ultrasonicator and a Kermanlar mechanical shaker.

2.2. Reagents and Chemicals

High purity standard of triticonazole (131983-72-7) was purchased from Sigma Aldrich (St. Louis, Missouri, USA) and 1000 mg L^{-1} of stock standard solution of triticonazole was prepared by dissolving in acetonitrile (Merck- Darmstadt, Germany). Analytical grade substances such as acetonitrile, formic acid, ammonium hydroxide (25%) were also supplied from Merck (Darmstadt, Germany). Preparation of different concentrations of triticonazole standard solutions was gravimetrically carried out.

2.3. Sample Preparations

Two brand of oat samples were supplied from a supermarket chain, İstanbul, Türkiye. Firstly, the oat samples were pounded in a mortar to pulverize to increase surface area. Then, 2.50 g of each pounded oat samples was weighed on an analytical balance and transferred into a sample tube and completed to final weight of 40 g with acetonitrile. In order to achieve an efficient mixing process, a mixing time of 10 minutes with an ultrasonicator and a mixing time of 15 minutes with a mechanical shaker were applied, respectively. Particulate-free samples were obtained by performing filtration process with a Whatman Qualitative Filter Papers (diameter: 12.5 cm, pore size: 11 µ) and then with a syringe filter (0.45 µm). The prepared samples were spiked within the linear range of the method and analyzed under their related optimum conditions.

3. Results and Discussion

All measurements were made on the basis of triple measurements and standard deviations were calculated to evaluate the precision and repeatability of the presented method. Calibration plot was obtained by using the average of peak areas of each measurement. Recovery percentages of triticonazole were calculated to check the both accuracy/applicability of the presented method.

3.1. Analytical Figures of Merit

Operational and chromatographic conditions mentioned in Section 2.1 were applied and the total run time required for triticonazole elution was completed within 6.5 mins. The signal of triticonazole was acquired within 4.6 min of the total run time. Linear regression analysis was carried out within 3.23-53619 µg kg^{-1}. Calibration plot were developed by the peak areas of the analyte obtained at each concentration value versus increasing concentrations of the analyte. As illustrated in Table 1, an excellent linearity was achieved for the analyte (R^2=1.000) over a wide concentration range. The values used to evaluate the analytical performance of the detection system such as limit of detection (LOD) and limit of quantification (LOQ) were calculated on the basis of lowest concentration of triticonazole, gives a signal to noise ratio (S/N) ≥3 and ≥10, respectively, as below and given in Table 1:

$$LOD = 3 * SD/Slope \quad (1)$$

$$LOQ = 10 * SD/Slope \quad (2)$$

Table 1. Analytical performance of HPLC-UV system and comparison to other methods in the literature

Method	LOD, µg kg^{-1}	LOQ, µg kg^{-1}	RSD%	Linear Range, µg kg^{-1}	R^2
HPLC-UV	0.87	2.90	12.5	3.23-53619 (y=76.526x+528.8)	1.000
LC-MS/MS [33]	1.54	5.15	-	-	0.999
LC-MS/MS [34]	3.00 (ng g^{-1})	10.00 (ng g^{-1})	4.09	2.50-200 (ng g^{-1})	0.997
DLLME/GC-FID [35]	11 (µg L^{-1})	36 (µg L^{-1})	-	50-80000 (µg L^{-1})	0.993

The standard deviation value (SD) was calculated by measuring the lowest concentration of the calibration plot seven times and was used to obtain the LOD and LOQ values of the system as 0.87 and 2.90 µg kg^{-1}, respectively. The RSD% value, calculated from the results obtained, represents the precision and indicated satisfactory repeatability.

Figure 1 shows overlay chromatograms obtained for triticonazole in the linear range under the optimum conditions.

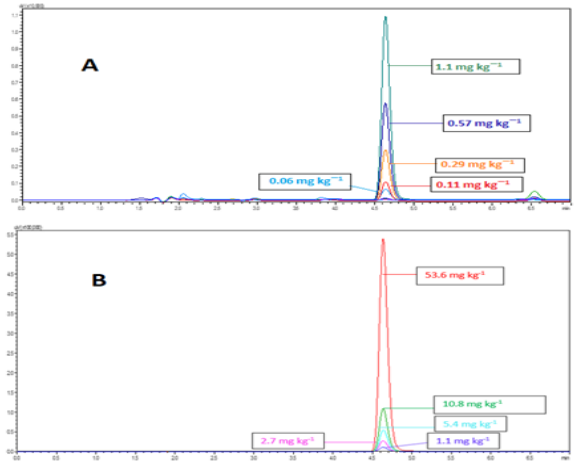


Figure 1. Overlay chromatograms of triticonazole in the linear range under the optimum conditions: A) The linear range of 3.23 µg kg^{-1}-1.1 mg kg^{-1}. B) The linear range of 1.1-53.6 mg kg^{-1}.

3.2. Real Sample Application

In order to ensure accuracy and applicability of the presented method, real sample application was carried out by two different oat samples. Matrix matching strategy was preferred to minimize effect of matrix interferences caused by matrix components. The method mentioned in Section 2.3 were applied to both of the oat samples belonging to different brands. The samples were prepared by spiking different concentrations of triticonazole within the linear range into the sample matrix prior to HPLC-UV detection. Measurements of unspiked sample solutions were also performed and no detectable analytical signal was observed for triticonazole in the samples at the retention time interval.

The percentage recovery results obtained from the spiked samples are given in Table 2. The results demonstrated the capability of the presented method for accurate quantification of trace levels of triticonazole in different oat samples and were found in the range of 92.9 and 101.4%. Overlay chromatograms of 1.0 mg kg^{-1} triticonazole standard solution and two different oat samples at spiked concentration of 1.0 mg kg^{-1} are demonstrated in Figure 2. As shown in Figure 2, matrix effect can overestimate or underestimate the results. Since there was no signal for the unspiked analysis, the high peak height/area recorded for the sample suggested overestimation or false positive for the analyte. It was therefore imperative to use matrix matching to mitigate the matrix effects and obtain accurate quantification results for the analyte.

Table 2. Percent recoveries of triticonazole in real samples

Samples	Spiked concentration, µg kg^{-1}	Recovery%
Oat-Brand A	58	95.1 ± 0.6
	110	96.9 ± 0.1
	275	101.4 ± 0.5
	549	101.0 ± 0.2
	1121	99.5 ± 0.3
	2826	100.3 ± 0.1
Oat-Brand B	54	92.9 ± 0.3
	110	96.3 ± 0.5
	268	99.7 ± 0.4
	552	100.5 ± 0.3
	1075	100.7 ± 0.2
	2895	100.2 ± 0.6

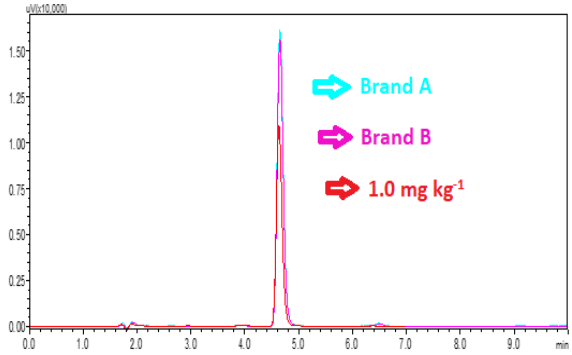


Figure 2. Overlay chromatograms of 1.0 mg kg^{-1} triticonazole standard solution and two different oat samples at spiked concentration of 1.0 mg kg^{-1}.

4. Conclusion and Suggestions

The proposed study has been performed to develop a simple and cheap determination method for the sensitive determination of triticonazole by HPLC-UV and to confirm the accuracy/applicability of the method by recovery studies. Very low detection/quantification limits (0.87 and 2.90 µg kg^{-1}, respectively) and quite wide linear range (3.23-53619 µg kg^{-1}) have been achieved for the analyte. Two different oat samples from different brands were used for real sample applications and satisfying recovery percentages (92.9 - 101.4%) for different spiked concentrations within linear range of the developed

method were obtained. The determination of triticonazole was carried out in a short analysis time period. As a result, the presented method was ensured various remarkable profits such speediness and easiness to detect triticonazole in oat samples.

Statement of Research and Publication Ethics

The authors of the paper declare that they comply with the scientific, ethical and quotation rules of BEU Fen Bilimleri Dergisi in all processes of the paper and that they do not make any falsification on the data collected. In addition, they declare that BEU Fen Bilimleri Dergisi and its editorial board have no responsibility for any ethical violations that may be encountered, and that this study has not been evaluated in any academic publication environment other than BEU Fen Bilimleri Dergisi.

References

[1] M. Eddleston, "Poisoning by pesticides," *Medicine (Abingdon)*, vol. 48, no. 3, pp. 214–217, 2020.

[2] M. A. Hassaan and A. El Nemr, "Pesticides pollution: Classifications, human health impact, extraction and treatment techniques," *Egypt. J. Aquat. Res.*, vol. 46, no. 3, pp. 207–220, 2020.

[3] N. K. Yetim, E. Hasanoğlu Özkan, C. Özcan, and N. Sarı, "Preparation of AChE immobilized microspheres containing thiophene and furan for the determination of pesticides by the HPLC-DAD method," *J. Mol. Struct.*, vol. 1222, no. 128931, p. 128931, 2020.

[4] E. Marzi Khosrowshahi, M. R. Afshar Mogaddam, M. A. Farajzadeh, Y. Javadzadeh, and M. Nemati, "Application of magnetic benzamide-coated SiC nanocomposite in effervesces-assisted DSPE combined with DLLME for the extraction of pesticides from fruit juice samples," *Microchem. J.*, vol. 181, no. 107749, p. 107749, 2022.

[5] L. Torović, G. Vuković, and N. Dimitrov, "Pesticide residues in fruit juice in Serbia: Occurrence and health risk estimates," *J. Food Compost. Anal.*, vol. 99, no. 103889, p. 103889, 2021.

[6] Q. Zhao *et al.*, "Sensitive detection of trace-level organophosphorus pesticides in fruit juices using a novel pH-responsive fluorescence probe based on 4-(morpholinomethyl) phenyl) boronic acid," *Microchem. J.*, vol. 169, no. 106541, p. 106541, 2021.

[7] A. Uclés, S. Herrera López, M. Dolores Hernando, R. Rosal, C. Ferrer, and A. R. Fernández-Alba, "Application of zirconium dioxide nanoparticle sorbent for the clean-up step in post-harvest pesticide residue analysis," *Talanta*, vol. 144, pp. 51–61, 2015.

[8] S. P, R. Thasale, D. Kumar, T. G. Mehta, and R. Limbachiya, "Human health risk assessment of pesticide residues in vegetable and fruit samples in Gujarat State, India," *Heliyon*, vol. 8, no. 10, p. e10876, 2022.

[9] M. Tudi *et al.*, "Agriculture development, pesticide application and its impact on the environment," *Int. J. Environ. Res. Public Health*, vol. 18, no. 3, p. 1112, 2021.

[10] P. Nicolopoulou-Stamati, S. Maipas, C. Kotampasi, P. Stamatis, and L. Hens, "Chemical pesticides and human health: The urgent need for a new concept in agriculture," *Front. Public Health*, vol. 4, 2016.

[11] D. Abong'o, S. Wandiga, I. Jumba, V. Madadi, and H. Kylin, "Impacts of pesticides on human health and environment in the River Nyando catchment, Kenya," *International Journal of Humanities, Arts, Medicine and Sciences*, vol. 2, pp. 1–14, 2014.

[12] A. Sharma *et al.*, "Worldwide pesticide usage and its impacts on ecosystem," *SN Appl. Sci.*, vol. 1, no. 11, 2019.

[13] M. A. Farajzadeh, M. R. A. Mogaddam, and A. A. Aghdam, "Comparison of air-agitated liquid-liquid microextraction technique and conventional dispersive liquid-liquid micro-extraction for determination of triazole pesticides in aqueous samples by gas chromatography with flame ionization detection," *J. Chromatogr. A*, vol. 1300, pp. 70–78, 2013.

[14] T. Goto, Y. Ito, H. Oka, I. Saito, H. Matsumoto, and H. Nakazawa, "Simple and rapid determination of N-methylcarbamate pesticides in citrus fruits by electrospray ionization tandem mass spectrometry," *Anal. Chim. Acta*, vol. 487, no. 2, pp. 201–209, 2003.

[15] C. Ozcan, U. K. Cebi, M. A. Gurbuz, and S. Ozer, "Residue analysis and determination of IMI herbicides in sunflower and soil by GC–MS," *Chromatographia*, vol. 80, no. 6, pp. 941–950, 2017.

[16] J. Nie, P. Yaro, K. He, and S. Zeng, "Development of a novel LC-MS/MS method for quantitation of triticonazole enantiomers in rat plasma and tissues and application to study on toxicokinetics and tissue distribution," *J. Pharm. Biomed. Anal.*, vol. 172, pp. 78–85, 2019.

[17] Y. Xu *et al.*, "A possible but unrecognized risk of acceptable daily intake dose triazole pesticides exposure—bile acid disturbance induced pharmacokinetic changes of oral medication," *Chemosphere*, vol. 322, no. 138209, p. 138209, 2023.

[18] Q. Zhang, B. Gao, M. Tian, H. Shi, X. Hua, and M. Wang, "Enantioseparation and determination of triticonazole enantiomers in fruits, vegetables, and soil using efficient extraction and clean-up methods," *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, vol. 1009–1010, pp. 130–137, 2016.

[19] D. L. Roman *et al.*, "Assessment of the effects of triticonazole on soil and human health," *Molecules*, vol. 27, no. 19, p. 6554, 2022.

[20] M. K. Draskau *et al.*, "In vitro and in vivo endocrine disrupting effects of the azole fungicides triticonazole and flusilazole," *Environ. Pollut.*, vol. 255, no. Pt 2, p. 113309, 2019.

[21] S. E. Romniou, K. Nana, M. Dasenaki, E. Komaitis, and C. Proestos, "Development and validation of pesticide residues determination method in fruits and vegetables through liquid and gas chromatography tandem mass spectrometry (LC-MS/MS and GC-MS/MS) employing modified QuEChERS method and a centrifugal vacuum concentrator," *Agriculture*, vol. 12, no. 11, p. 1936, 2022.

[22] M. A. Farajzadeh, M. R. A. Mogaddam, and H. Ghorbanpour, "Development of a new microextraction method based on elevated temperature dispersive liquid–liquid microextraction for determination of triazole pesticides residues in honey by gas chromatography-nitrogen phosphorus detection," *J. Chromatogr. A*, vol. 1347, pp. 8–16, 2014.

[23] Q. Tan *et al.*, "Stereoselective quantification of triticonazole in vegetables by supercritical fluid chromatography," *Talanta*, vol. 164, pp. 362–367, 2017.

[24] T. Tang, K. Qian, T. Shi, F. Wang, J. Li, and Y. Cao, "Determination of triazole fungicides in environmental water samples by high performance liquid chromatography with cloud point extraction using polyethylene glycol 600 monooleate," *Anal. Chim. Acta*, vol. 680, no. 1–2, pp. 26–31, 2010.

[25] Z. Zhang *et al.*, "Simultaneous enantioselective determination of the chiral fungicide prothioconazole and its major chiral metabolite prothioconazole-desthio in food and environmental samples by ultraperformance liquid chromatography-tandem mass spectrometry," *J. Agric. Food Chem.*, vol. 65, no. 37, pp. 8241–8247, 2017.

[26] C. Ozcan, "Determination of organochlorine pesticides in some vegetable samples using GC-MS," *Pol. J. Environ. Stud.*, vol. 25, no. 3, pp. 1141–1147, 2016.

[27] O. Kuzukiran, I. Simsek, T. Yorulmaz, B. Yurdakok-Dikmen, O. Ozkan, and A. Filazi, "Multiresidues of environmental contaminants in bats from Turkey," *Chemosphere*, vol. 282, no. 131022, p. 131022, 2021.

[28] T. Balkan and H. Karaağaçlı, "Determination of 301 pesticide residues in tropical fruits imported to Turkey using LC–MS/MS and GC-MS," *Food Control*, vol. 147, no. 109576, p. 109576, 2023.

[29] A. Jedrzejczuk, K. Góralczyk, K. Czaja, P. Struciński, and J. K. Ludwicki, "High performance liquid chromatography: application in pesticide residue analysis," *Rocz. Panstw. Zakl. Hig.*, vol. 52, no. 2, pp. 127–138, 2001.

[30] S. Nawaz, "PESTICIDES AND HERBICIDES | Residue Determination," in *Encyclopedia of Food Sciences and Nutrition*, Elsevier, 2003, pp. 4487–4493.

[31] Z. Lin, Y. Zhang, Q. Zhao, Y. Cui, A. Chen, and B. Jiao, "In-situ decomposed nanofluids dispersive liquid-phase microextraction for detection of seven triazole fungicidets in fruit juices and tea drinks," *Microchem. J.*, vol. 173, no. 106973, p. 106973, 2022.

[32] R. Fu, T. Ren, Y. Zhang, Q. Lu, X. Guo, and X. Di, "Hexafluoroisopropanol-based supramolecular solvent for liquid phase microextraction of triazole fungicides in drinking water and beverages," *Microchem. J.*, vol. 191, no. 108842, p. 108842, 2023.

[33] M. Kanbolat, T. Balkan, and K. Kara, "Verification of QuEChERS method for the analysis of pesticide residues in some fruits grown in Tokat province, Turkey, and their risk assessment," *Tarım bilim. derg.*, vol. 29, no. 2, pp. 573–588, 2023.

[34] A. Elmastas, A. Umaz, V. Pirinc, and F. Aydin, "Quantitative determination and removal of pesticide residues in fresh vegetables and fruit products by LC-MS/MS and GC-MS/MS," *Environ. Monit. Assess.*, vol. 195, no. 2, p. 277, 2023.

[35] M. A. Farajzadeh, D. Djozan, M. R. A. Mogaddam, and M. Bamorowat, "Extraction and preconcentration technique for triazole pesticides from cow milk using dispersive liquid-liquid microextraction followed by GC-FID and GC-MS determinations," *J. Sep. Sci.*, vol. 34, no. 11, pp. 1309–1316, 2011.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 923-938 YEAR: 2024
DOI: 10.17798/bitlisfen.1471912

A Parameter-Oriented FFT Signal Processing App

Sıtkı AKKAYA[1*],
[1]*Department of Electrical and Electronics Engineering of Faculty of Engineering and Natural Sciences, Sivas University of Science and Technology, Sivas, Turkey*
(ORCID: 0000-0002-3257-7838)

Keywords: Power Quality, FFT, main frequency, sampling frequency, measurement window, visible thinking.

Abstract
In power systems, noise, harmonics, and interharmonics arise in electrical signals due to varying sources and loads, affecting signal purity. Continuous monitoring and accurate analysis of electrical signals are mandatory. The Fast Fourier Transform (FFT) continuously analyzes electrical signals using sliding windows per the IEC-61000-4-7 standard. Parameters from this analysis are compared with threshold values specified in the IEEE-1159 standard. However, variable conditions and factors like sampling frequency, measurement window, main frequency, additional component frequencies, and Signal-to-Noise Ratio (SNR) cause measurement errors. These challenges complicate accurate measurement, leading to errors in preventive measures and control procedures. Understanding the effects of these parameters and improving methods is crucial. The Visible Thinking pedagogical framework is effective in this achievement. This study highlights the importance of parameter selection for FFT and investigates FFT responses to different parameters with synthetical and experimental signal examples. It also presents measurement errors due to signal changes, and a basic interface design shows these errors. Small changes, like a 1/2000 shift in sampling frequency, a 0.5 Hz shift in fundamental frequency, or a 1/1000 difference in the measurement window, cause significant errors. These findings underscore the need for careful parameter selection for accurate computation and signal monitoring, showing the need for FFT method improvements to adapt to changing conditions.

1. Introduction

Within the power systems domain, strict adherence to 50 Hz frequency and 220 V_{rms} standards determined in Turkey is paramount, as the main frequency and amplitude serve as fundamental determinants [1]-[4]. Nevertheless, these critical parameters are susceptible to fluctuations attributed to diverse system components [5]-[8], encompassing nonlinear loads, illumination systems, electrical motors, arc furnaces, and welding operations [9]-[12]. Substantial scholarly inquiry has meticulously scrutinized the complexities inherent in this phenomenon, conducting rigorous analyses and taxonomies [5]–[13]. Standards such as IEEE 1159 and 1459 are dedicated to elucidating power quality parameters and facilitating the revealing evaluation of the measurements in succession. Moreover, detailed guidelines delineating the testing and assessment of disturbances within power systems are expounded in the standards, and an analysis window for power systems is prescribed as 0.2 seconds [1]–[4]. Incorporating disturbance components into the primary signal poses inherent risks and may introduce measurement errors, impacting systems and individuals. These erroneous measurements rather exacerbate the complexity of the issue, impeding the precise determination of values pertaining to components that deserve to be eliminated.

The periodic surveillance and power systems signals analysis are indispensable endeavors. The efficacy of robust adjustment mechanisms, by compensation of filtering, hinges upon the compatibility between signal parameters and the chosen analytical approach.

[*]Corresponding author: sakkaya@sivas.edu.tr
Received: 23.04.2024, Accepted: 23.12.2024

According to the article in [14], the Visible Thinking pedagogical framework is effective in achieving a good complement to traditional lecture–tutorial systems. Moreover, the other study on programming, which is a difficult area to understand due to its notional, proved the positive effect on visible thinking [15]. Indeed, there are limited similar studies for FFT and harmonics for electrical systems directly. For example, a mobile application to teach simple harmonic motions to high school students was made in [16]. A study in [17], presented the VisualHarmony program to learn the harmony effects in musical components by using harmonic analysis. A study in [18], shows that Mobile apps can be used experimentally for physics education. In this study, the ready-made programs are utilized to implement the experiments. In another study, an application of Fourier Theory for optics was implemented for engineering mathematics education [14]. The positive effect of visible thinking is shown in the article. An education tool for signals and systems achieved in MATLAB proposed different transform methods roughly and with some exercises [19]. Different smartphone applications like the Signal Generator, the Function Generator as a tone generator, and the Oscilloscope, the FFTWave as a sound oscilloscope, were performed for teaching the Fourier Series (F.S.) in [20] without any additional programming. It was revealed that the learning rate of F.S. was improved with this method based on Visible Thinking, as shown in the article. A basic signal processing app for sounds was performed with MATLAB [21]. An educational software interface for power electronic applications carried out in [22] provided amplitude monitoring and FFT analysis coarsely. A Fourier Series app based on Android for undergraduate education was performed in [23]. A software based on ISE (Integrated Software Environment) was prepared for harmonics monitoring for FPGA applications in [24]. Another MATLAB app for harmonics was made in [25], and it gave the harmonics results of THD (Total Harmonics Distortion).

As understood from the literature, all solve different problems, and only some approach FFT or harmonics and interharmonics of electrical systems roughly. Consequently, the present study tries to conduct a comprehensive comparative evaluation of parameters with different specifications. In this study, a MATLAB-based software is proposed that enables the generation of harmonic or interharmonic electrical signals, comparative samples with a predefined interface and detailed FFT analysis with different parameters such as sampling frequency (Hz), length (bit), frequency (Hz), main amplitude (V), inserted component frequency, inserted component amplitude (V), SNR level (dB) according to user definition. Moreover, subinterfaces of predetermined comparative parameters based on synthetic and experimental signals are conducted to make the effects of the FFT parameters easy for users to understand.

This manuscript is organized by the following. Section two provides the methodology, including the power signal model and the signal processing model of FFT. The third section details the proposed app that originated on FFT parameters. Time and frequency domain responses obtained from the app are indicated in the next section. The results and effects of the different parameters of FFT are also discussed in this section. A summary of the study, contributions, and future studies are presented in the conclusion.

2. Material and Method

The main amplitude and frequency have crucial roles in power systems engineering, and they are widely acknowledged for their significance. Optimal operational performance within these systems is contingent upon the fidelity of the transmitted signal to a pure sinusoidal waveform characterized by the fundamental frequency. Nonetheless, odd-order harmonics, particularly those within the initial three orders specified by the IEC 61000-4-7 standard, can potentially disrupt these fundamental parameters. Such harmonics or interharmonics may manifest with varying amplitudes, thereby introducing an array of noise manifestations.

2.1. Signal Model

Eq. (1) presents a comprehensive representation of a signal with inserted components. Within this mathematical formulation, the amplitude of the main component is explicitly denoted as $\alpha_{main} = 1$ while the ensuing coefficients $\alpha_{inserted}$ delineate the amplitude of the corresponding inserted harmonic / interharmonic component. The frequencies of the components are also called with similar indices. The term $noise_{SNR}(t)$ signifies the Additive White Gaussian Noise (AWGN) integrated within the signal framework.

$$x(t) = \alpha_{main}\sin(2\pi f_{main}t) + \alpha_{inserted}\sin(2\pi f_{inserted}t) + noise_{SNR}(t) \quad \textbf{(1)}$$

2.2. Signal Processing with FFT

Eq. (2) presents FFT for x called by y with length N:

$$y(k) = \sum_{j=1}^{N} x(j) W_N^{(j-1)(k-1)}, W_N = e^{(-2\pi i)/n} \quad \textbf{(2)}$$

All transformations with different parameters will be done with FFT methods in this study and will not be defined in detail [14]. An input signal as x(t) modeled by Eq. (1) and FFT of it are available in Figure 1. a. and b. consecutively.

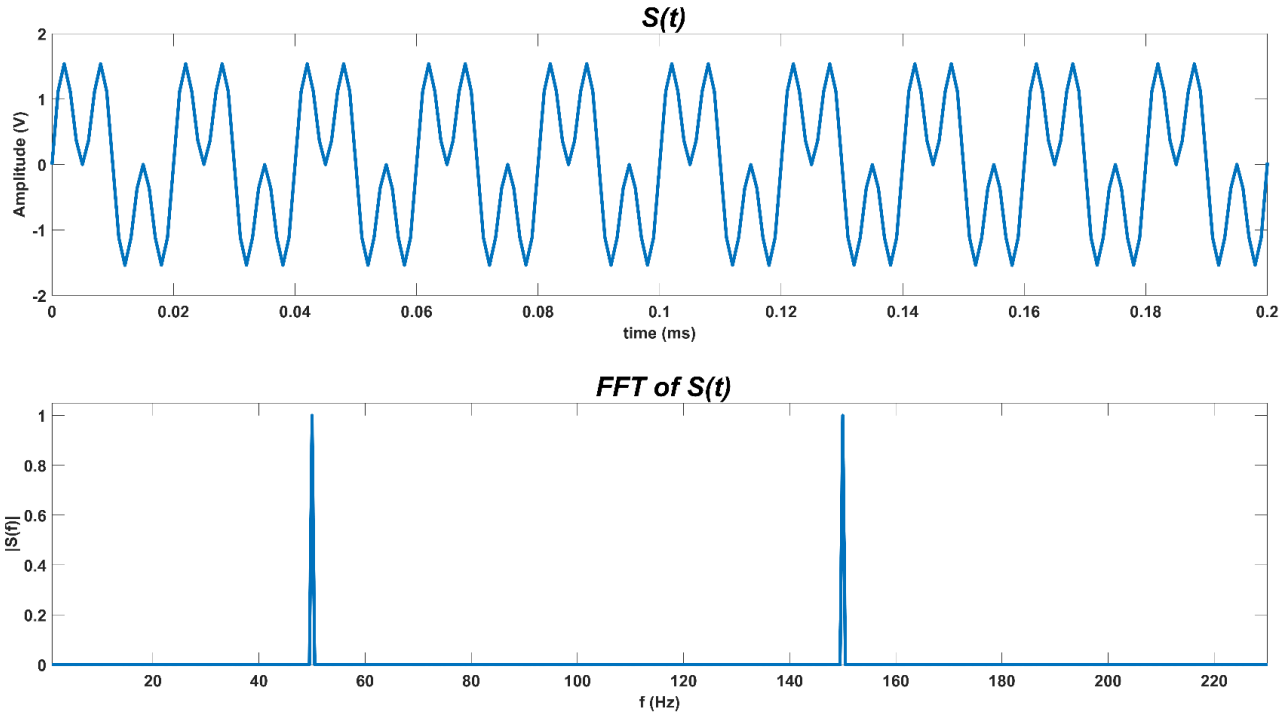


Figure 1. Input Signal and FFT of It

3. Proposed App for Parameter-Oriented FFT

In this study, a MATLAB-based software is proposed that enables the generation of harmonic or interharmonic electrical signals, comparative samples with a predefined interface and detailed FFT analysis with different parameters such as sampling frequency (Hz), length (bit), frequency (Hz), main amplitude (V), inserted component frequency, inserted component amplitude (V), SNR level (dB) according to user definition.

This study proposed an app with three main subprograms, as revealed in Figure 2. The first is a user-defined input interface that allows the user to input phase and time variability. The second provides comparative predefined synthetic signals based on different parameters like sampling frequency, signal length, frequencies of main and inserted components, and SNR level. The last one shows similar signals based on the experimental dataset.

The program is set based on the default parameters available in Table 1. Moreover, the resolution of the FFT and window also important for an accurate measurement are given in Table 1. All parameters are explained why these values are selected as default.

- For this table, the sampling frequency of the signal is selected as $f_s = 1000$ Hz, which provides the 10^{-3}sec according to most basic studies for simplicity.
- The length of the signal is gotten as *L=1500* default. However, it is arranged to different values and performed according to the standard of IEC 61000-4-30 in the simulation studies.
- Window and the resolution effects of the parameters can be obtained from the parameters above. So, they are also performed according to the related values consisting of standard values as $win = 0.2\ sec$ and $res = 5\ Hz$.
- The frequencies of the components f_0 and $f_{inserted}$ are selected with the most common values as 50 and 150 Hz. These are performed for the most common frequency ranges.
- SNR can be encountered between 0 to infinity, so it is arranged as inf, and it can be arranged between 0 to inf.

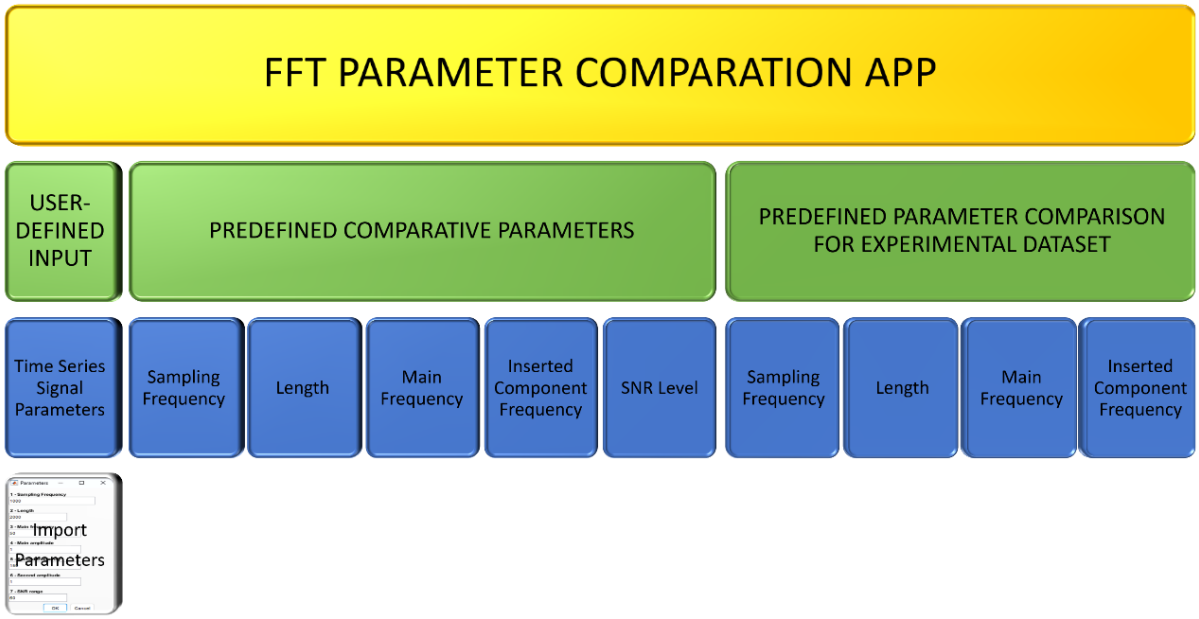


Figure 2. Block Structure of the App

Table 1. Default Parameters of the systems

Default Parameters of the systems	
Sampling Frequency (Hz)	$f_s = 1000$
Sampling period (sec)	$T = 1/f_s$
Length of signal (bit)	$L = 1500$
Window (sec)	$win = L * T$
Resolution (Hz)	$res = f_s/L$
Main Frequency (Hz)	$f_0 = 50$
Inserted Component Frequency (Hz)	$f_{inserted} = 150$
SNR Level (dB)	∞

3.1. Input Option

The program presented allows for user login and predefined operations. The interface in Figure 3 offers the options of selecting User-Defined Input, Predefined Comparative Parameters or Predefined Comparative Experimental Parameters by selecting in order of 1, 2 or 3.

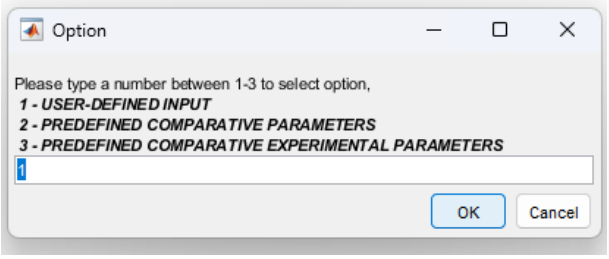


Figure 3. Input Selection Sub-Interface

3.1.1. User-Defined Input

The user-defined input interface that appears in Figure 4 makes an easier arrangement of parameters like sampling frequency, length, main frequency, main amplitude, inserted component frequency, inserted component frequency, and SNR level, successively as seen in the figure.

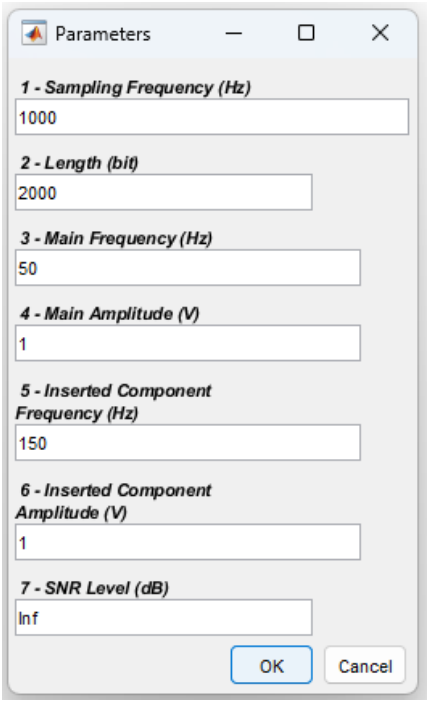


Figure 4. User-Defined Input Sub-Interface

3.1.2. Predefined - Comparative Parameters

The Predefined Parameter interface shown in Figure 5, is usable to see comparative results that align with the selected parameter while the others are default parameters. The details of the results will be observed in the next section.

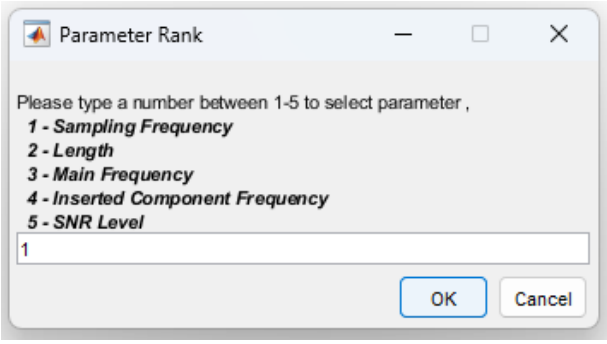


Figure 5. Parameter Selection Sub-Interface for Predefined Comparative Parameters

3.1.3. Predefined - Comparative Experimental Parameters

The Predefined Experimental Parameter interface shown in Figure 6, is usable to see comparative results that align with the selected parameter while the others are default parameters. The details of the results will be observed in the next section.

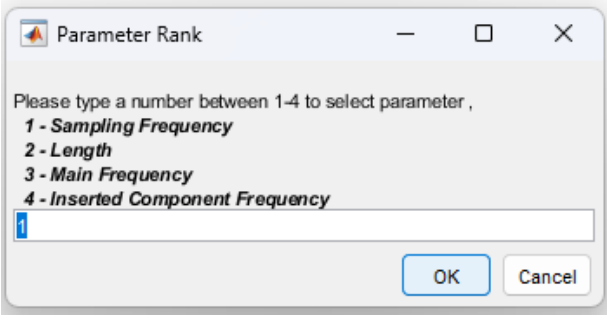


Figure 6. Parameter Selection Sub-Interface for Predefined Comparative Experimental Parameters

4. Results And Discussion

In this study, an application for FFT is based on the parameters. The computation was performed with MATLAB 2024a on a computer with Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz 2.90 GHz, 64-bit processor, 32 GB RAM. The elapsed time is 0.071373 seconds, which is less than a window length of 0.2 seconds. That means it leads to only 0.071373 seconds delaying the process of a window, and it can be tolerated until the other window acquisition. In this section, the comparative results for predefined values are obtained for the "5" different parameters with figures and tables in the case of synthetical and experimental datasets as follows.

4.1. Synthetical Results

4.1.1. Sampling Frequency

Figure 7 demonstrates the effects of the different sampling frequencies. It is clear that the appropriate sampling frequency selection is important for accurate measurement in the frequency spectrum. The results of this comparison are given in Table 2. Both components are measured with 0% error in the case of $f_s = 1000$, while important errors occur in terms of frequency and amplitude under the other cases. These errors depend on frequency resolution.

Table 2. Results for different sampling frequencies

	Main Component			***Inserted Component***		
Sampling Frequency, f_s	1000	2000	2994	1000	2000	2994
Frequency Resolution (Hz)	0.6667	1.3333	1.996	0.6667	1.3333	1.996
Nominal Amplitude (V_{pu})	1	1	1	1	1	1
Nominal Frequency (Hz)	50	50	50	150	150	150
Measured Amplitude (V_{pu})	1	0.647	0.996	1	0.637	0.961
Error (%)	**0**	**-35.3**	**-0.4**	**0**	**-36.3**	**-3.9**
Measured Frequency (Hz)	50	49.33	49.9	150	150.67	149.7
Error (%)	**0**	**-1.34**	**-0.2**	**0**	**+0.45**	**-0.2**

4.1.2. Length

Figure 8 demonstrates the effects of the different signal lengths. The appropriate length selection is important for accurate measurement, as shown in this figure. The results of the comparison are available in Table 3. Both components are measured with 0% error under the case of $L = 1500$ and 200, while important errors occur in terms of frequency and amplitude under the case of $L = 1001$. Because the frequency resolution is 0.999 that leads to a -0.1 % and -0.45% frequency deviation for both components. This result is related directly to the measurement window.

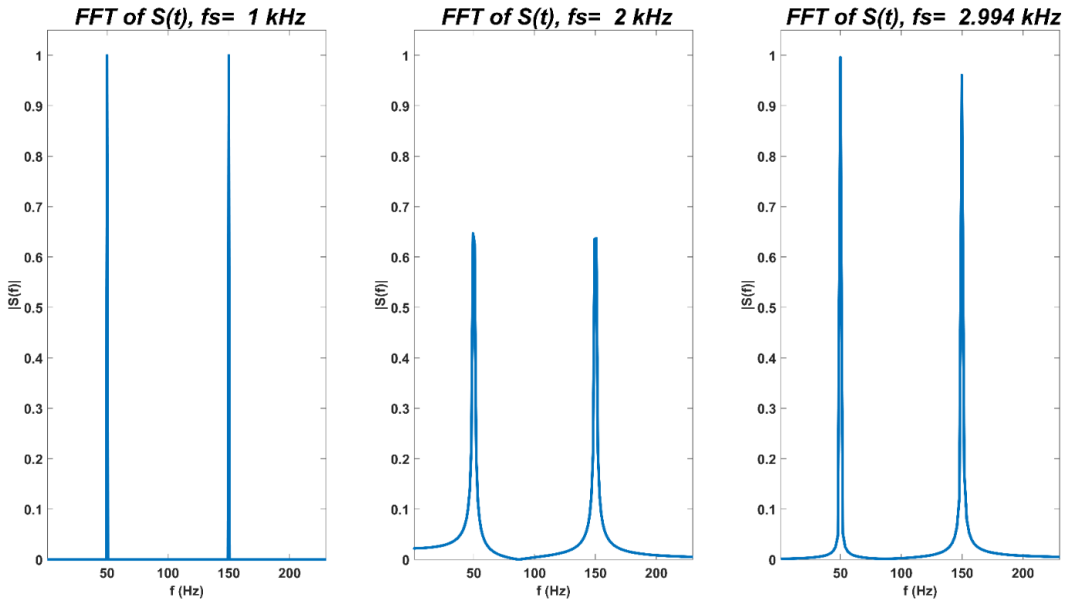


Figure 7. Effects of different sampling frequencies

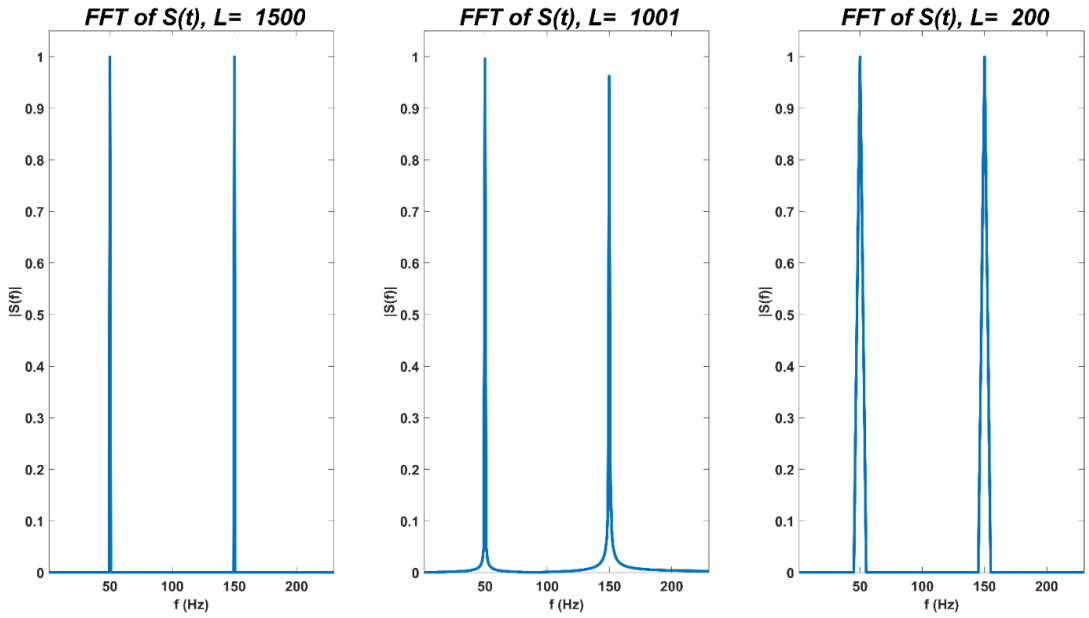


Figure 8. Effects of different signal lengths

Table 3. Results for different signal lengths

	Main Component			***Inserted Component***		
Length of Signal, L	1500	1001	200	1500	1001	200
Frequency Resolution (Hz)	0.667	0.999	5	0.6667	0.999	5
Nominal Amplitude (V_{pu})	1	1	1	1	1	1
Nominal Frequency (Hz)	50	50	50	150	150	150
Measured Amplitude (V_{pu})	1	0.996	1	1	0.962	1
Error (%)	**0**	**-0.4**	**0**	**0**	**-3.8**	**0**
Measured Frequency (Hz)	50	49.95	50	150	149.85	150
Error (%)	**0**	**-0.1**	**0**	**0**	**-0.45**	**0**

4.1.3. Main Frequency

Figure 9 demonstrates the effects of the main frequency value. An appropriate main frequency value is also important for accurate measurement, as shown in this figure. The results of this effect are presented in Table 4. Both components are measured with 0% error in the case of $f_{main} = 50\,Hz$, while important errors occur in terms of frequency and amplitude under the cases of $f_{main} = 49.5\,Hz$, and 50.5 Hz for the main component. The inserted component is measured accurately in all cases. Because the frequency value of the main component affects itself independently from the other components.

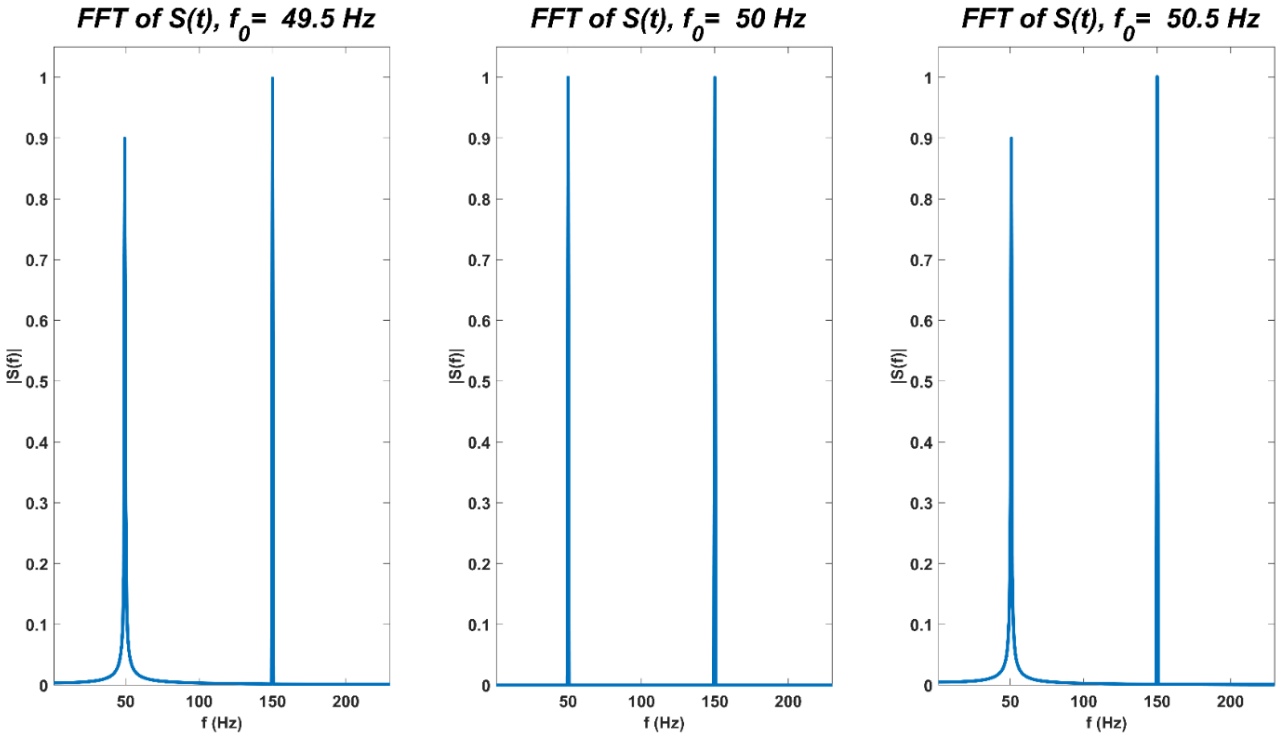


Figure 9. Effects of different main frequencies

Table 4. Results for different main frequencies

	Main Component			*Inserted Component*		
Main Frequency, f_0 (Hz)	49.5	50	50.5	150	150	150
Frequency Resolution (Hz)	0.667	0.667	0.667	0.667	0.667	0.667
Nominal Amplitude (V_{pu})	1	1	1	1	1	1
Nominal Frequency (Hz)	50	50	50	150	150	150
Measured Amplitude (V_{pu})	0.8998	1	0.8998	1	1	1
Error (%)	**-10.02**	**0**	**10.02**	**0**	**0**	**0**
Measured Frequency (Hz)	49.333	50	50.667	150	150	150
Error (%)	**-1.334**	**1**	**1.334**	**0**	**0**	**0**

4.1.4. Inserted Component Frequency

Figure 10 demonstrates the effects of the inserted component frequency value. An appropriate inserted component frequency value is also important for accurate measurement, as shown in this figure. The results of this effect are presented in Table 5. Both components are measured with 0% error in the case of $f_{inserted} = 100\ Hz$, and 150 Hz, while important errors occur in terms of frequency and amplitude under the cases of $f_{inserted} = 149\ Hz$ for just the inserted component. The main component is measured accurately in all cases. Similar to that, in the cases of main frequency, the value of the inserted component frequency only affects itself independently from the other components.

Table 5. Results for different inserted component frequencies

	Main Component			*Inserted Component*		
Inserted Component Frequency, $f_{inserted}$ (Hz)	50	50	50	150	149	100
Frequency Resolution (Hz)	0.667	0.667	0.667	0.667	0.667	0.667
Nominal Amplitude (V_{pu})	1	1	1	1	1	1
Nominal Frequency (Hz)	50	50	50	150	150	150
Measured Amplitude (V_{pu})	1	1	0	1	0.667	1
Error (%)	**0**	**0**	**0**	**0**	**-33.3**	**0**
Measured Frequency (Hz)	50	50	50	150	148.67	150
Error (%)	**0**	**0**	**0**	**0**	**-0.89**	**0**

4.1.5. SNR Level

Figure 11 demonstrates the effects of SNR level for the inserted noise. The noise level is also important for accurate measurement as seen in this figure. The results of this effect are presented in Table 6. Both components are measured with almost errorless under the case of $SNR = 50\,dB$, and $30\,dB$ while important errors occur in terms amplitude in the case of $SNR = 3\,dB$ for both components.

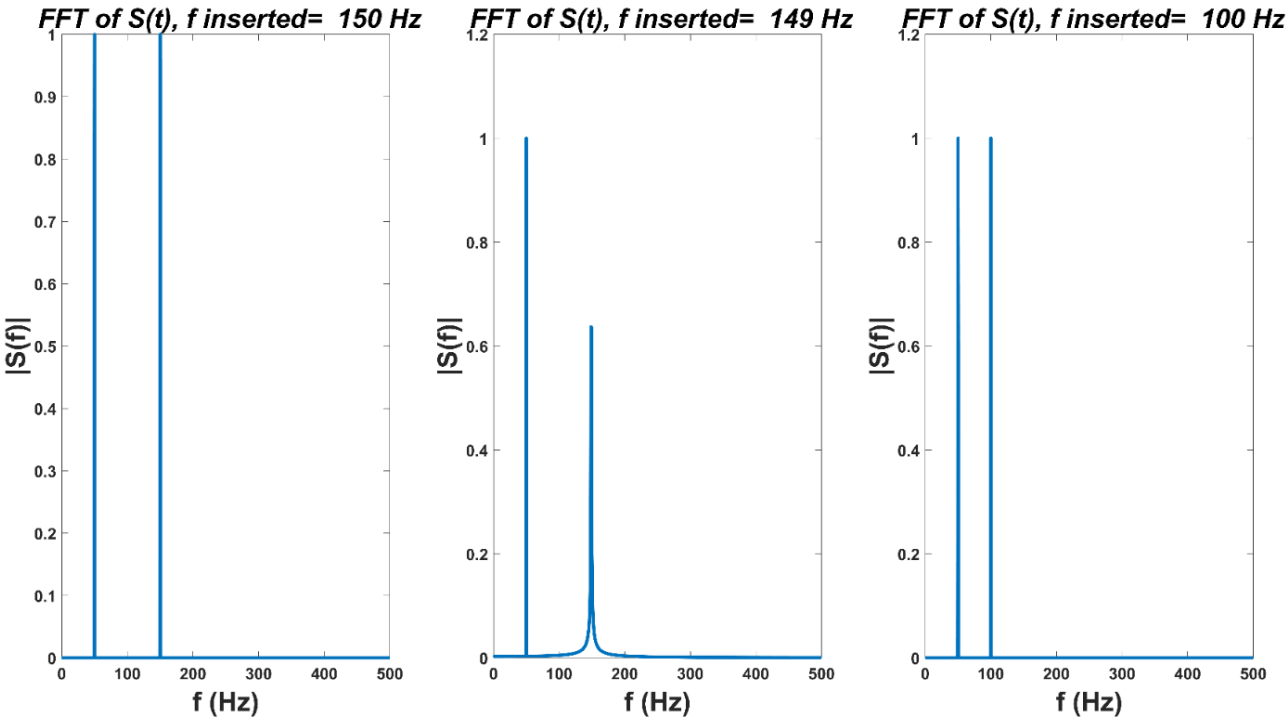


Figure 10. Effects of different inserted component frequencies

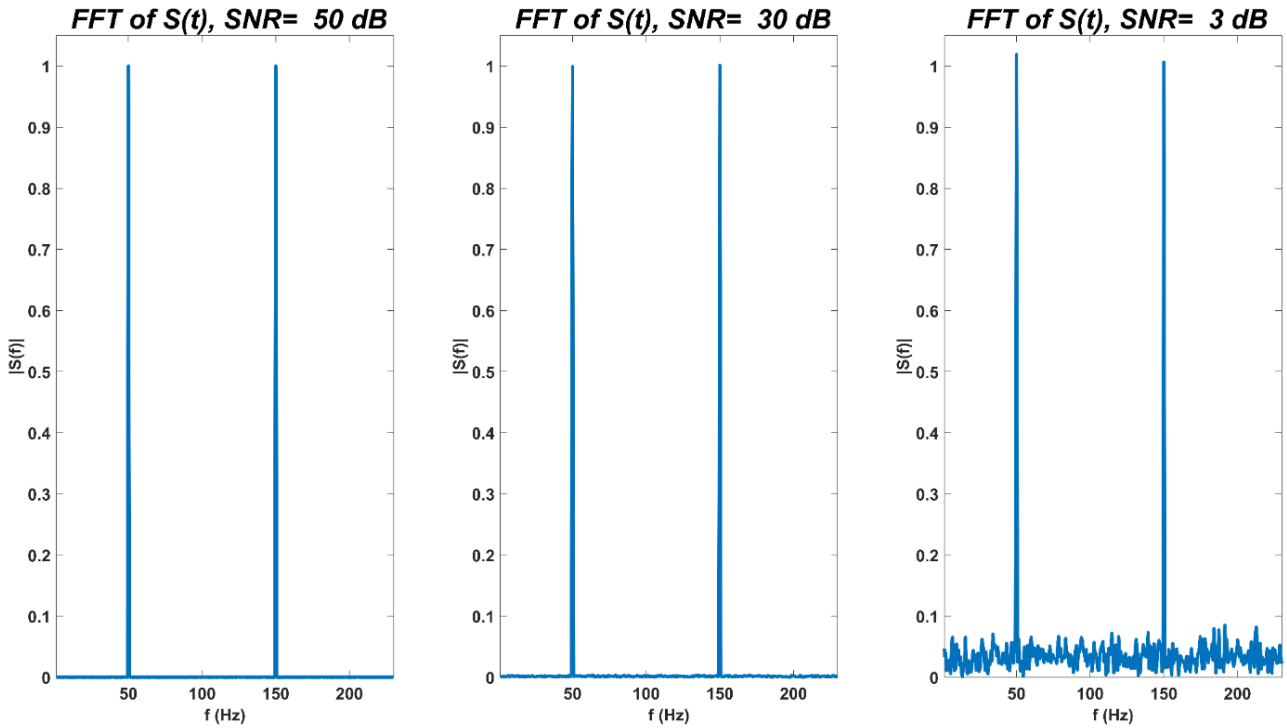


Figure 11. Effects of different SNR levels

Table 6. Results for different SNR levels

	Main Component			*Inserted Component*		
SNR (dB)	50	30	3	50	30	3
Frequency Resolution (Hz)	0.667	0.667	0.667	0.667	0.667	0.667
Nominal Amplitude (V_{pu})	1	1	1	1	1	1
Nominal Frequency (Hz)	50	50	50	150	150	150
Measured Amplitude (V_{pu})	0.9999	0.9991	0.9857	1.00003	1.0004	0.957
Error (%)	**-0.01**	**-0.093**	**-1.433**	**0.003**	**0.041**	**-4.32**
Measured Frequency (Hz)	50	50	50	150	150	150
Error (%)	**0**	**0**	**0**	**0**	**0**	**0**

4.2. Experimental Results

An experimental dataset is generated by an adjustable signal generator in the laboratory. According to the ability of the generator, all comparisons are carried out except for the SNR level as follows. Default parameters for the experimental dataset are selected based on the properties and IEC 6100-4-7 standard recommendations of window 0.2 secs, as in Table 7.

Table 7. Default parameters of the systems for experimental dataset.

Default parameters of the systems	
Sampling Frequency (Hz)	$f_s = 5000$
Sampling period (sec)	$T = 1/f_s$
Length of signal (bit)	$L = 1000$
Window (sec)	$win = L * T$=0.2 sec
Resolution (Hz)	$res = f_s/L$ = 5 Hz
Main Frequency (Hz)	$f_0 = 50$
Inserted Component Frequency (Hz)	$f_{inserted} = 150$
SNR Level (dB)	*Natural Grid Noise*

4.2.1. Sampling Frequency

Figure 12 demonstrates the effects of the different sampling frequencies for experimental time series data. The appropriate sampling frequency selection is important for accurate measurement in the frequency spectrum. The results of this comparison are given in Table 8. Both components are measured with a low error rate in the cases of $f_s = 1000$ and 5000 Hz, while important errors occur in terms of frequency and amplitude under the case of $f_s = 7150$. These errors depend on frequency resolution. The measured amplitudes of the components are *1.024* and *0.9832* V_{pu} because of the natural grid noise of the signal. This is also true for the other cases of the experimental parameters as follows.

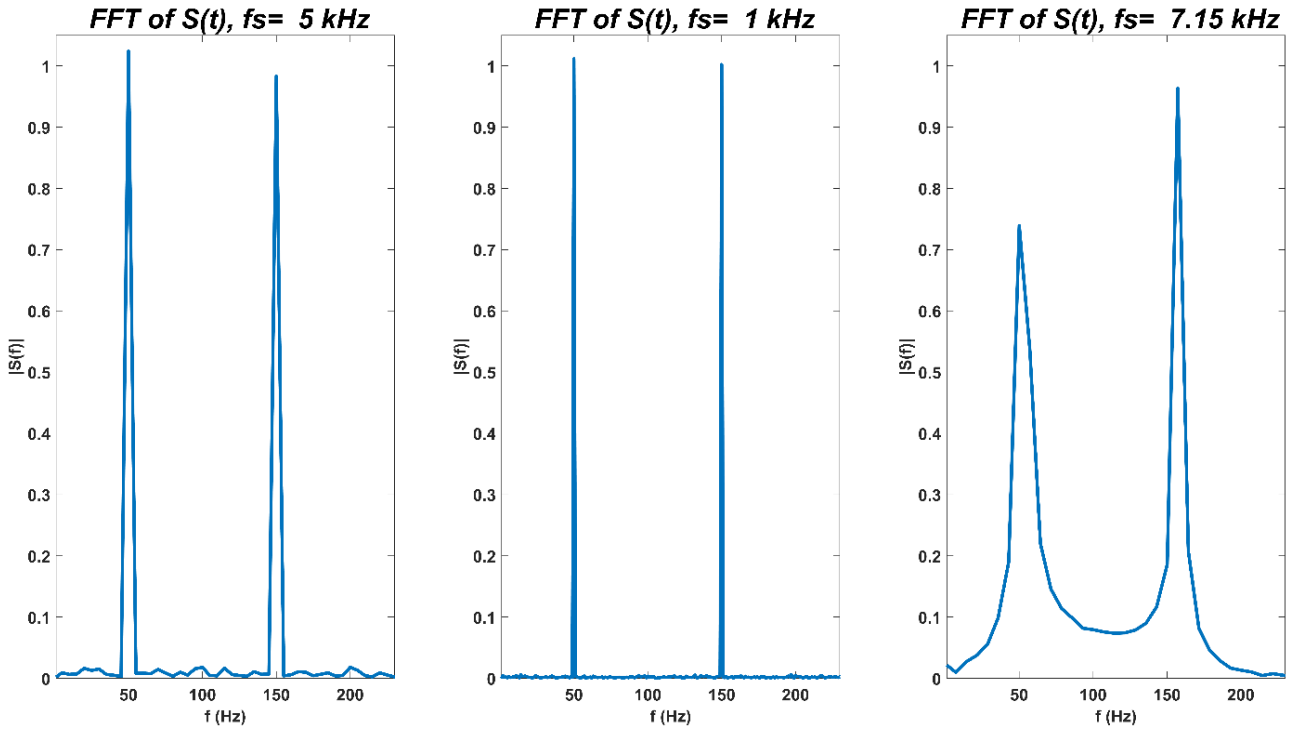


Figure 12. Effects of different sampling frequencies for experimental dataset

Table 8. Results for different sampling frequencies for experimental dataset

	Main Component			*Inserted Component*		
Sampling Frequency, f_s	5000	1000	7150	5000	1000	7150
Frequency Resolution (Hz)	5	1	7.15	5	1	0.696
Nominal Amplitude (V_{pu})	1	1	1	1	1	1
Nominal Frequency (Hz)	50	50	50	150	150	150
Measured Amplitude (V_{pu})	1.024	1.012	0.739	0.983	1.002	0.9636
Error (%)	**+2.4**	**+1.17**	**-26.1**	**-1.68**	**+0.22**	**-3.64**
Measured Frequency (Hz)	50	50	50.05	150	150	157.3
Error (%)	**0**	**0**	**+0.1**	**0**	**0**	**+3.65**

4.2.2. Length

Figure 13 gives the effects of the different signal lengths for experimental data. The appropriate length selection is important for accurate measurement as shown in this figure. The results of the comparison are available in Table 9. Both components are measured with a low error rate under the case of $L = 1000$ and 15000, while important errors occur in terms of frequency and amplitude under the case of $L = 14371$. Because the frequency resolution is 0.6958 that leads to a 0.202 % frequency deviation for both components. This result is related directly to the measurement window.

Table 9. Results for different signal lengths for the experimental dataset

	Main Component			*Inserted Component*		
Length of Signal, L	1000	15000	14371	1000	15000	14371
Frequency Resolution (Hz)	5	0.6667	0.696	5	0.667	0.696
Nominal Amplitude (V_{pu})	1	1	1	1	1	1
Nominal Frequency (Hz)	50	50	50	150	150	150
Measured Amplitude (V_{pu})	1.0240	1.005	0.97	0.983	1.0038	0.72
Error (%)	**+2.4**	**+0.5**	**-3**	**-1.68**	**+0.38**	**-28**
Measured Frequency (Hz)	50	50	50.101	150	150	150.303
Error (%)	**0**	**0**	**+0.202**	**0**	**0**	**+0.202**

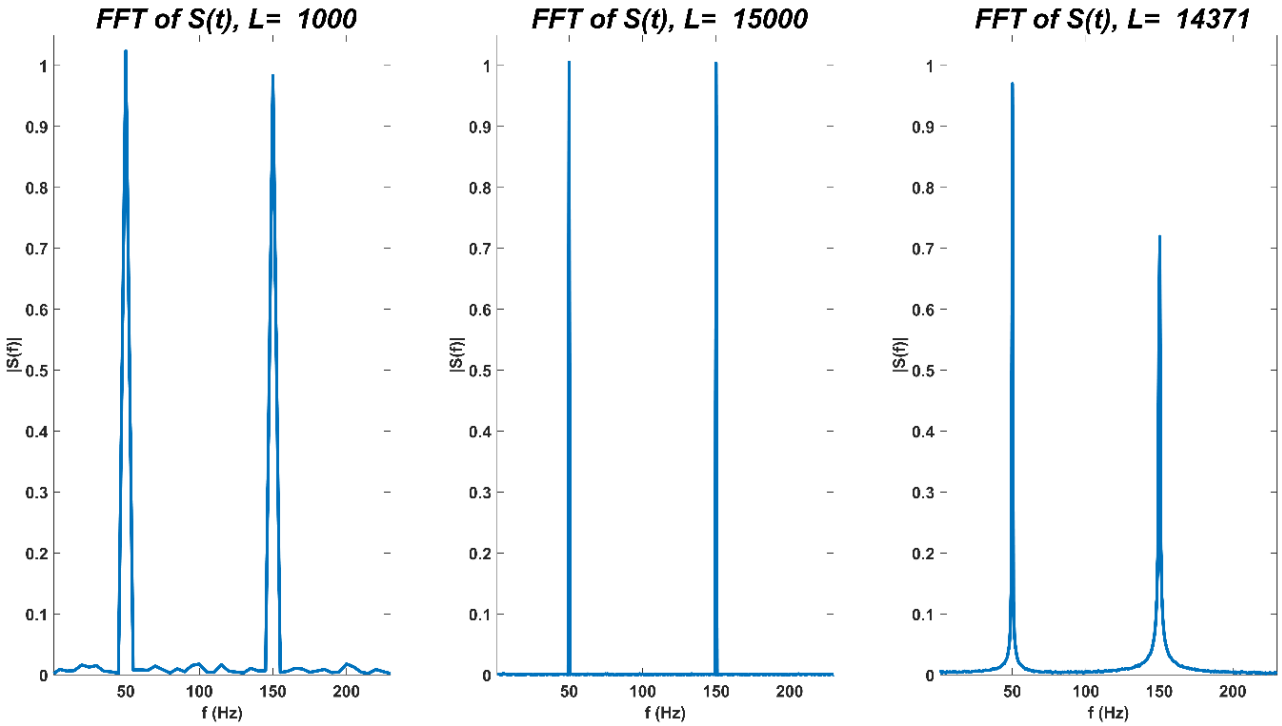


Figure 13. Effects of different main frequencies for experimental dataset

4.2.3. Main Frequency

Figure 14 illustrates the effects of the main frequency value in the case of experimental data. An appropriate main frequency value is also important for accurate measurement, as shown in this figure. The results of this effect are presented in Table 10. Both components are measured with a low error rate in the case of $f_{main} = 50\ Hz$. In contrast, important errors occur in terms of frequency and amplitude under the cases of $f_{main} = 49.5$ and $50.5\ Hz$ for the main component. The inserted component is measured accurately in all cases. Because the frequency value of the main component affects itself independently from the other components.

Table 10. Results for different main frequencies for experimental data

	Main Component			***Inserted Component***		
Main Frequency, f_0 (Hz)	49.5	50	50.5	150	150	150
Frequency Resolution (Hz)	5	5	5	5	5	5
Nominal Amplitude (V_{pu})	1	1	1	1	1	1
Nominal Frequency (Hz)	50	50	50	150	150	150
Measured Amplitude (V_{pu})	1.002	1.0240	0.99	1.005	0.9832	1.007
Error (%)	**+0.2**	**+2.4**	**+1**	**0.5**	**-1.68**	**+0.7**
Measured Frequency (Hz)	50	50	50	150	150	150
Error (%)	**+1.0101**	**1**	**-0.990**	**0**	**0**	**0**

4.2.4. Inserted Component Frequency

Figure 15 reveals the effects of the inserted component frequency value for experimental data. An appropriate inserted component frequency value is also important for accurate measurement, as shown in this figure. The results of this effect are presented in Table 11. Both components are measured with a low error rate in the case of $f_{inserted} = 100\ Hz$, and 150 Hz, while important errors occur in terms of frequency and amplitude under the cases of $f_{inserted} = 149\ Hz$ for just the inserted component. The main component is measured accurately in all cases. Similar to that in the cases of main frequency, the value of the inserted component frequency only affects itself independently from the other components.

Table 11. Results for different inserted component frequencies for experimental data

	Main Component			***Inserted Component***		
Inserted Component Frequency, $f_{inserted}$ (Hz)	50	50	50	150	149	100
Frequency Resolution (Hz)	5	5	5	5	5	5
Nominal Amplitude (V_{pu})	1	1	1	1	1	1
Nominal Frequency (Hz)	50	50	50	150	150	150
Measured Amplitude (V_{pu})	1.024	1.028	1.016	0.983	0.951	1.0021
Error (%)	**+2.4**	**+2.81**	**+1.61**	**-1.68**	**-4.1**	**+0.21**
Measured Frequency (Hz)	50	50	50	150	150	150
Error (%)	**0**	**0**	**0**	**0**	**+1.0067**	**0**

4.3. Comparison of the proposed method

The proposed study on harmonic- / interharmonic-signal generation and analysis with different parameters comprehensively on electrical systems has been compared with 11 different studies in Table 12 is established with It outperforms in this field in terms of parameters like interharmonic-, length effect, frequency deviation effect, and noise effect. Furthermore, all parameters are assured by the proposed method.

All results show that values of the parameters, such as sampling frequency, signal length, amplitudes and frequencies of main and inserted components, and SNR level, are effective for the correct measurement. These are presented for the user predefined and comparatively in this app to understand the effects. If one would like to generate and analyze a user-defined signal with additional components, he/she can use the app's user-defined input interface.

This study offers users different choices, such as training, generating, and analyzing a signal with different interfaces of the proposed app.

The proposed method's limitation is that only one harmonic or interharmonic can be inserted into the fundamental frequency signal at any noise level of AWGN instead of more than one related component.

Table 12. Comparison of proposed and similar studies in the literature

Methods	Fields	FFT	Harmonics	Interharmonics	Length Effect	Frequency Deviation Effect	Noise Effect	Generation	Analysis	Examples	Programming Language
[14]	Optics	✔ Fourier Theory	✔	✖	✖	✖	✖	✔	✔	✔	Java
[16]	Physics	✖	✔	✖	✖	✖	✖	✖	✖	✖	Android
[17]	Music	✖	✔	✖	✖	✖	✖	✖	✔	✖	Java
[18]	Physics	✖	✖	✖	✖	✖	✖	✔	✔	✖	Android
[19]	Signals and Systems	✔	✔	✖	✖	✖	✖	✔	✔	✔	MATLAB
[20]	Sounds	✔ Fourier Series	✔	✖	✖	✖	✖	✔	✔	✖	✖
[21]	Sounds	✔	✖		✖	✖	✖	✖	✔	✖	MATLAB
[22]	Power Electronics	✔	✔	✖	✖	✖	✖	✔	✔	✖	Labview
[23]	Electrical and Electronics Engineering	✔ Fourier Series	✔	✖	✖	✖	✖	✔	✔	✖	Android
[24]	Electrical Systems	✔	✔	✖	✖	✖	✖	✔	✔	✖	ISE
[25]	Electrical Systems	✔ Fourier Series	✔	✖	✖	✖	✖	✖	✔	✖	MATLAB
Proposed Method	Electrical Systems	✔	✔	✔	✔	✔	✔	✔	✔	✔	MATLAB

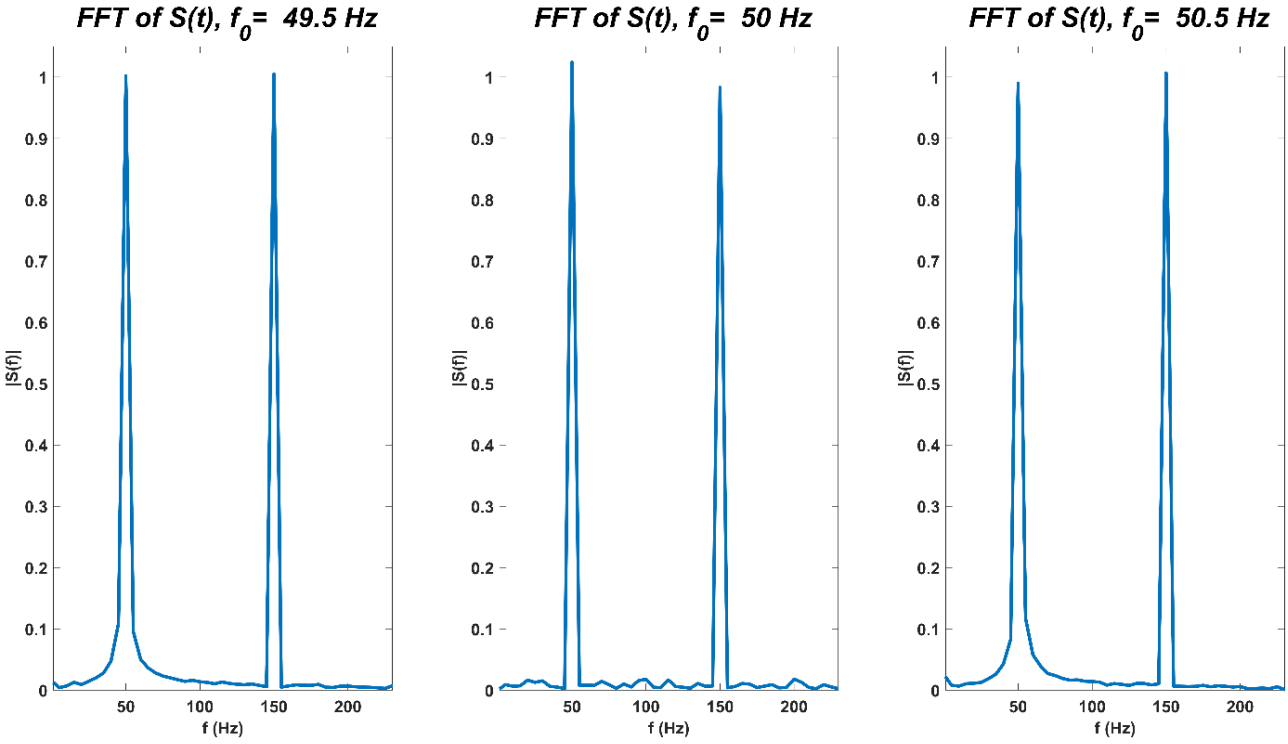


Figure 14. Effects of different lengths for experimental dataset

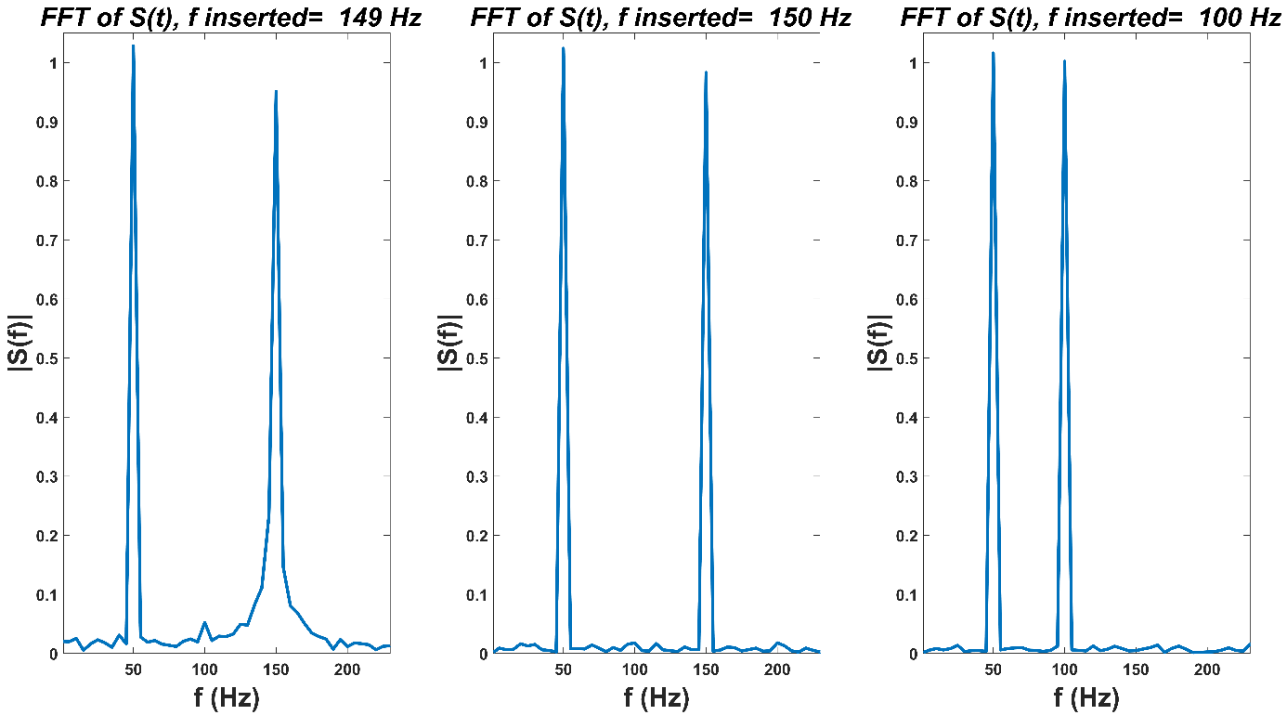


Figure 15. Effects of different inserted component frequencies for experimental dataset

5. Conclusion and Suggestions

The conclusion section should be stand-alone. The aim of the study and its significant results should be given briefly in a concrete way. In addition, suggestions and opinions that are requested to be conveyed to the readers regarding the results of the study can be stated.

In summary, this paper presents an enhanced comparative investigation concerning the FFT based on related parameters, encompassing the analysis of power signals containing disturbance components

such as harmonics / interharmonics, and noise with not only synthetical signals but also experimental signals. The findings of the FFT analysis, involving parameters including α_{main}, $\alpha_{inserted,}$ f_{main}, $f_{inserted}$, f_s, L, and SNR of noise, reveal that alterations in the mentioned specific parameters exert a localized impact solely on individual component values. In contrast, variations in the remaining FFT parameters affect the entirety of the measured values across all processed components. Moreover, it was found that small changes in the sampling frequency of 1/2000, the fundamental frequency of 0.5 Hz, or the measurement window of 1/1000 caused very large errors. Furthermore, the paper introduces a rudimentary application facilitating the comparison of these parameters and examining resultant outcomes. This research contributes to advancing the understanding of disturbance components and noise within signal-processing contexts. Additionally, it highlights important parameters that affect the analysis and effectiveness of FFT-based study across different parameter circumstances. Furthermore, users can easily understand the effects of the FFT parameters employing subinterfaces of predetermined comparative parameters based on synthetic and experimental data. This study offers different choices to users, such as training, generating, and analyzing a signal with different interfaces of the proposed app. Although this application has been designed for power system signals, it also applies to biomedical, audio, and other electrical signal studies.

Acknowledgment

The author would like to thank the Sivas University of Science and Technology, Department of Electrical Electronics Engineering, Smart Grids Laboratory.

Statement of Research and Publication Ethics

The study complies with research and publication ethics.

References

[1] “General guide on harmonics and interharmonics measurements and measuring instruments for power supply networks and attached devices used for the measurements”, *IEC Standard 61000-4-7*, 2008.

[2] “Testing and measurement techniques - Power quality measurement methods”, *IEC Standard 61000-4-30*, 2003.

[3] "IEEE Standard Definitions for the Measurement of Electric Power Quantities Under Sinusoidal, Nonsinusoidal, Balanced, or Unbalanced Conditions," *IEEE Std 1459-2010*, pp.1-50, 19 March 2010, doi: 10.1109/IEEESTD.2010.5439063

[4] "IEEE Recommended Practice for Monitoring Electric Power Quality, "*IEEE Std 1159-2019*, pp.1-98, 13 Aug. 2019, doi: 10.1109/IEEESTD.2019.8796486

[5] S. Akkaya and Ö. Salor, “Flicker Detection Algorithm Based on the Whole Voltage Frequency Spectrum for New Generation Lamps – Enhanced VPD Flickermeter Model and Flicker Curve,” *Electric Power Components and Systems*, vol. 0, no. 0, pp. 1–15, 2022, doi: 10.1080/15325008.2021.2011487.

[6] S. Akkaya and Ö. Salor, “New flickermeter sensitive to high-frequency interharmonics and robust to fundamental frequency deviations of the power system,” *IET Science, Measurement and Technology*, vol. 13, no. 6, 2019, doi: 10.1049/iet-smt.2018.5338.

[7] S. Akkaya and Ö. Salor, “A new flicker detection method for new generation lamps both robust to fundamental frequency deviation and based on the whole voltage frequency spectrum,” *Electronics (Switzerland)*, vol. 7, no. 6, 2018, doi: 10.3390/electronics7060099.

[8] S. Akkaya and Ö. S. Durna, “Enhanced spectral decomposition method for light flicker evaluation of incandescent lamps caused by electric arc furnaces,” *Journal of the Faculty of Engineering and Architecture of Gazi University*, vol. 2018, no. 18–2, pp. 987–1005, 2018, doi: 10.17341/gazimmfd.460497.

[9] S. Akkaya, "A Review of the Experimental Studies on Analysis of Power Quality Disturbances," in *Pioneer and Contemporary Studies in Engineering*, 2023, pp. 453–477.

[10] S. Akkaya, "An Overview of the Empirical Investigations into the Classification of Power Quality Disturbances," in *Pioneer and Contemporary Studies in Engineering*, 2023, pp. 409–430.

[11] S. Akkaya, "A Conspectus of PQD Analysis," in *ICAENS*, 2023, pp. 325–329. [Online]. Available: https://as-proceeding.com/index.php/icaens/article/view/1015/950 . [Accessed: Dec. 11, 2004].

[12] S. Akkaya, "Empirical Investigations: Power Quality Disturbance Classification," in *ICAENS*, 2023, pp. 320–324. [Online]. Available: https://as-proceeding.com/index.php/icaens/article/view/1014/949 . [Accessed: Dec. 11, 2004].

[13] S. Akkaya, E. Yüksek, and H. M. Akgün, "A New Comparative Approach Based on Features of Subcomponents and Machine Learning Algorithms to Detect and Classify Power Quality Disturbances," *Electric Power Components and Systems*, 2023, doi: 10.1080/15325008.2023.2260375.

[14] K. H. Cheong and J. M. Koh, "Integrated virtual laboratory in engineering mathematics education: Fourier theory," *IEEE Access*, vol. 6, pp. 58231–58243, 2018, doi: 10.1109/ACCESS.2018.2873815.

[15] Y. Y. Zhuang, Y. H. Lin, M. Liyanawatta, A. H. Saputro, Y. D. Utami, and J. H. Wang, "An interactive programming learning environment supporting paper computing and immediate evaluation for making thinking visible and traceable," *Interactive Learning Environments*, 2023, doi: 10.1080/10494820.2023.2212709.

[16] B. Pambayun, J. V. D. Wirjawan, H. Herwinarso, A. Wijaya, B. Untung, and E. Pratidhina, "Designing Mobile Learning App to Help High School Students to Learn Simple Harmonic Motion," *International Journal on Social and Education Sciences*, vol. 1, no. 1, 2019.

[17] D. Malandrino, D. Pirozzi, and R. Zaccagnino, "Learning the harmonic analysis: is visualization an effective approach?," *Multimed Tools Appl*, vol. 78, no. 23, pp. 32967–32998, Dec. 2019, doi: 10.1007/s11042-019-07879-5.

[18] D. Buongiorno and M. Michelini, "Experimental use of mobile apps in physics education," *AAPP Atti della Accademia Peloritana dei Pericolanti, Classe di Scienze Fisiche, Matematiche e Naturali*, vol. 99, 2021, doi: 10.1478/AAPP.99S1A22.

[19] F. Vatansever and N. A. Yalcin, "e-Signals&Systems: A web-based educational tool for signals and systems," *Computer Applications in Engineering Education*, vol. 25, no. 4, pp. 625–641, Jul. 2017, doi: 10.1002/cae.21826.

[20] C. S. Wang, "Teaching Fourier series with tone experiments based on smartphone applications," *Computer Applications in Engineering Education*, vol. 31, no. 5, pp. 1358–1371, Sep. 2023, doi: 10.1002/cae.22644.

[21] P. C. P. S. Andreas Spanias, "A New Signal Processing Course for Digital Culture," in *Frontiers in Education 2015 : launching a new vision in engineering education*, IEEE, 2015.

[22] A. Marquez, J. I. Leon, L. G. Franquelo, and S. Vazquez, "Educational Hardware/Software Interface for Power Electronic Applications," in *6th IEEE International Conference on e-Learning in Industrial Electronics*, IEEE, 2012.

[23] M. J. C. S. Reis, S. Soares, S. Cardeal, R. Morais, E. Peres, and P. J. S. G. Ferreira, "FouSE: An android tool to help in the teaching of fourier series expansions in undergraduate education," in *CSEDU 2013 - Proceedings of the 5th International Conference on Computer Supported Education*, 2013, pp. 166–171. doi: 10.5220/0004401101660171.

[24] B. Erişti, Ö. Yıldırım, H. Erişti, and Y. Demir, "An FPGA-based System for Real-time Monitoring of Voltage Harmonics," in *19th IMEKO TC 4 Symposium and 17th IWADC Workshop Advances in Instrumentation and Sensors Interoperability*, 2013, pp. 677–682.

[25] Z. Khan, M. K. Karim, and M. M. Ashraf, "Design Of A Prototype Model For Harmonics Estimation Of Real-Time Current/Voltage Waveforms By Using MATLAB App Designer At Laboratory Level," in *MDSRC 2021*, 2021, pp. 1–9.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 939-949 YEAR: 2024
DOI: 10.17798/bitlisfen.1472913

Determination of Biomass Potential of Muş Province

Okan AKTAN[1], Ömer ARSLAN[2*]

[1]*KOSGEB Muş Provincial Directorate, Muş, Türkiye*
[2]*Muş Alparslan University, Muş Faculty of Engineering and Architecture, Mechanical Engineering Department, Türkiye*
(ORCID: 0000-0002-0704-0182) (ORCID: 0000-0003-4190-5271)

Keywords: Renewable Energy, Muş, Biomass, Animal Waste, Vegetable Waste, Domestic Waste

Abstract

Fossil resources currently supply the majority of the world's energy needs. Even if fossil fuel sources are gradually depleted, the need for energy will continue due to population increases and the widespread use of life-improving technology. Biomass is one of these unique and sustainable energy sources. Anywhere there is life- human, animal, or plant-biomass may be a clean, readily available, and sustainable energy source. Furthermore, Türkiye, which depends on foreign energy and should reduce its carbon footprint, must benefit from this energy source to the maximum extent. Since the city of Muş has an economy based on agriculture and animal husbandry, it is a province with high biomass potential. Therefore, it was chosen as the subject of this study. Furthermore, with this study, the biomass potential of Muş province was tried to be determined for the first time in the literature. The theoretical annual amount of energy obtained from plant, animal, forest, and municipal waste in Muş province is 222,066.3 TOE. However, the economically realizable amount of the theoretical potential was calculated to be 8935.20 TOE. It has been estimated that this economical biomass energy can meet 27.15% of the city's electricity consumption. This contributes significantly to the socio-economically underdeveloped city economy and provides the utilization of waste that would harm the city's environment.

1. Introduction

Humans rely on both underground and surface energy resources to meet their energy needs. Among these sources, fossil fuels are the most widely consumed. However, due to the limited and finite nature of fossil fuel reserves and the environmental issues they cause, many countries are turning to renewable energy sources [1-2]. This shift has sparked significant interest among government officials, academic researchers, and industry experts who aim to enhance energy supply. A primary goal of renewable energy production is to reduce anthropogenic greenhouse gas emissions contributing to climate change while promoting social and economic development. It includes improving energy accessibility, enhancing energy security, and achieving better public health outcomes. Biomass, a versatile renewable energy source, plays a crucial role in reducing fossil fuel emissions, particularly in sectors where it is challenging to cut carbon emissions. Solid biomass is any plant matter that can be used directly as fuel or processed into other forms before combustion [3]. The diversity of biomass sources for energy generation such as forest management practices, cultivated biomass crops, animal waste, agricultural residues, and biomass from industrial, municipal, and agricultural solid waste highlights the potential of this resource [4].

Like many fossils fuel-poor countries, Türkiye has started to work hard to meet most of its energy needs from local biomass and other renewable energies. This trend is increasing since biomass energy can be used as thermal energy or converted into electrical energy and is environmentally friendly. About six-percent of the electrical energy produced in

*Corresponding author: o.arslan@alparslan.edu.tr
Received: 25.04.2024, Accepted: 23.12.2024

the world is produced from renewable energy sources. Biomass energy is the most widely used renewable energy type after hydroelectric and wind energy in the production of electrical energy. Especially in OECD and European countries, there has been a significant increase in the evaluation of biomass energy and other renewable energy sources due to the dramatic increase in incentives and grants to encourage using renewable energy sources [5]. In the report titled "Fuels of 2050," announced in 2011 by the European Union, detailed information about the use of different biomass fuels in sea, land, and air vehicles is given. In the Vision 2030 document, the target of a 25% increase in the use of biomass energy until 2030 is stated [6]. Biogas, which is one of the biomass types, is an alternative fuel to natural gas because it is used not only for electricity generation but also for heating purposes. Sweden plans to use biogas as an alternative to natural gas. In Sweden, the use rate of biogas as a fuel in vehicles in the country has increased in the last ten years. In addition, biogas energy has been used in trains in Sweden since 2005 [7].

Muş has been particularly preferred for biomass potential analysis in this study since it is predicted that the city's biomass potential may be attractive for investment due to the agriculture and animal husbandry-based economy of Muş province. Within the scope of this study, the biomass potential of Muş was determined by considering the amount of municipal and domestic waste as well as animal, plant, and forest wastes of Muş. Then, a comparison was made with the values in the literature. In addition, in this study, monthly electricity consumption, natural gas consumption, and petroleum products consumption in Muş province were calculated. Thus, it has been tried to determine how much biomass energy, an alternative to fossil fuels, can meet the energy consumed by the city. As a province of Türkiye, which is foreign-dependent in energy, the contribution that Muş can offer to the country by making maximum use of the biomass energy potential has been tried to be revealed. It is also hoped that this study can serve as a reference guide that can contribute to academic studies and business people who want to invest in renewable energy.

2. Material and Method

The biomass potential of a region is determined by the amount of vegetal, animal, domestic, municipal, and forestry waste belonging to that region. The biomass potential created by the vegetal entity is calculated by taking into account the waste rate of the plants and the unit calorific values of these plants. When looking at animal-derived waste, calculations are made according to the type of animal, the amount of waste it will create, and the amount of energy that can be obtained from these wastes. The amount of domestic and municipal waste is proportional to the size of the city's human population. In forestry assets, the amount of energy obtained from tree species and wastes such as bark and leaves arising from trees is calculated. In this study, the BEPA calculation system was used. Türkiye Biomass Energy Potential Atlas (BEPA) is a GIS application that can dynamically present on a map, graphically and numerically, the potential to produce how much electricity and how much biofuel from which biomass source in Türkiye and in which regions of the country these resources are concentrated. In this study, data regarding the vegetal, animal, forest, and domestic waste potential of Muş province are given in tables to help understand the subject. In addition, all formulas used in BEPA biomass energy potential calculations are given in detail, and an attempt is made to understand the method used in the calculations. In this study, calculations were taken from the BEPA system. In the BEPA system, calculations are updated according to the latest data.

3. Biomass Potential of Muş Province

Muş is a province located in the Eastern Anatolia Region. Its area is 8,196 km^2, corresponding to 1.1% of Türkiye's area. It is located between 39 29' and 38 29' north latitudes and 41 06' and 41 47' east longitudes. Its population is 411,117. It has the Muş Plain, Türkiye's third largest inland plain. It covers an area of approximately 1650 km^2, 30 kilometers wide and 80 kilometers long. Muş province borders the Ahlat and Adilcevaz districts of Bitlis and the Patnos and Tutak districts of Ağrı from the east. It borders Erzurum's Karayazı, Hınıs, Tekman, and Karaçoban districts from the north, Bingöl's Karlıova and Solhan districts, and Diyarbakır's Kulp districts from the west. It borders Batman's Sason and Bitlis' Güroymak and Mutki districts from the south. Muş province was established on the northern slope of Kurtik Mountain [8].

Muş province is an agricultural and animal husbandry city. It includes the Muş Plain, Bulanık, Liz, and Malazgirt plains within the borders of Muş province and is a city with a high animal population as it has large-scale pasture areas. Therefore, the city's agriculture and livestock potential are high. This section will discuss the city's plant, animal, forest assets, and domestic and municipal waste potential. Not all plant and animal wastes can be used to obtain biomass energy. Only a certain percentage of this waste is used.

This section will discuss the city's plant, animal, forest assets, and domestic and municipal waste potential.

3.1. Animal Biomass Potential of Muş Province

Eastern Anatolia is one of the regions with high livestock breeding potential in Türkiye, and Muş is one of the provinces with high livestock breeding potential in this region. Since the amount of animal waste increases directly proportional to the number of animals, biomass energy potential also increases. Due to the advantage of having large pasture lands in Muş province, farmers mainly engage in pasture livestock farming.

The number of sheep, goats, and cattle in Muş province is over 2 million. Data on animal numbers are shared in Table 1. Due to the large capacity of pasture lands in Muş province, sheep and goats are widely bred.

Table 1. Animal population of Muş Province between 2018-2022 [9]-[10].

Animal species	2018	2019	2020	2021	2022
Small ruminant	1,041,102	1.084.591	1,235,552	1,250,000	1.205.000
Cattle	306,542	325,067	331,881	335,798	301.000
Buffalo	7,341	7,078	7,031	7,591	-
Horse	2,629	2,619	2,430	2,171	-
Donkey	2,601	2,625	2,327	2,091	-
Mule	342	348	347	311	-
Poultry (total)	511,128	503,169	732,592	467,007	346.773
Chicken	305,462	300,908	422,591	277,047	-
Turkey	76,484	72,615	110,227	64,234	-
Duck	36,428	35,610	43,472	23,268	-
Goose	92,754	94,036	156,302	102,458	-

The animal population of Muş Province as of the end of 2022 is illustrated in Figure 1.

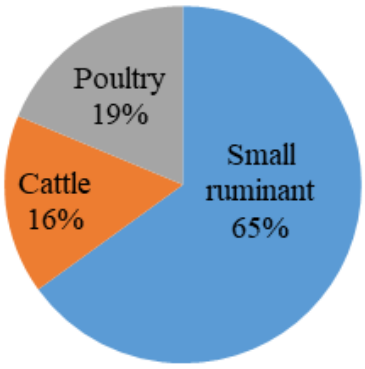


Figure 1. Animal population of Muş Province at the end of 2022 [10].

Table 2 analyzes the changes in the animal population within Muş Province from 2010 to 2023, with corresponding data illustrated in Figure 2. Figure 2 depicts variations in the animal population, demonstrating fluctuations rather than a consistent trend of either growth or decline.

Table 2. Animal population change in Muş Province between 2010-2023 [10].

Years	Cattle	Small ruminant
2010	236,330	1,004,831
2011	247,302	1,016,089
2012	276,507	1,001,228
2013	294,998	925,551
2014	285,146	1,085,197
2015	290,521	984,070
2016	302,215	1,021,142
2017	306,508	1,049,367
2018	306,542	1,041,102
2019	325,067	1,084,591
2020	331,881	1,235,552
2021	335,798	1,250,000
2022	301,000	1,205,000
2023	235,507	1,051,461

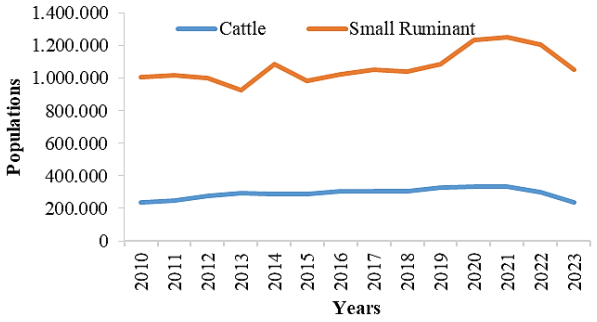


Figure 2. Changes in cattle and small ruminant populations between 2010-2023 years in Muş Province

When calculating the amount of animal waste, the waste yield for cattle is taken as 10-20 kilograms per day. However, the daily waste amount of goats and sheep is accepted as 2 kg. As can be seen in Table 3, animal biogas potential was calculated by taking into account animal species, waste collection rate, and biogas yield [8].

Table 3. Animal species and biomass (biogas) energy yield [11].

Animal species	Live weight (kg)	Amount of fresh waste		TS(%)	VS(%)	Availability duration of stay in the stable (%)	Biogas yield (L/kg.UK)
		Percentage of weight (%)	Manure (kg/day)				
Cattle	135-800	5-6	10-20	5-25	75-85	65 (milk), 25(meat)	200-350
Small ruminants	30-75	4-5	2	30	20	13	100-310
Chicken(egg) Chicken(meat)	1.5-2.0	3-4	0.08-1.00	10-35 50-90	70-75 60-80	99	310-620 550-650

Formulas (1) and (2) below were used to calculate animal waste.

$$\sum AM_{yıllık} = AMx\ HS \quad (1)$$

$$\sum Q_{top.yanma} = \sum AM_{yıllık}\ x\ Hu \quad (2)$$

Here;

$Q_{top.yanma}$ = Energy obtained from total combustion

$AM_{yıllık}$ = Annual waste amount

HS = Number of Animals

Hu = Lower calorific value of animal waste

3.2. Crop Production Potential of Muş Province

As seen in Table 4, Muş province has a land area of 866,833 hectares. The proportion of agricultural lands among these lands is 357,342 ha. 278,500 ha of agricultural land is suitable for the use of agricultural machinery. The non-arable area of Muş province is 21,331 ha [9].

Table 4. Land types and ratios in Muş Province [12].

Land type	Area (ha)	Percentage in land (%)
Farmland	357,342	41.2
Meadow	56,985	6.6
Pasture	351,176	40.5
Forest	79,999	9.2
Unsuitable for agriculture	21,331	2.5
Total	866,833	100

As seen in Table 5, grain and plant products rank first in the use of agricultural lands in Muş province, with an area of 2,095,225 decares. Fruit beverages, spices, and vegetable plants follow these. The table shows that there is not ornamental plant cultivation in Muş province. Additionally, the land left fallow is 281,530 decares

Table 5. Usage status of agricultural lands in Muş Province [13]

Agricultural areas (decare)	
Fruits, beverage and spice plants	13,415
Fallow	281,530
Vegetables	41,088
Ornamental plants	0
Cereals and other plant products	2,095,225

Table 6 shows the crop production data of Muş province, covering the years 2020-2022. Wheat is the most significant grain production in Muş. The wheat production area corresponds to 1,115,402 decares. Then, alfalfa, barley, sugar beet, sainfoin, vetch, silage corn, safflower seeds, chickpeas, dry beans, oil sunflower seeds, tobacco, and potatoes are produced, respectively. Besides grains and other plant products, mostly vegetables are produced in Muş. We can list these vegetables, from highest production to lowest, as watermelon, table tomato, melon, cucumber, white cabbage, green beans, bell pepper,

green pepper, tomato paste, onion, eggplant, onion, and zucchini. When the fruits, beverages, and spice plants produced in Muş province are examined, table grapes take the first place with an area of 3,848 decares. These are followed by walnuts, apples, pears, and cherries [11].

Table 6. Crop production amounts in Muş Province [14].

	2020		2021		2022	
Products	**Cultivation area(ha)**	**Production (ton)**	**Cultivation area(ha)**	**Production (ton)**	**Cultivation area(ha)**	**Production (ton)**
Clover	49,427	1,234,224	49,599	1,077,623	46,590	965,447
Sugar beet	6,603	372,670	5,581	277,558	7,313	406,612
Wheat (other)	111,540	222,970	124,254	199,798	125,000	431,061
Sainfoin-(green herb)	5,438	106,200	82,964	82,964	3,695	49,695
Corn-slage	2,315	114,190	2,008	96,994	2,299	125,685
Vetch (green grass)	3,578	60,634	3,452	50,107	2,877	36,467
Barley	24,710	56,107	26,923	33,069	34,897	97,706
Sunflower(oil)	777	1,712	2,936	4,085	3,517	7,934
Haricot bean	1,236	2,087	1,521	2,575	1,995	3,812
Chickpeas	1,797	2,633	1,736	1,361	14	1,742
Potatoes	129	3,904	145	4,087	33	967
Red lentil	7	10	7	6	76	114
Corn-grain	54	308	145	961	162	1,942
Onion	61	513	64	996	23	582

The biomass energy potential obtained from agricultural wastes was calculated using the formula (3). Thus, the total biomass potential that can be obtained from the plant products of Muş province has been revealed.

$$\sum Q_{top.yanma} = \sum AM_{yıllık}\ x\ Hu\ x\ AK\ x\ KO. \quad (3)$$

Here;

$Q_{top.yanma}$ = Energy obtained from total combustion

$AM_{yıllık}$ = Annual waste amount

AK = Literature waste coefficient

KO = Availability rate

Hu = Lower calorific values of agricultural wastes

While calculating the biomass potential from plant products, the usability and unit heat values of the products shown in Table 7 were considered.

Table 7. Unit calorific values of some plant products [11]

Products	Wastes	Utilization (%)	Unit calorific value (MJ/kg)
Wheat	Straw	15	17.9
Barley	Straw	15	17.5
Rye	Straw	15	17.5
Oat	Straw	15	17.4
Corn	Stalk	60	18.5
	Somek	60	18.4
Rice	Straw	60	16.7
	Shell	80	12.98
Tabacco	Stalk	60	16.1
Cotton	Stalk	60	18.2
	Crispy waste	80	15.65
Sunflower	Straw	60	14.2
Peanut	Straw	80	20.74
	Shell	80	20.74
Soya	Straw	60	19.4

3.3. Municipal Waste Biomass Potential of Muş Province

Waste such as leaves from parks and gardens, grass, sewage waste, waste from municipalities, garbage from houses, urban waste, and waste from industrial facilities are evaluated within this scope. While conducting biomass potential analysis, domestic and municipal wastes were included.

The biomass potential of municipal waste was calculated using the formula (4).

$$\sum Q_{top.yanma} = \sum AM_{yıllık}\ x\ H_u\ x\ AO. \quad (4)$$

Here;

$Q_{top.yanma}$ = Energy obtained from total combustion

$AM_{yıllık}$ = Annual waste amount

AO = Proportion of organic waste in municipal waste

H_u = Lower calorific value of municipal wastes

3.4. Waste Biomass Potential of Muş Province

Due to the continental climate structure of the city, biomass potential based on forest products is low. There is a total of 78,426 ha of forest area within the borders of Muş, including 48,775 ha of degraded forest area and 29,651 ha of normal forest area. Fuelwood generated from production activities in the forests in these limited areas can be evaluated as forest waste, such as abandoned roots, thin branches, trees disintegrated during transportation, and dried cones and leaves [15].

The biomass energy potential obtained from forest waste can be calculated using the formula (5).

$$\sum Q_{top.yanma} = AM_{yıllık}\ \text{x}\ H_u\ x\ AO \quad (5)$$

Here;

$Q_{top.yanma}$ = Energy obtained from total combustion

$AM_{yıllık}$ = Annual waste amount

AO = Proportion of forest organic waste

H_u = Lower calorific value of forest wastes

4. Theoretical Biomass Energy Potential of Muş Province

In order to calculate the realistic amount of biomass energy in plant production, it is necessary to collect the waste left over as a result of plant production. Because when full access to the waste generated in plant production is not provided, less biomass energy can be obtained compared to plant production.

As seen in Table 8, the theoretical biomass energy potential of Muş province was 222,067.3 TOE/year. Of this biomass energy potential, 189,412 TOE/year consists of plant wastes 23,125.9 TOE/year animal wastes, 9,175 TOE/year municipal wastes, and 354.4 TOE/year forest product wastes. In addition, Figure 3 graphically shows the proportional comparison of the theoretical biomass energy potential that can be produced from different waste sources in the city. The economically applicable energy potential of this total is calculated as 8,935.20 TEP per year (Table 9) and Figure 4 shows the economically feasible biomass potential from waste at district scale (TOE/year).

Table 8. Theoretical biomass energy potential to be obtained from waste in Muş Province [16].

Vegetable waste (TOE/year)	Animal waste (TOE/year)	Municipal waste (TOE/year)	Forest waste (TOE/year)	Total (TOE/year)
189,412	23,125.90	9,175	354.40	222,067.30

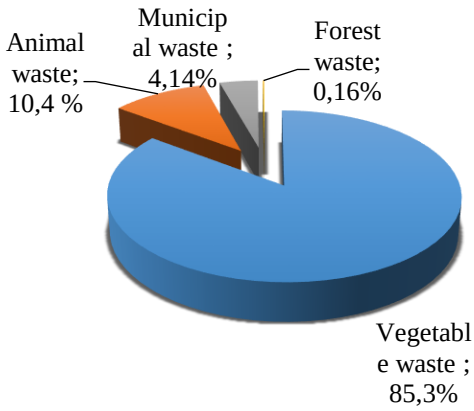


Figure 3. Proportional comparison of theoretical biomass energy potential that can be produced from different waste sources in Muş Province.

Table 9. Economically feasible biogas potential that can be produced from waste at the district scale in Muş Province [16].

Province/District	Type	Animal population (quantity) Amount of crop production (Ton)	Amount of waste (Ton)	Economic energy equivalent (TOE/Year)
Central District	Cattle	65,508	461,803.5	888.2
	Small ruminant	481,750	486,212.9	88.4
	Poultry	156,063	8,604	522.8
	Vegetable waste	1,064,876	140,182	758.1
	Forest waste	291.67		59.07
	Municipal waste	25,412		163.9
Bulanık	Cattle	104,298	684,234	1,235.6
	Small ruminant	143,186	148,284.2	27.5
	Poultry	95,748	4,649.4	282.5
	Vegetable waste	335,225	108,918.3	775.6
	Forest waste	291.67		59.07
	Municipal waste	10,583		68.3
Hasköy	Cattle	24,245	169,703.1	324.7
	Small ruminant	42,420	41,077.1	7.2
	Poultry	37.80	1,969.9	119.7
	Vegetable waste	147,814	4,787.2	2.7
	Forest waste	291.67		59.07
	Municipal waste	3,383.4		78.2
Korkut	Cattle	25,164	172,226.8	324.6
	Small ruminant	109,174	109,621.2	19.9
	Poultry	10,195	489.1	29.7
	Vegetable waste	257,114	30,179.6	20.3
	Forest waste	291.67		59.07
	Municipal waste	3,392.4		78.4
Malazgirt	Cattle	65,230	465,905.8	900.2
	Small ruminant	41,394	43,224.7	8
	Poultry	114,958	5,555.8	337.6
	Vegetable waste	393,973.63	114,088.1	571.9
	Forest wste	291.67		59.07
	Municipal waste	6,743.8		155.8
Varto	Cattle	27,369	179,551.6	324.6
	Small ruminant	223,178	234,462.9	43.9
	Poultry	76,184	4,369.7	265.6
	Vegetable waste	193,334	9,992	62.3
	Forest waste	291.67		59.07
	Municipal waste	4,095.6	461,803.5	94.6
Total				8,935.2

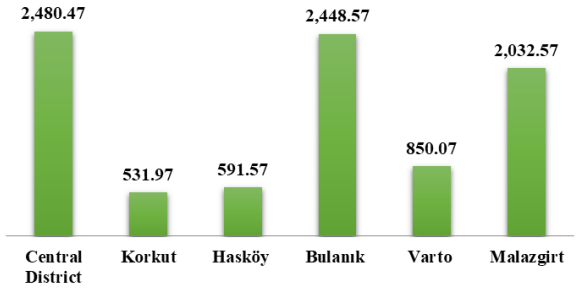


Figure 4. The economically feasible biomass potential from waste at district scale (TOE/year).

4.1 Comparison of the Biomass Potential of the Provinces in the TRB2 Region (Van, Mus, Bitlis, Hakkari)

This section analyzes the economically viable biomass potential derived from waste in the TRB2 region. Among the provinces in the TRB2 region, Muş has the highest biomass production potential, with an economically viable capacity of 8,935.2 tons of oil equivalent (TOE) annually, as illustrated in Table 10 and Figure 5. Following Muş, the provinces of Van, Bitlis, and Hakkari rank sequentially in terms of their respective biomass potentials.

Table 10. The Economically feasible biomass potential from waste in the TRB2 Region [16].

Potential of Van (TOE/year)	Potential of Muş (TOE/year)	Potential of Bitlis (TOE/year)	Potential of Hakkari (TOE/year)
6735.40	8935.20	4512.90	1744.30

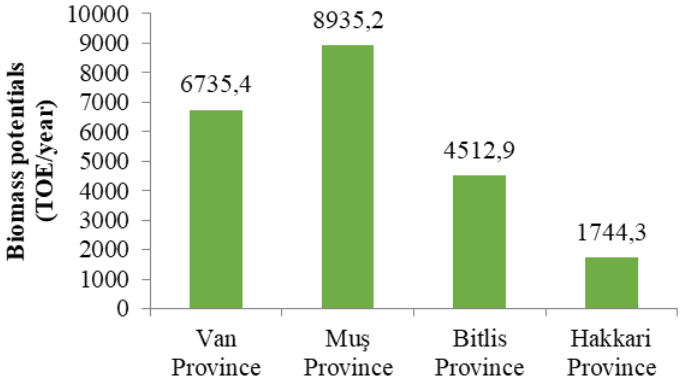


Figure 5. Comparison of the economically feasible biomass potentials (TOE/year) of provinces within the TRB2 Region

5. Energy Consumption Data for Muş Province

As seen in Table 11, the total electrical energy consumption in Muş province in 2022 was 381,772 MWh [17].

Table 11. Distribution of electricity consumption by area in Muş Province [17].

Province	Lighting (MWh)	Public and commercial services and others (MWh)	Residential (MWh)	Industrial (MWh)	Agricultural activities (MWh)	Total (MWh)
Muş	29,841	121,624	161,038	66,086	3,182	381,772

Table 12 shows the natural gas consumption of Muş province. The city's total natural gas consumption in 2022 was 183,672 million sm^3 [18].

Table 12. Natural gas Consumption of Muş Province [18].

Pipe gas	LNG	CNG	Other products	Total
182,433	0,933	0,305	0	183,672

As seen in Table 13, the consumption of petroleum products in Muş province was 56,246.072 tons [19].

Table 13. Petroleum products consumption of Muş Province [19].

Petrol (Ton)	Diesel (Ton)	Fuel oil (Ton)	Aviation fuels (Ton)	Other products (Ton)	Total (Ton)
3,727	48,024	177.94	4,317.146	0	56,246.086

Muş province energy production and consumption data are expressed in different units according to the usage area and sector. Expressing these data in the same unit will provide a clearer understanding of energy consumption and production potential. The unit commonly used for these different energy units is the tonne equivalent unit of oil. According to the definition of the International Energy Agency/Organization for Economic Co-operation and Development (IEA/OECD), it is the amount of energy produced by burning one ton of crude oil. The equivalent of one ton of oil in standard units is 10 Gcal, or 41.868 GJ, or 11.625 MWh. Since crude oil has different origins, different amounts of heat are obtained when burned. Conversion coefficients to TOE, depending on the calorific values of all energy resources in Türkiye, were determined by the regulation on Increasing Efficiency in the Use of Energy Resources and Energy dated October 25, 2008. The approximate equivalent of 1 TOE in other types of energy is given below [20].

1 TOE corresponds approximately to the following values in practice.

- 11600 kWh electrical energy
- 1200 m3 natural gas
- 3 tons of lignite
- 1 ton of fuel oil
- 1.6 tons of hard coal

Energy consumption data for Muş province is converted into tons of oil equivalent (TOE), as shown in Table 14. According to Table 14, 381,772 MWh of electrical energy was consumed annually in Muş province, which corresponds to 32,911.38 TOE. As for natural gas, 183,672 million sm3 was consumed annually, which corresponds to 153,060.00 TOE. Petroleum products were consumed annually at 56,246,090 tons, which corresponds to 56,246,090 TOE. The total energy consumed in the city is 242,217.47 TOE.

Table 14. Annual energy consumption in Muş Province by energy types.

Energy type	Amount/Unit	TOE equivalent
Petroleum	54,246.090/Ton	56,246.090
Natural gas	183,672/ million sm^3	153,060.00
Electricity	381,772 MWh	32,911.38
Total		242,217.47

The energy consumption data of Muş province, given in the table above, is also shown graphically in Figure 6.

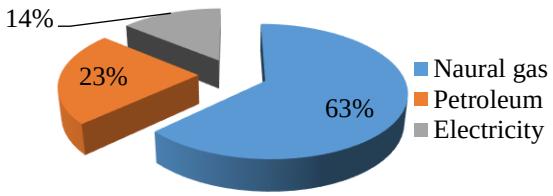


Figure 6. Energy consumption rates in Muş Province by energy types.

The theoretical annual amount of energy obtained from plant, animal, forest, and municipal waste in Muş province is 222,066.3 TOE. The economically feasible amount of this potential is 8935.20 TOE.

6. Conclusion and Evaluation

Since Türkiye is a developing country with a rapidly increasing population, energy demand also increases in parallel with this increase. On the other hand, since it is a country dependent on foreign energy, the burden of energy costs increases on the budget. Therefore, the country needs to make maximum efforts to utilize local energy resources. In addition, the fact that these local energy sources are renewable will greatly contribute to the country's fulfillment of its carbon emission obligations and prevent energy-related environmental pollution on-site. Biomass energy sources, which have a significant potential among renewable energy sources, are increasingly used worldwide and in our country. With this thesis study, the biomass energy potential of Muş province was tried to be revealed.

Muş has the potential for animal, vegetable, forest, and municipal waste that can be used as biomass energy. The city of Muş has a significant population in cattle and sheep farming. While Muş ranks 15th in Türkiye regarding cattle population, it ranks 12th in the country regarding sheep farming [8]. It also promises great potential in terms of agricultural production. Muş Plain, the third largest interior plain of Türkiye, Bulanık, Liz, and Malazgirt plains confirm the existence of this potential. With the implementation of the Alparslan II irrigation project shortly, it will be possible to utilize this potential on a larger scale.

This study calculated the theoretical annual amount of energy obtained from plant, animal, forest, and municipal waste in Muş as 222,066.3 TOE. However, the economically feasible amount was determined as 8,935.20 TOE. The ratios of this economic energy potential to meet the energy consumption of different sectors of the city are shown in Figure 7. As can be seen in Figure 7, it meets

27.15% of the electrical energy needs of Muş province, 5.84% of its natural gas needs, and 15.89% of petroleum fuel products. It covers 3.69% of the city's total energy consumption.

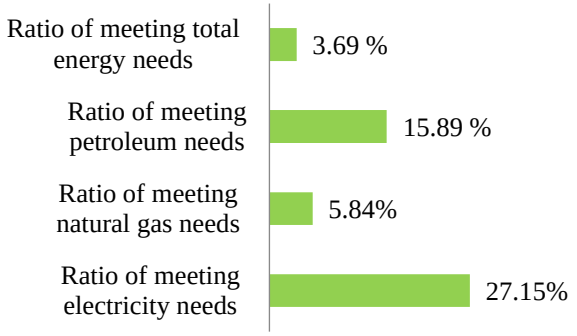


Figure 7. The capacity of Muş's biomass potential to meet the city's energy consumption

There are currently no biogas, bioethanol, or biodiesel facilities in Muş province. Considering the production possibilities and climatic conditions of Muş province, constructing biomass facilities with low investment costs and high efficiency and whose installation is publicly supported will accelerate the establishment of these facilities in the city. By increasing the number of these facilities, a certain amount of Muş's energy needs will be met from local renewable energy sources. Utilizing the biomass potential of Muş province will contribute significantly to both the economic development of the city and the supply of energy needs without harming the environment. Therefore, it is thought that intense efforts are needed to inform and encourage potential entrepreneurs about incentives and grant support for biomass energies. It is hoped that this study will help support this effort.

7. Suggestions for Future Work

The production of biogas from organic waste has the potential to significantly benefit the economy of the rural area of Muş while addressing environmental issues related to waste. However, despite this potential, no biomass facility has been established in Muş province due to a lack of interest and support from the public and relevant stakeholders. With the support of the TRB2 region development agency (Eastern Anatolia Development Agency, DAKA) and the Agricultural and Rural Development Support Institution (TKDK), conducting research, feasibility studies, and pilot projects on a local scale can help increase interest among farmers and the energy sector in this initiative.

Muş province has considerable potential for utilizing animal waste. Currently, animal dung is often burned as a low-calorie fuel or left to decompose in open areas for use as a low-nutrient fertilizer. However, converting animal manure into biogas can produce fuel with significantly higher energy content, along with high-nutrient biofertilizer as a valuable byproduct. This approach not only adds more value but also addresses environmental issues associated with outdoor manure storage. To maximize this potential, it is essential to raise awareness among farmers about the benefits of biogas and to secure support and incentives from relevant authorities. Establishing pilot biogas facilities in select villages or towns, funded by rural development grants, could serve as an exemplary initiative. Such pilot programs could also encourage broader adoption of biogas facilities across the TRB2 region.

Finally, local manufacturing of the equipment and units used in biogas facilities can significantly reduce installation costs. This would promote the widespread use of these facilities and ensure their economic feasibility, thereby boosting confidence in the potential for biogas production in Muş.

Contributions of the authors

Ö. Arslan: Methodology, Formal analysis, visualization, data curation, writing & editing and resources. O. Aktan: Investigation and resources.

Acknowledgement

This article is based on the master's thesis of Okan Aktan, titled "Biomass Potential of Muş Province." The research was conducted under the supervision of Dr. Ömer Arslan at the Mus Alparslan University Institute of Science, within the Department of Nuclear Energy and Energy Systems, in Muş, Türkiye, in the year 2023.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study complies with research and publication ethics.

References

[1] S. Roy, P. K. Dikshit, K. C. Sherpa, A. Singh, S. Jacob, and R. C. Rajak, "Recent nanobiotechnological advancements in lignocellulosic biomass valorization: A review," *Journal of Environmental Management*, vol. 297, p. 113422, 2021.

[2] M. Saleem, "Possibility of utilizing agriculture biomass as a renewable and sustainable future energy source," *Heliyon*, vol. 8, e08905, 2022.

[3] M. A. Helal, N. Anderson, Y. Wei, and M. A. Thompson, "Review of Biomass-to-Bioenergy Supply Chain Research Using Bibliometric Analysis and Visualization," *Energies*, vol. 16, p. 1187, 2023.

[4] W. Liu, J. Wang, T. L. Richard, D. S. Hartley, S. Spatari, and T. A. Volk, "Economic and life cycle assessments of biomass utilization for bioenergy products," *Biofuels. Bioprod. Biorefining*, vol. 11, pp. 633–647, 2017.

[5] C. Karaca, A. Ulutaş, and M. Eşgünoğlu, "Türkiye'de Optimal Yenilenebilir Enerji Kaynağının COPRAS Yöntemiyle Tespiti ve Yenilenebilir Enerji Yatırımlarının İstihdam Artırıcı Etkisi," *Maliye Dergisi*, vol. 172, pp. 111–132, 2017.

[6] Dünya Enerji Konseyi Türk Milli Komitesi, *Enerji Raporu 2012*, 2012.

[7] A. Özen, M. Ü. Şaşmaz, and E. Bahtiyar, "Türkiye'de Yeşil Ekonomi Açısından Yenilenebilir Bir Enerji Kaynağı: Rüzgar Enerjisi," *Karamanoğlu Mehmetbey Üniv. Sos. ve Ekon. Araştırmalar Derg.*, vol. 2015, no. 1, p. 85, 2015.

[8] Ö. Arslan, *Muş İli SWOT Analizi: Kentin Gelişim Potansiyelini Ortaya Çıkarma Bağlamında*. Bursa: Ekin Yayınevi, 2021.

[9] Muş İl Tarım ve Orman Müdürlüğü, *Muş İl Tarım ve Orman Müdürlüğü 2022 Brifingi*, 2022.

[10] TÜİK, "Muş İli Hayvan Varlığı," 2024.

[11] İ. Yokuş, "Sivas İlindeki Hayvansal Atıkların Biyogaz Potansiyeli," M.S. thesis, Tarım Makinaları Anabilim Dalı, Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 2011.

[12] Türkiye İstatistik Kurumu (TÜİK), "Muş İline Ait Arazi Kullanım Çizelgesi," 2020. [Online]. Available: https://www.tuik.gov.tr/. [Accessed: Mar. 29, 2023].

[13] Türkiye İstatistik Kurumu (TÜİK), "Muş İline Ait Tarım Arazilerinin Kullanım Alanları," 2021. [Online]. Available: https://www.tuik.gov.tr/. [Accessed: Mar. 29, 2023].

[14] Muş İl Tarım ve Orman Müdürlüğü, *Muş İl Tarım ve Orman Müdürlüğü 2023 Brifingi*, 2023.

[15] Elazığ Orman Genel Müdürlüğü, *Muş İline Ait Orman Varlığı*, 2023.

[16] BEPA, *Türkiye Biyokütle Enerji Potansiyeli Atlası,* 2023. [Online]. Available: https://bepa.enerji.gov.tr/. [Accessed: Mar. 6, 2023].

[17] EPDK, *Elektrik Piyasası 2022 Yılı Piyasa Gelişim Raporu*, 2023.

[18] EPDK, *Doğal Gaz Piyasası 2022 Yılı Sektör Raporu*, 2023.

[19] EPDK, *Petrol Piyasası 2021 Yılı Sektör Raporu*, 2022.

[20] Yaşar Üniversitesi, *Ton Eşdeğer Petrol (TEP)*, 2015.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 950-968 YEAR: 2024
DOI: 10.17798/bitlisfen.1482747

Q Learning Based PSO Algorithm Application for Inverse Kinematics of 7-DOF Robot Manipulator

Murat Erhan ÇİMEN[1*]

[1]*Sakarya University of Applied Sciences, Faculty of Technology, Department of Electrical and Electronic Engineering, Sakarya, Türkiye*
(ORCID: 0000-0002-1793-485X)

Keywords: Robot kinematics, Reinforcement learning, Q learning, PSO.

Abstract

Solving inverse kinematics problems is one of the fundamental challenges in serial robot manipulators. In this study, a learning-based algorithm was developed to minimize the complexity of solving the inverse kinematics problem for a 7-degree-of-freedom serial manipulator. The parameters of the Particle Swarm Optimization algorithm, modified with Q-learning, a reinforcement learning technique, are updated depending on the states. This approach aimed to increase the efficiency of the algorithm in finding solutions. In the simulation studies, two different end positions of the robot, measured in meters, were used to compare the performance of the proposed algorithm. The location error of the proposed algorithm was statistically compared, and meaningful results were obtained regarding the reliability of the outcomes through Wilcoxon analysis. The simulation results demonstrated that the reinforcement learning-based particle swarm optimization algorithm can be effectively used for inverse kinematics solutions in serial robot manipulators.

1. Introduction

Nowadays, robots have entered many different areas in various sectors and have provided many conveniences to our lives [1], [2]. Robots, in particular, have become an important factor in industrial systems due to reasons such as their adaptation to different places, their ability to perform different tasks like humans, and their widespread use [1], [2], [3], [4]. When looked at the industry, there are many types of robots, large and small, fixed, autonomous, serial or parallel [2], [5], [6], [7]. Sectors and companies using this technology can become the pioneers of the sector as they can increase their production capacity and profit margins. Therefore, issues such as position control, motion control, acceleration or thrust controls, and structural designs of robots are examined by many different researchers [4], [8].

In industry, kinematic equations of robots must be derived and designs must be made before robots can be used [9]. Therefore, it can be said that the derivation of advanced and inverse kinematic equations in the control of robots is the first stage [4]. Thanks to these equations, the angle between the robot's joints or axes or the distance between the linear distances can be used to determine the position and direction of the robot, thanks to advanced kinematic equations. Or, given the direction and the positions it needs to go, the angles or distances between the joints in the robot can be calculated using inverse kinematic equations [1].

Analytical and iterative methods could be used to solve the inverse kinematics problem in robots. Solutions made with analytical methods become difficult in robots with high degrees of freedom due to the increase in degrees of freedom depending on the motors used in the robot [1], [10]. It requires a good background, skill and ability, especially in deriving and solving the inverse kinematic equations of robotic arms with higher degrees of freedom. On the other hand, the inverse kinematics problem can be solved with iterative methods. However, since the computation burden is high, their solutions take time.

Reinforcement learning is one of the machine learning methods [11], [12], [13]. Reinforcement learning is a type of learning based on perceiving the environment in which any agent (individual or person) is located, interacting with it,

* Corresponding author: muratcimen@subu.edu.tr
Received: 12.05.2024, Accepted: 25.09.2024

regulating its behavior according to this interaction, and taking action in accordance with its goal [13], [14], [15], [16], [17], [18]. The learning process is reward-based work. Therefore, the agent adopts the behavior or action that provides the most reward in order to achieve his goal. He gets used to the action that gives the least reward or punishment and learns to do that action less. For example, while a player makes moves or follows strategies in which he earns more points in any game, he is careful not to make strategies in which he earns fewer points or loses the game. In terms of reinforcement learning, the player becomes an agent (individual or person) and learns the game by interacting with the environment. Then he makes the moves that will earn him the most points. Therefore, reinforcement learning can be applied to many optimization problems in the game industry, industrial control applications, image processing applications, path planning or industry.

Optimization is the process of determining variables to minimize or maximize a certain objective criterion [19], [20], [21], [22], [23], [24], [25]. Optimization is encountered in almost every field, from logistics to finance, from chemistry to health, from machinery to electrical and electronics, from health to tourism, from automotive to construction, from medicine to food and retail, from education to social sciences [19], [26], [27], [28], [29]. The reason is that in the problem to be applied, there is a goal criterion that is desired to be maximized or minimized, and there are variables that will ensure this. In addition to classical optimization methods, there are swarm-based optimization algorithms inspired by nature. These algorithms were developed inspired by the behavior of living creatures in nature and are more successful in swarm-based global search [19], [24]. These algorithms can be given as Genetic Algorithm [30], [31], Particle Swarm Optimization (PSO) [20], [32], [33], Firefly Optimization (FA) [34], [35], Cuckoo Search Optimization [36], Gray Wolf Optimization [37], Flower Pollination Optimization [38], Whale Optimization Algorithm [35], [39] and many more. For instance, PSO is one of the well-known swarm-based optimization algorithms inspired by nature. PSO was developed by taking inspiration from the natural behaviors of fish and birds, such as finding food and escaping from predators [20], [32], [33]. In PSO, swarm experience and each particle's (fish or bird) speed are repeatedly used to solve optimization problem. By this means, algorithm try to find best particle in swarm. Overall, PSO is a basic and easy to apply any scientific or engineering optimization problem. Moreover, setting only a few parameters is sufficient. On the contrast, this algorithm suffers both complexity with high dimension, is very sensitive to particles optimism and particles' speed [3], [33], [40]. Especially, there are many studies to improve the performance of PSO algorithm. These studies are usually about parameter tuning or control of parameters. Parameter tuning is usually performed before the optimization process [41], [42], [43], [44]. The parameter tuning approach produces more efficient and successful results in some problems compared to the simple PSO method. However, its disadvantage is that some adaptation features are lost when the algorithm with the adjusted parameters is run. The other method, parameter control, adapts the algorithm to the specified conditions while the algorithm is running and a more dynamic optimization process is performed. In general, historical experience [45], [46], small test period [47], fuzzy logic [48], [49] and reinforcement learning methods [18], [50] are used in control-based PSO. However, there's no longer a need to manually design rules and fine-tune parameters, significantly reducing the burden for users. While parameters in Reinforcement Learning (RL) are important, experiments have demonstrated that a single set of parameters can yield good results across various algorithms and test functions. Consequently, in practice, adjusting RL parameters is often unnecessary. Additionally, by leveraging past experiences, the algorithm's applicability becomes broader, and its effectiveness improves. For instance, Liu et al. proposed the reinforcement learning method in the PSO method and applied this method to some multi-objective problems. In this adaptive method, the Q table has four states. These states, which can be adjusted adaptively, are created from the best particle among the particles used by the PSO algorithm and the particle in the swarm. The output of the Q table is determined as the inertia and correlation coefficients, which are the parameters of PSO [18]. Xu and Pi similarly tried to increase the performance of Q learning method by associating it with PSO [51]. However, unlike [18], they used different topologies while creating the states of Q table and added the diversity within the swarm in addition to the performance of the algorithm while creating the reward function. Xu and his colleagues similarly applied the reinforcement learning method to PSO algorithm. Unlike [18], they developed the algorithm by training an artificial intelligence model via deep deterministic policy (DDP) method. They applied DDP method and set its inputs as iteration, diversity and reward value obtained in the previous iteration. They determined the output as the parameters of PSO, inertia and correlation coefficients [50].

In the literature, inverse kinematic problems were usually solved by conventional methods [52]. It is a much more complicated and time-consuming solution because of non-linear equations [1]. These studies are still ongoing. Düzgün proposed new methods for solving the kinematics of robots in his doctoral thesis [2]. In addition to these methods, especially in recent years, artificial intelligence, fuzzy logic, metaheuristic optimization algorithms or reinforcement learning methods have been used to solve this problem. Koker et al. presented the inverse kinematic solution of a 3-joint robot using artificial intelligence [53]. Özüdoğru, in his master's thesis, used constructive artificial neural network to determine the joint angles and trajectory of industrial robots according to the target position [1]. Alamdar et al. presented an alternative solution to the inverse kinematic problem with the found logic-based ANFIS method [54]. The basic logic of the studies [53], [1], [54] is to collect data and determine the joint angles or joint distances of the robot using machine intelligence according to the collected data. In the study in [53], constructive neural networks, which are more successful in training, were used in [1], unlike the study. This proposed structure was implemented in real time on an industrial robot and successful results were obtained. On the other hand, in [54], they used ANFIS to be an alternative solution to solving kinematic problems. The superiority of ANFIS; It can produce successful results especially in situations containing uncertainty, as well as other artificial intelligence models. The reason for this is that it fuzzifies the fuzzy logic inputs by means of membership functions and then performs the clarification process according to the rules. The disadvantage is that many rules and membership functions need to be selected appropriately. Segoto et al. have collected synthetic data consisting of angles, speeds, torque from an industrial robot and trained it on an artificial neural network model. Then, they compared the performance of the trained model with machine learning logic [8]. Jin et al. have modeled a robot with six degrees of freedom. They used interpolation methods to determine joint angles [55]. However, Dereli transformed the inverse kinematic problem of a 7-axis robot with more joints into an optimization problem and then used the Artificial Bee Algorithm (ABC) algorithm to solve this problem [3]. In addition to this study, in [7], Quantum based PSO algorithm was applied to the same problem and more efficient and successful results were obtained. Quantum based PSO algorithm is the development of particles by affecting the quantum physics. Due to this feature, it has produced more successful results in finding the global search optimal than PSO algorithm. In addition, Quantum based PSO algorithm was compared to ABC algorithm in solving the inverse kinematic problem of a 7-DOF robot. Advantages of Quantum based PSO algorithm are the shorter computation time, fewer iterations and the number of particles. In the study conducted in [56], a similar process was performed with PSO for the target position determined by using the kinematic equations of a 7-DOF robot. However, the difference of this study is that the processing time and trajectories of the manipulator were optimized. In [57], an optimization study was performed to reduce the energy cost of a robot with a different objective criterion. Another machine learning method used in the control of robots is the reinforcement learning method. The advantage of reinforcement learning is that the robot is controlled by learning the behavior of the robot according to a reward function from a model determined by the researcher without needing the exact kinematic equation of the robot. In this direction, Avery et al. in their study, they carried out a study on the movement, speed and path planning of a robot with the incremental reinforcement learning method [6] In their study, Hou and Li proposed a reinforcement learning method for a 6-axis robot to grasp objects used in daily life that it recognizes in the image, and they achieved successful results [58].

In this study, a reinforcement learning-based particle swarm optimization algorithm was developed to solve a seven-axis robot inverse kinematics problem, and the problem was solved with particle swarm optimization and Q learning-based Particle Swarm optimization algorithm. Differences of the proposed model from [18], [51] are parameters and golden ratio has been inserted. The performance of the proposed reinforcement learning based particle swarm optimization algorithm was statistically compared with the mean value and standard deviation over different swarm sizes, iteration numbers and parameters. It has been observed that the proposed Q learning based Particle Swarm Optimization Algorithm produces successful results. Briefly, main contributions of the study as follows;

1. Q learning PSO algorithm (PSO-RL-Q) was developed using golden ratio that is used while calculating states,
2. PSO-RL-Q was applied to solve inverse kinematics of 7-DOF Robot Manipulator,
3. Statistical analysis and Wilcoxon test for results of proposed PSO-RL-Q were made to evaluate the performance.

2. Material and Method

2.1. Partical Swarm Optimization

Particle Swarm Optimization, proposed by Kennedy and Eberhart in 1995, is an algorithm inspired by the behavior of swarm of birds and fish [33]. It is an algorithm developed by modeling the behavior of each individual in the swarm, such as finding food and avoiding predators. Each individual in the swarm has a position (x_t) and speed (v_t) [19], [33], [59], [60]. Each individual interacts with other individuals in the swarm and iteratively updates its position and speed. This process takes place as in Equation 1.

$$v_{t+1} = \mathrm{w}v_t + c_1 r_1(p_{best} - x_t) + c_2 r_2(g_{best} - x_t) \quad (1)$$

$$x_{t+1} = x_t + v_t$$

As seen in Equation 1, the speeds of individuals are affected by g_{best}, which is in the best position in the swarm, and p_{best}, which is in the best position in the iteration. These effects are weighted with correlation coefficients (c_1, c_2) and random values between 0-1 in each iteration (r_1, r_2). In addition, the speed (x_t) of each individual is multiplied by a weight factor (w) and reflected in the next speed. The next location information of the individuals (x_{t+1}) is also updated by collecting their current location (x_t) and speed (v_t) [19], [33], [59], [60].

2.2. Q Learning

Q learning algorithm, one of the machine learning methods, dates back to Bellman's studies on optimal control theory, that is, in the 1950s [11], [61]. Bellman's work, which tried to solve the dynamic optimization problem to find the control signal in discrete systems, formed the basis of the Q learning algorithm in later years. Q learning algorithm basically requires an agent to interact with the environment according to its own knowledge and experience. After this interaction, according to the data collected from the environment, the agent learns the environment according to a reward/punishment value and decides its next move according to the current situations[12], [62]. As shown in Figure 1, the Agent generates a movement signal from the environment according to the situation. According to this signal, the agent moves within the environment and moves to the next state. The next state also produces a reward signal [14].

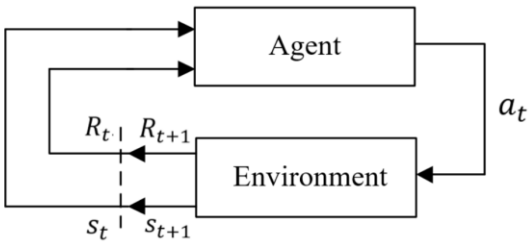


Figure 1. Environment and Agent Interaction.

Q learning method and State-Action-Reward-State-Action (SARSA) methods are model-independent or model-free reinforcement learning methods [13], [17]. Reinforcement learning also has a feature that learns mostly from behavior. It achieves this during the interaction of an agent with the environment. The determined agent interacts with the environment in a certain way and produces an output called reward by evaluating the outputs of the environment. Depending on the current state of the agent, its interaction with the environment, and the reward for the agent's next state, the agent that learns the environment begins to act in a way that will earn a higher reward at each step. They can be preferred especially in applications that are difficult to model, as they can learn by experiencing the results of actions rather than using a model [14]. In particular, reinforcement learning is provided by the Bellman equation suggested by Richard Bellman and used by Watkin in reinforcement learning, and the actions and outcomes depending on the situations are learned. In this learning method, the system is learned in terms of situations and action and reward value rather than a specific model. The simplest version of the Bellman equation used in Q learning is given in Equation 2. s_t used in the equation is for the instantaneous state at time t, a_t is for the instantaneous action at time t, s_{t+1} is for the instantaneous state at time t+1, $R(s_t, a_t)$ is for the reward value of s_t and a_t, $Q(s_t, a_t)$ is for the value of s_t and a_t. state of training value, α is called the learning factor, and γ is called the discount factor. $\max_a Q(s_{t+1}, a)$ is the value at which the maximum Q value is produced according to the α action in case s_{t+1}. By performing this process in each iteration, Q values are updated [11], [12], [14], [17].

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha \left[R(s_t, a_t) + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right] \quad (2)$$

Additionally, the epsilon greedy method is applied to determine Q values. This method allows the Agent to visit all possible states during learning.

Algorithm 1. Q learning Pseudocode

Input:
State (s)
Action (a_t)
Learning rate (α)
Discount factor (γ)
Reward $R(s_t, a_t)$
Updated table $Q(s_t, a_t)$
Output:
Selected action according to updating table $Q(s_t, a_t)$
For iter=1, Max_iteration do
  Initialise state s_t
  For t=1, Max_iteration do
    Choose a_t with ϵ greedy probability
    Execute a_t and observe state s_{t+1} and reward r_t
    Update table $Q(s_t, a_t) = Q(s_t, a_t) + \alpha \left[R(s_t, a_t) + \gamma \underset{a}{max} Q(s_{t+1}, a) - Q(s_t, a_t)\right]$
  End for
End for

Thus, the q values that will maximize the reward by learning the environment better are determined. The pseudo code of the Q learning algorithm is given in Figure 1. This method is an off policy and the learning agent learns the value function based on the current action derived from the policy currently in use. The pseudocode of the Q learning algorithm is given in Algorithm 1.

2.3. Kinematic Model of the Robot

The homogeneous transformation matrix is expressed by Equation 3. Transformation matrices of the serial robot manipulator are obtained by the David Hartembert method (DH) [4], [9]. The DH parameters listed in Table 1 represent the connection length (l_i), connection angle (α_i), connection offset (d_i) and joint angles (θ_i). Transformation matrices of adjacent link coordinate frames are obtained according to the DH parameters of the robotic manipulator given in Figure 2 and Equation 1. Equation 5 is obtained by multiplying the six transformation matrices in Equation 4. Lengths are in meters and angles are in degrees [3].

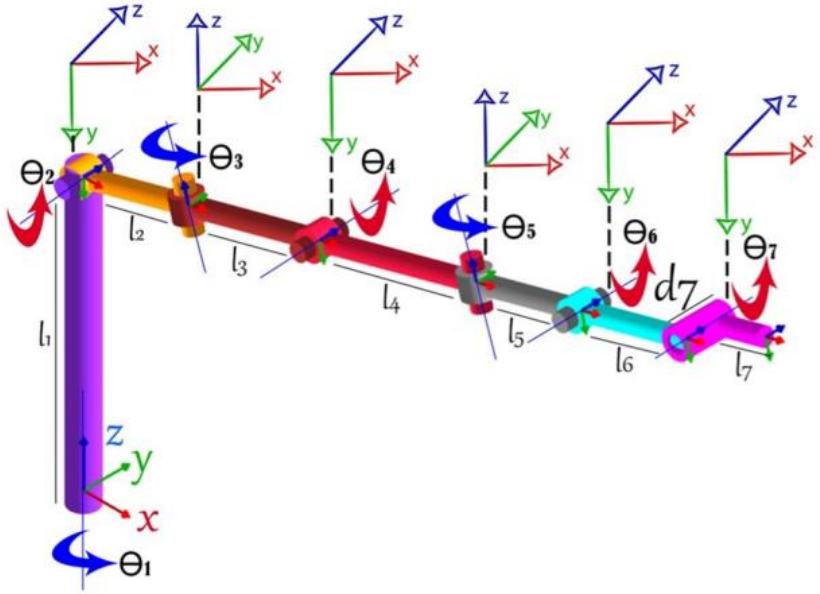


Figure 2. 7-DOF Robot

$$ {}^{i-1}_{i}T\left(cos(\theta_i)\right) = \begin{bmatrix} cos(\theta_i) & -cos(\alpha_i)sin(\theta_i) & sin(\alpha_i)sin(\theta_i) & l_i cos(\theta_i) \\ sin(\theta_i) & cos(\alpha_i)cos(\theta_i) & -cos(\theta_i)sin(\alpha_i) & l_i sin(\theta_i) \\ 0 & sin(\alpha_i) & cos(\alpha_i) & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3} $$

$$ {}^{1}_{0}T = \begin{bmatrix} cos(\theta_1) & 0 & -sin(\theta_1) & 0 \\ sin(\theta_1) & 0 & cos(\theta_1) & 0 \\ 0 & -1 & 0 & l_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad {}^{2}_{1}T = \begin{bmatrix} cos(\theta_2) & 0 & sin(\theta_2) & l_2 cos(\theta_2) \\ sin(\theta_2) & 0 & -cos(\theta_2) & l_2 sin(\theta_2) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} $$

$$ {}^{3}_{2}T = \begin{bmatrix} cos(\theta_3) & 0 & sin(\theta_3) & l_2 cos(\theta_3) \\ sin(\theta_3) & 0 & -cos(\theta_3) & l_2 sin(\theta_3) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad {}^{4}_{3}T = \begin{bmatrix} cos(\theta_4) & 0 & sin(\theta_4) & l_4 cos(\theta_4) \\ sin(\theta_4) & 0 & -cos(\theta_4) & l_4 sin(\theta_4) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} $$

$$ {}^{5}_{4}T = \begin{bmatrix} cos(\theta_5) & 0 & -sin(\theta_5) & l_5 cos(\theta_5) \\ sin(\theta_5) & 0 & cos(\theta_5) & l_5 sin(\theta_5) \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad {}^{6}_{5}T = \begin{bmatrix} cos(\theta_6) & -sin(\theta_6) & 0 & l_6 cos(\theta_6) \\ sin(\theta_6) & cos(\theta_6) & 0 & l_6 sin(\theta_6) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4} $$

$$ {}^{7}_{6}T = \begin{bmatrix} cos(\theta_7) & -sin(\theta_7) & 0 & l_7 cos(\theta_7) \\ sin(\theta_7) & cos(\theta_7) & 0 & l_7 sin(\theta_7) \\ 0 & 0 & 1 & d_7 \\ 0 & 0 & 0 & 1 \end{bmatrix} $$

$$ {}^{0}_{7}T = {}^{0}_{1}T(\theta_1){}^{2}_{1}T(\theta_2){}^{4}_{3}T(\theta_3){}^{5}_{4}T(\theta_4){}^{6}_{5}T(\theta_5){}^{7}_{6}T(\theta_6) = \begin{bmatrix} n_x & o_x & a_x & P_x \\ n_y & o_y & a_y & P_y \\ n_z & o_z & a_z & P_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5} $$

2.3. Objective Function

In the application of reinforcement learning based Particle swarm optimization to inverse kinematics equations, each individual (x_t) is the joint variables $(\theta_1^o, \theta_2^o, \theta_3^o, \theta_4^o, \theta_5^o, \theta_6^o, \theta_7^o)$ of the 7-axis serial robot manipulator [3]. In order to provide an optimal solution, the end effector reaching the target position is achieved through optimal adjustments. The robot arm has many destination paths from its starting point to its destination. The important issue at this point is that the manipulator reaches the target with minimum error with the fitness function.

$$ E_r = \sqrt{(P_x - P_x')^2 + \left(P_y - P_y'\right)^2 + (P_z - P_z')^2} \tag{6} $$

The positions to be calculated P_x', P_y', P_z' given in Equation 6 represent the target position of the end effector and E_r represents the error between P and P'. The aim of this study is to minimize the E_r error value.

2.3. Q Learning Based Particle Swarm Optimization

Q learning method is one of the Reinforcement Learning algorithms [13], [14]. By integrating Q learning into particle swarm optimization, it is aimed to update the parameters depending on the situation and thus increase the performance of the Particle Swarm optimization method. In this direction, situations were determined to be used within the Q learning method. In order to determine these states, $f_{min}, f_{max}, v_{min}, v_{max}, d_{min}, d_{max}$ values to be used in the states are determined. The pseudocode for performing this operation is given in Algorithm 2.

Algorithm 2. Pseudocode for calculation of $f_{min}, f_{max}, v_{min}, v_{max}, d_{min}, d_{max}$ values

$f_{min} = func(x(t)), f_{max} = func(x(t)), v_{min} = \|v(t)\|, v_{max} = \|v(t)\|,$
$d_{min} = \|x(t) - p\|, d_{max} = \|x(t) - p\|$
For t=1, maximum generation **do**
  If $f_{min} > func(x(t))$
    $f_{min} > func(x(t))$
  End If
  If $f_{max} < func(x(t))$
    $f_{max} = func(x(t))$
  End If
  If $v_{min} > \|v(t)\|$
    $v_{min} = \|v(t)\|$
  End If
  If $v_{max} < \|v(t)\|$
    $v_{max} = \|v(t)\|$
  End If
  If $d_{min} > \|x(t) - p\|$
    $d_{min} = \|x(t) - p\|$
  End If
  If $d_{max} < \|x(t) - p\|$
    $d_{min} = \|x(t) - p\|$
  End If
End for

Values have been generated by comparing the $f_{min}, f_{max}, v_{min}, v_{max}, d_{min}, d_{max}$ produced with Psoudecode with the x, d, v variables used in PSO, and corresponding states would be created. This part is inspired by golden section search. The golden ratio is a ratio that can be found in the shape and structure of countless living and non-living entities in nature. The golden ratio is a numerical ratio that was discovered by the ancient Egyptian and Greek civilizations and has been applied to works of art such as sculpture, painting, and architecture for centuries. The value of the golden ratio is $\emptyset = \frac{1+\sqrt{5}}{2} \cong 1.61803$. As seen in Figure 3, there is a large piece |AC| (L) and a small piece |BC| (S) between points A and B. As seen in Figure 3 and in Equality 7, the ratio of the large piece to the small piece is a situation where it is equal to a fixed value ratio. When Equation 7 is arranged and Equation 8 is obtained, the ratio of the small piece |BC| to |AC| is calculated as 0.61803. In this study, the value $\emptyset - 1 = 0.61803$ was used from this value.

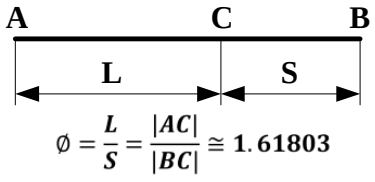


Figure 3. Golden ratio

$$\emptyset = \frac{L}{S} = \frac{|AC|}{|BC|} \cong \frac{1+\sqrt{5}}{2} \cong 1.61803 \quad (7)$$

$$\frac{1}{\emptyset} = \frac{S}{L} = \frac{2}{1+\sqrt{5}} = \emptyset - 1 \cong 0.61803 \quad (8)$$

When environment and nature are observed, the golden ratio is encountered in many places. One of the first places where the golden ratio is used in architecture is seen in Figure 4 and Figure 5. Or, as a few examples from nature itself, the proportions of snail shells, plants and human limbs are also places that contain the golden ratio, as seen in Figure 6 [63].

The golden ratio, known for centuries, is a method that has been applied to optimization problems before [64]. The golden ratio is aimed to be developed as a new and highly efficient method by adding it to the Q learning algorithm used with PSO. Therefore, in this study, the state assignment was carried out by comparing the objective value, speed and distance to the local best position of the particles or individuals used in the states produced for the swarm element, with the golden ratio. For example, according to the objective criterion to be produced for the herd element at index t, the value of $state_1$ is calculated using Equation 9. The ε value used in the equation was chosen as a small value of 10^{-3} so that the expression does not become infinite.

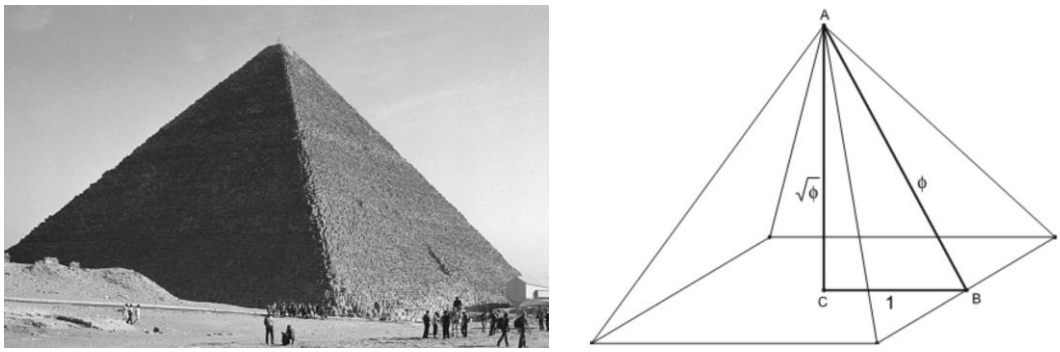


Figure 4. Pyramids in Egypt

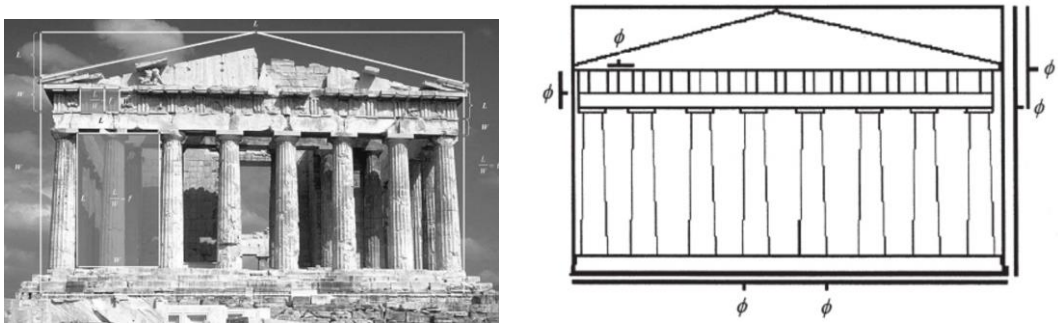


Figure 5. Parthenon in Athens, Greece

Figure 6. Snail shell, plant and human

Then, $state_2$ for the speeds of the particles was created as in Equation 10. Additionally, the situation obtained according to the distance of the particles to the best position is given in Equation 11.

$$state_1(\mathrm{t}) = \begin{cases} 1 & \dfrac{f(x_t) - f_{min}}{f_{max} - f_{min} + \varepsilon} < \emptyset - 1 \\ 0 & else \end{cases} \quad (9)$$

$$state_2(\mathrm{t}) = \begin{cases} 1 & \dfrac{\|v(t)\| - v_{min}}{v_{max} - v_{min} + \varepsilon} < \emptyset - 1 \\ 0 & else \end{cases} \quad (10)$$

$$state_3(\mathrm{t}) = \begin{cases} 1 & \dfrac{\|d(t)\| - d_{min}}{d_{max} - d_{min} + \varepsilon} < \emptyset - 1 \\ 0 & else \end{cases} \quad (11)$$

The parameters of the PSO algorithm will be updated according to $state_1$, $state_2$ and $state_3$ used to create the states of the Q learning algorithm. Algorithm 3 is given to better express this process. The main working logic of the algorithm is that while PSO is running, it creates actions using the Q learning table and tries to find the optimum point by updating the parameters of the PSO algorithm according to the situations.

Algorithm 3. PSO-RL-Q learning Psoude Code

1: Initialize population and reset Q table, determine f
2: **For** t= 1,maximum generation **do**
3: $f_{min}, f_{max}, v_{min}, v_{max}, d_{min}, d_{max}$ calculate
4: $state_1 = []$, $state_2 = []$, $state_3 = []$ ve $states = []$
5: **For** i= 1, population size **do**
6: $state_1$(i), $state_2$(i), $state_3$(i) calculate
$states(i) = [state_1, state_2, state_3]$
7: **If** rand()>0.5
8: Action(i)=rand_int(0-2)
9: **Else**
10: Specify $Action(i) = \arg\left(\underset{a}{max} Q(s_t, a), Q(s_t, a)\right)$
11: **If End**
Specify w, c_1 and c_2 with respect to Action(i)
12:
$$\omega = \begin{cases} 0.05 & Action(i) = 0 \\ 0.50 & Action(i) = 1 \\ 0.95 & Action(i) = 2 \end{cases}, c_1 = \begin{cases} 0.95 & Action(i) = 0 \\ 0.50 & Action(i) = 1, \\ 2.00 & Action(i) = 2 \end{cases}$$
$$c_2 = \begin{cases} 0.95 & Action(i) = 0 \\ 0.50 & Action(i) = 1 \\ 2.00 & Action(i) = 2 \end{cases}$$
13: $v_{i.d}(t+1) = wv_{i.d}(t) + c_1 r_1 \left(p_i(t) - x_{i.d}(t)\right) + c_2 r_2 \left(p_a(t) - x_{i.d}(t)\right)$
14: $x_{i.d}(t+1) = x_{i.d}(t) + v_{i.d}(t)$
15: **End For**
16: $Reward = [\ \ ], next\_state = [\ \ ]$
17: **For** t=1, population size **do**
18: $f_{min}, f_{max}, v_{min}, v_{max}, d_{min}, d_{max}$ update
19: $Reward$ (i)$= -f(x_t)$
20: $state_1$(i), $state_2$(i), $state_3$(i) calculate
$next\_states(i) = [state_1, state_2, state_3]$
21: Apply Equation 2' and update $Q(states(i), Action(i))$ Table
22: **End For**
23: **End For**

In this context, when the psoudecode of the PSO-RL-Q algorithm is examined, the population to be used in PSO is first created. Then, the function f is determined and initial values are assigned to the Q table. In 2nd line, a for loop was created to repeat the operations in the algorithm for the maximum number of generations. In order to be used in calculating the states within the loop, $f_{min}, f_{max}, v_{min}, v_{max}, d_{min}, d_{max}$ values and variables to be used in the algorithm are assigned. Then, the states are calculated within the loop. After that, the control signal was determined depending on the rand() probability. According to this process, $Action$s are determined for each particle. According to the determined action, the parameters (ω, c_1, c_2) to be used in PSO were determined and the positions and speeds of each individual were updated. Afterwards, variables were created for Reward and next_state states. These variables are used to determine reward values according to the positions of the individuals in the swarm, to determine the status of the individuals according to their new positions, and to update the Q table. Therefore, a loop was created at the 17th line to update each individual. In this cycle, the negative value of each individual's goal criterion value is determined as reward. This is due to minimization. Subsequently, the states of each individual were calculated according to their x, d, v values, and next_states were created. Then, the value in the Q Table is updated according to the state and action of the relevant individual.

A Q table must be created to run the algorithm. This created Q table depicted the states, actions and rewards. The sample Q table obtained after running the algorithm is given in Table 1. This

table contains states and actions. The values of the states obtained depending on the states and actions are given in the table.

3. Results and Discussion

In order to test the performance of the PSO algorithm and PSO- RL-Q algorithm in different parameters, the inverse kinematics problem of a 7-axis serial robot manipulator was used. In this study, the position of the end effector was chosen randomly. In this direction, Dereli et al. used the DH method to derive the kinematic equations of this robot and used the parameters in Table 2.

The parameters used to compare the performance of the proposed PSO- RL-Q algorithm with different criteria such as number of iterations and swarm size are given in Table 3. As can be seen, while the control parameters ω , c_1, c_2 for PSO_par_1 were $0.05, 0.95, 0.95$, the tests were carried out using the iteration numbers as 100, 1000 and the Swarm size as 50 and 100. Similarly, it has been implemented in PSO_par_2, PSO_par_3, and PSO-RL-Q algorithm. To test performances of algorithms, two different points were considered for the final position of the manipulator. According to these points, each algorithm was run independently 30 times. Then, the results obtained were compared statistically. Additionally, their statistical reliability was compared to test whether there were significant results.

Table 1. A Q table results for population size 50 and maximum iteration 1000 for PSO-RL-Q

$States$	$Action = 0$ $\omega = 0.05$ $c_1 = 0.50$ $c_2 = 0.95$	$Action = 1$ $\omega = 0.50$ $c_1 = 0.50$ $c_2 = 0.95$	$Action = 2$ $\omega = 0.95$ $c_1 = 2.00$ $c_2 = 2.00$
(0,0,0)	-2.2268×10^{-17}	-2.0845×10^{-17}	-3.7241×10^{-17}
(0,0,1)	-1.0336×10^{-16}	-1.0861×10^{-6}	-3.2223×10^{-7}
(0,1,0)	-7.3675×10^{-17}	-2.7739×10^{-6}	-4.7972×10^{-6}
(0,1,1)	-3.1769×10^{-11}	-1.5064×10^{-16}	-1.6349×10^{-9}
(1,0,0)	-0.4615	-0.3071	-0.3664
(1,0,1)	-0.9919	-0.87698	-1.1141
(1,1,0)	-0.2333	-0.33376	-0.3182
(1,1,1)	-1.6299	-2.3279	-2.2797

Table 2. DH parameters of 7-DOF Robot

i	$l_i(m)$	$\alpha_i(°)$	$d_i(m)$	$\theta_i(°)$
1	$l_1 = 0.5$	-90	0	$-180 < \theta_1 < 180$
2	$l_2 = 0.2$	90	0	$-90 < \theta_2 < 30$
3	$l_3 = 0.25$	-90	0	$-90 < \theta_3 < 120$
4	$l_4 = 0.3$	90	0	$-90 < \theta_4 < 90$
5	$l_5 = 0.2$	-90	0	$-90 < \theta_5 < 90$
6	$l_6 = 0.2$	0	0	$-90 < \theta_4 < 60$
7	$l_7 = 0.2$	0	$d_7 = 0.05$	$-90 < \theta_4 < 90$

Table 3. Parameters of Algorithms

Method	Parameters	Iteration Numbers	Swarm Size
PSO_par_1	$\omega = 0.05,\quad c_1 = 0.95, c_2 = 0.95$	100, 1000	50,100
PSO_par_2	$\omega = 0.50,\quad c_1 = 0.50, c_2 = 0.50$	100, 1000	50,100
PSO_par_3	$\omega = 0.95,\quad c_1 = 2.00, c_2 = 2.00$	100, 1000	50,100
PSO-RL-Q	$\begin{cases} \omega = 0.05, & c_1 = 0.95, c_2 = 0.95 & a_t = 0 \\ \omega = 0.50, & c_1 = 0.50, c_2 = 0.50 & a_t = 1 \\ \omega = 0.95, & c_1 = 2.00, c_2 = 2.00 & a_t = 2 \end{cases}$	100, 1000	50, 100

Table 4. Optimal joint angle values produced by PSO_par_1, PSO_par_2, PSO_par_3, PSO_Q algorithms according to target positions by swarm size 50, iteration number 100.

Position	Method	θ_1^o	θ_2^o	θ_3^o	θ_4^o	θ_5^o	θ_6^o
[-25.00 100.00 50.00]	PSO_par_1	-39.9982	-40.3681	54.3798	18.5928	7.3264	45.6443
	PSO_par_2	-0.07114	35.5679	3.5509	-78.1808	-21.9988	51.0210
	PSO_par_3	-73.2728	-90.0	13.9466	-20.2021	12.5424	-15.0000
	PSO-RL-Q	-16.0189	-10.2971	-1.4765	-33.4392	72.7653	22.1168
[-30.00 20.00 80.00]	PSO_par_1	-77.7382	-23.6192	68.1264	32.7416	110.5017	56.5476
	PSO_par_2	-30.6456	14.0067	-90.0	89.9859	-52.3304	39.1834
	PSO_par_3	-17.6197	-90.0	-90.0	90.0	119.9999	69.9999
	PSO-RL-Q	9.8517	-31.9813	-85.8088	69.2948	-78.6991	-12.5035

The joint angles obtained for the inverse kinematic analysis results in the case of swarm size 50, iteration number 100, according to the determined target positions [-25.00 100.00 50.00] and [-30.00 20.00 80.00] are listed in Table 4. In general, it has been observed that some algorithms produce joint angles at different angles. The reason for this is that Equation 6 can be minimized, at different angles of joint if the robot has many joints of robot.

At Table 5, Px, Py and Pz positions produced by the algorithms for 2 different locations, location errors of the algorithms and processing times are given. Although the proposed PSO-RL-Q method produced the lowest error in terms of position error, it showed lower performance than other methods in terms of computation time. In addition, the best values produced by the algorithms according to the iterations are shown in Figure 7. The best results of these algorithms, which contain random parameters, were close during iterations. However, when the iteration continued, the dynamics of these algorithms differentiated the results. The part where this will be noticed best is the part where statistical analysis will be performed. In for that, the minimum and maximum mean and standard deviation values produced by the algorithms as a result of 30 different simulation studies where the number of swarm is 50 and the iteration is 100 are given in Table 6. The results are written with the best values in bold. However, when examined in terms of mean value, the proposed PSO-RL-Q algorithm produced more successful results than others.

Table 5. Optimal joint angle values produced by PSO_par_1, PSO_par_2, PSO_par_3, PSO_Q algorithms according to target positions by swarm size 50, iteration number 100.

Target Position (cm)	Method	Px (cm)	Py (cm)	Pz (cm)	Position Error	Computation Time (sec)
[-25.00 100.00 50.00]	PSO_par_1	-24.99	99.99	50.00	9.4954×10^{-5}	1.40824
	PSO_par_2	-25.00	100.00	49.99	2.2551×10^{-14}	1.87351
	PSO_par_3	-24.05	99.92	50.82	1.2634×10^{-2}	**1.33919**
	PSO-RL-Q	-24.99	100.00	49.99	$\mathbf{8.0251\times 10^{-16}}$	4.51963
[-30.00 20.00 80.00]	PSO_par_1	-29.48	31.23	79.25	1.5306×10^{-2}	1.11302
	PSO_par_2	-30.00	29.99	80.00	3.7180×10^{-15}	1.13652
	PSO_par_3	-29.97	29.82	79.92	1.9437×10^{-3}	**1.05694**
	PSO-RL-Q	-30.00	30.00	80.00	$\mathbf{1.7554\times 10^{-15}}$	3.70225

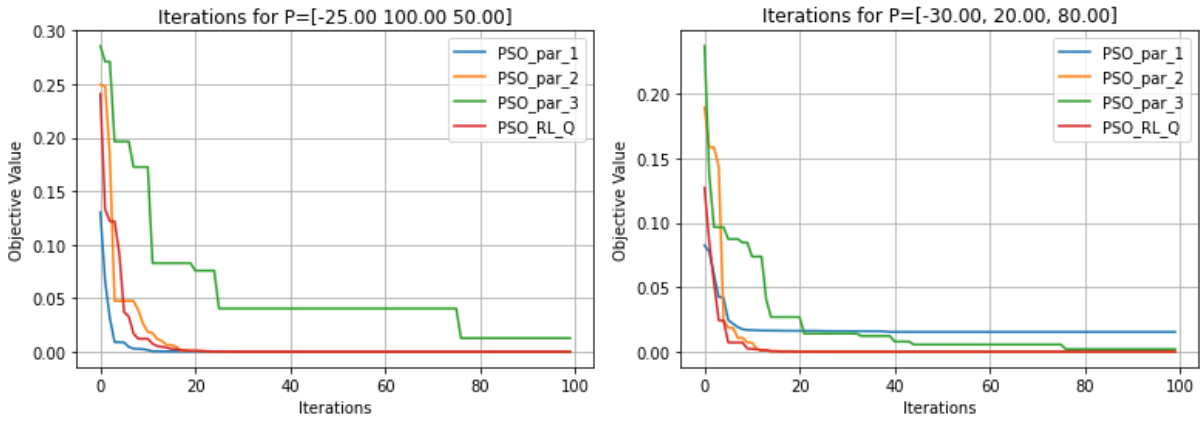


Figure 7. Objective values along the iterations for Swarm Size 50, iteration number 100

For statistical reliability, the Wilxocon test was applied to the trials. The results obtained are given for different locations in Table 7 and Table 8. From the pairwise comparison of the results, it is seen that, since the significance value is lower than 0.05, there is a significant difference in the results except for PSO_par_1/PSO_par_1, PSO_par_2/PSO_par_2, PSO_par_4/PSO_par_4 and PSO-RL-Q/PSO_par_3 results.

It has been observed that the PSO algorithm is successful in solving the inverse kinematics problem of the 7-DOF robot. At the same time, it has been observed that changing the parameters of PSO or adding reinforcement learning into PSO increases the performance of the algorithm. However, when the number of iterations and swarm sizes of the algorithms are increased, the processing time becomes longer because the search takes longer and with more particles. At the same time, the potential to find its global minimum is higher. For this case, only for the [-25.00 100.00 50.00] position, the algorithms were applied to the solution of the inverse kinematic problem with the swarm size being 100 and the number of iterations being 1000, and the results were analyzed. The results obtained according to the iterations are depicted in Figure 4. As can be seen from these results, since the number of swarms and the number of iterations are high, all algorithms converge to the same result, which is the global result.

Table 6. Statistical results for swarm size 50, iteration number 100.

Target Position (cm)	Method	Minimum	Max	Mean	Standard deviation
[-25.00 100.00 50.00]	PSO_par_1	9.4954×10^{-5}	0.2327	4.8891×10^{-2}	3.0914×10^{-2}
	PSO_par_2	2.2484×10^{-14}	0.0332	8.7174×10^{-3}	3.2777×10^{-3}
	PSO_par_3	1.2634×10^{-2}	0.1744	3.5619×10^{-2}	4.4413×10^{-3}
	PSO-RL-Q	$\mathbf{7.8504\times 10^{-17}}$	$\mathbf{1.2567\times 10^{-7}}$	$\mathbf{2.2654\times 10^{-8}}$	4.7879×10^{-9}
[-30.00 20.00 80.00]	PSO_par_1	1.5306×10^{-2}	0.3715	9.5057×10^{-2}	1.7430×10^{-1}
	PSO_par_2	3.8298×10^{-15}	$\mathbf{9.5388\times 10^{-2}}$	2.3801×10^{-2}	1.0253×10^{-2}
	PSO_par_3	1.9437×10^{-3}	0.1011	2.4124×10^{-2}	2.3077×10^{-2}
	PSO-RL-Q	**0.0**	0.1162	$\mathbf{2.1863\times 10^{-2}}$	1.1943×10^{-2}

Table 7. Wilxocon test results for [-25.00 100.00 50.00], swarm size 50, iteration number 100.

	PSO_par_1	PSO_par_2	PSO_par_3	PSO-RL-Q
PSO_par_1	1.00	9.01×10^{-10}	8.58×10^{-33}	9.03×10^{-10}
PSO_par_2	9.01×10^{-10}	1.00	1.30×10^{-25}	0.0775
PSO_par_3	8.58×10^{-33}	1.30×10^{-25}	1.00	2.58×10^{-28}
PSO-RL-Q	9.03×10^{-10}	0.0775	2.58×10^{-28}	1.00

Table 8. Wilxocon test results for [-30.00 20.00 80.00], swarm size 50, iteration number 100.

	PSO_par_1	PSO_par_2	PSO_par_3	PSO-RL-Q
PSO_par_1	1.00	2.57×10^{-26}	3.92×10^{-13}	1.39×10^{-28}
PSO_par_2	2.57×10^{-26}	1.00	7.14×10^{-24}	0.0227
PSO_par_3	3.92×10^{-13}	7.14×10^{-24}	1.00	6.92×10^{-26}
PSO-RL-Q	1.39×10^{-28}	0.0227	6.92×10^{-26}	1.00

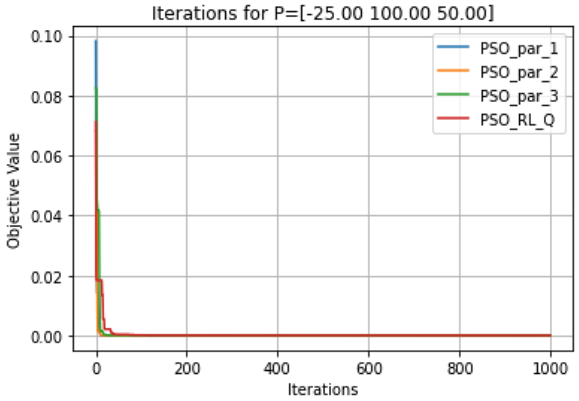


Figure 8. Objective values along the iterations for [-25.00 100.00 50.00] Position ve Swarm Size 100, iteration number 1000.

Furthermore, for a numerical comparison of the results obtained, the positions reached, position errors and processing times are demonstrated at Table 9. As can be seen, as the processing load increases, processing times vary between 9 and 14 seconds. Sürü sayısı ve iterasyon sayısı arttığında, all algorithms found the same global optimal solution within 30 different trials. However, the calculation times have changed due to differences in calculations and operations. Although the results are the same, as can be seen, the PSO_par_1 algorithm produced the best result, while the PSO-RL-Q algorithm was the slowest. As a result, it is more correct to choose the faster

method among the methods that show the same performance.

Statistical analysis results are demonstrated at Table 10. Since the results of 30 independent experiments reached the desired global value, the minimum, maximum, average value and standard deviation values were obtained as 0. The reason why all the results are 0 in Table 10, is because when the swarm numbers and iteration numbers of the algorithms were increased, all algorithms found the same global optimal solution within 30 different trials. Therefore, when the optimal solution was one, the error function, the mean value, minimum value, maximum value and standard deviation of the error function were found to be 0.

Table 9. Optimal joint angle values produced by PSO_par_1, PSO_par_2, PSO_par_3, PSO_Q algorithms according to target positions by swarm size 100, iteration number 1000.

Target Position (cm)	Method	Px (cm)	Py (cm)	Pz (cm)	Position Error	Computation Time (sec)
[-25.00 100.00 50.00]	PSO_par_1	-25.00	100.00	50.00	**0.0**	**9.47613**
	PSO_par_2	-25.00	100.00	50.00	**0.0**	11.7058
	PSO_par_3	-25.00	100.00	50.00	**0.0**	12.7068
	PSO-RL-Q	-25.00	100.00	50.00	**0.0**	13.2338

Table 10. Statistical results for swarm size 100, iteration number 1000.

Target Position (cm)	Method	Minimum	Max	Mean	Standard deviation
[-25.00 100.00 50.00]	PSO_par_1	**0**	**0**	**0**	**0**
	PSO_par_2	**0**	**0**	**0**	**0**
	PSO_par_3	**0**	**0**	**0**	**0**
	PSO-RL-Q	**0**	**0**	**0**	**0**

In addition, the Wilxocon test was performed to test whether there was a significant difference in the results. To perform this test, the results obtained by the algorithms in 30 different experiments were used. The algorithms always found the same error value with a high number of iterations and a high number of swarm. Therefore, there was no difference in the results obtained when high number of iterations and swarm sizes was used in the algorithms. The results obtained when the Wilxocon test is applied for the reliability of this statistical result are given in Table 11. In the results given in Table 11, it was seen that there was no significant difference in the results since all values were greater than 0.05. The reason why all the results are 1.00 in Table 11, is because when the swarm numbers and iteration numbers.

As the results are examined, changing the parameters of the PSO algorithm affects the performance of the algorithm, and the solution performance of the algorithm can be increased when a reinforcement learning-based method is added, as in this study. Ultimately, 7-DOF robot used in industry was turned into an inverse kinematic problem with optimization using advanced kinematic equations with DH to determine its angles and distances using optimization algorithms. So that, PSO-RL-Q algorithm distinctly showed successful results in finding the joint angles and distances of the robot within short iteration number than plain PSO algorithm.

Table 11. Wilxocon test results for [-25.00 100.00 50.00], swarm size 100, iteration number 100.

	PSO_par_1	PSO_par_2	PSO_par_3	PSO-RL-Q
PSO_par_1	1.00	1.00	1.00	1.00
PSO_par_2	1.00	1.00	1.00	1.00
PSO_par_3	1.00	1.00	1.00	1.00
PSO-RL-Q	1.00	1.00	1.00	1.00

4. Conclusion and Suggestions

In this study, a new optimization method was applied to find the joint angles of a 7-axis robot at the desired position. In this direction, first the kinematic equation of the robot was obtained, and then the PSO-based PSO_par_1, PSO_par_2, PSO_par_3 and reinforcement learning-based proposed PSO-RL-Q optimization method using golden section have been used to find the manipulator angle and distances. In order to compare the performances of the algorithms, 30 independent studies have been carried out and the results were compared statistically. Statistically, it has been observed that the proposed PSO-RL-Q algorithm produces more successful results than standard PSO algorithms when the number of iterations is 100 and swarm size is 50. From the results, while the swarm size of the algorithms is 50 and the number of iterations is 100, the PSO-RL-Q algorithm performed the best result by producing 2.2654×10^{-8} mean value for the position [-25.00 100.00 50.00]. Also, it produced more successful results than other algorithms by producing a mean value of 2.1863×10^{-2} for the position [-30.00 20.00 80.00]. In addition, the test results obtained were subjected to the Wilxocon test and it was observed that there was a significant difference in the results when the number of iterations was 50 and iteration number are 100. Since reinforcement learning works on a reward basis, it appears to be a method that can be easily applied to many different algorithms. In future studies, it is planned to apply these proposed methods and their variations to different optimization problems and systems.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] F. Özüdoğru, "Endüstriyel Robot Kolu Modelinin Hedef Konum Eklem Açilarinin Yapıcı Sinir Ağı Ile Kestirimi Ve Kontrollü Yörünge Uygulamasi," Yüksek Lisans, Elektrik Elektronik Mühendisliği, Tokat, 2020.

[2] E. Düzgün, "Paralel ve Hibrit Manipülatörlerin Ileri Kinematik Çözümü Için Yeni Metotlar Geliştirilmesi," Doktora, Fen Bilimleri Enstitüsü, Bursa, 2023.

[3] S. Dereli and R. Köker, "A meta-heuristic proposal for inverse kinematics solution of 7-DOF serial robotic manipulator: quantum behaved particle swarm algorithm," *Artif Intell Rev*, vol. 53, pp. 949–964, 2020.

[4] F. Aysal, İ. Çelik, E. Cengiz, and Y. Oğuz, "A comparison of multi-layer perceptron and inverse kinematic for RRR robotic arm," *Politeknik Dergisi*, vol. 27, no. 1, pp. 121–131, 2023.

[5] S. Hwang, H. Kim, Y. Choi, K. Shin, and C. Han, "Design Optimization Method for 7 DOF Robot Manipulator Using Performance Indices," *International Journal of Precision Engineering and Manufacturing*, vol. 18, no. 3, pp. 293–299, 2017.

[6] A. Avaei, L. van der Spaa, L. Peternel, and J. Kober, "An incremental inverse reinforcement learning approach for motion planning with separated path and velocity preferences," *Robotics*, vol. 12, no. 2, 2023.

[7] S. Dereli and R. Köker, "Simulation based calculation of the inverse kinematics solution of 7-DOF robot manipulator using artificial bee colony algorithm," *SN Appl Sci*, vol. 2, no. 1, p. 27, 2020.

[8] S. Baressi Šegota, N. Anđelić, M. Šercer, and H. Meštrić, "Dynamics Modeling of Industrial Robotic Manipulators: A Machine Learning Approach Based on Synthetic Data," *Mathematics*, vol. 10, no. 7, p. 1174, 2022.

[9] Z. Bingül and S. Küçük, *Robot Kinematiği*. Umuttepe Yayınları, 2019.

[10] H. Danaci, L. A. Nguyen, T. L. Harman, and M. Pagan, "Inverse Kinematics for Serial Robot Manipulators by Particle Swarm Optimization and POSIX Threads Implementation," *Applied Sciences*, vol. 13, 2023.

[11] C. J. Watkins and P. Dayan, "Q-learning," *Mach Learn*, vol. 8, pp. 279–292, 1992.

[12] J. Peng and R. J. Williams, "Incremental Multi-Step Q-Learning," 1996.

[13] İ. Tunç and M. Söylemez, "Fuzzy logic and deep Q learning based control for traffic lights," *Alexandria Engineering Journal*, vol. 67, pp. 343–359, 2023.

[14] M. E. Çimen, Z. Garip, Y. Yalçın, M. Kutlu, and A. F. Boz, "Self Adaptive Methods for Learning Rate Parameter of Q-Learning Algorithm," *Journal of Intelligent Systems: Theory and Applications*, vol. 6, no. 2, pp. 191–198, 2023.

[15] A. O. Köroğlu, A. E. Edem, S. N. Akmeşe, Ö. Elmas, I. Tunc, and M. T. Soylemez, "Agent-Based Route Planning with Deep Q Learning," in *13th International Conference on Electrical and Electronics Engineering (ELECO)*, 2021, pp. 403–407.

[16] A. Wang, H., Emmerich, M., & Plaat, "Monte Carlo Q-learning for General Game Playing," *arXiv preprint arXiv:1802.05944*.

[17] F. Candan, S. Emir, M. Doğan, and T. Kumbasar, "Takviyeli Q-Öğrenme Yöntemiyle Labirent Problemi Çözümü Labyrinth Problem Solution with Reinforcement Q-Learning Method," in *TOK2018 Otomatik Kontrol Ulusal Toplantısı*, 2048.

[18] Y. Liu, H. Lu, S. Cheng, and Y. Shi, "An adaptive online parameter control algorithm for particle swarm optimization based on reinforcement learning," in *IEEE congress on evolutionary computation (CEC)*, 2019, pp. 815–822.

[19] M. Çimen, "Hibrit ve Kaotik Metasezgisel Arama Algoritmalari Kullanarak Model Öngörülü Kontrol Yapıları Tasarımı," Doktora, Sakarya Uygulamalı Bilimler Üniversitesi, 2022.

[20] A. F. Boz and M. E. Çimen, "An interface design for controlling dead time systems using PSO, CS and FA algorithms," in *8th International Advanced Technologies Symposium (IATS'17), 19-22 October*, 2017.

[21] A. F. Boz and M. E. Çimen, "PID Controller Design Using Improved FireFly Algorithm," in *8th International Advanced Technologies Symposium (IATS'17), 19-22 October*, 2017.

[22] M. E. Çimen and A. F. Boz, "Parameter identification of a non-minimum phase second order system with time delay using relay test and PSO, CS, FA algorithms," *Journal of the Faculty of Engineering and Architecture of Gazi University*, vol. 34, no. 1, pp. 461–477, 2019, doi: 10.17341/gazimmfd.416507.

[23] Z. B. Garip, M. E. Cimen, D. Karayel, and A. L. I. F. Boz, "The chaos-based whale optimization algorithms global optimization," vol. 0, no. 1, pp. 51–63, 2019.

[24] A. Akgül, Y. Karaca, Pala MA, M. Çimen, A. Boz, and M. Yıldız, "Chaos Theory, Advanced Metaheuristic Algorithms and Their Newfangled Deep Learning Architecture Optimization Applications: A Review," *Fractals*, vol. 32, no. 3, 2024.

[25] A. Hossain and Z. Yılmaz Acar, "Comparison of New and Old Optimization Algorithms for Traveling Salesman Problem on Small, Medium, and Large-scale Benchmark Instances," *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, vol. 13, no. 1, pp. 216–231, 2024.

[26] M. E. Cimen, Z. Garip, A. F. Boz, and D. Karayel, "Firefly Algorithm and Particle Swarm Optimization for photovoltaic parameters identification based on single model," *ISMSIT 2018 - 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies, Proceedings*, 2018, doi: 10.1109/ISMSIT.2018.8567288.

[27] A. Karthikeyan, M. E. Cimen, A. Akgul, A. F. Boz, and K. Rajagopal, "Persistence and coexistence of infinite attractors in a fractal Josephson junction resonator with unharmonic current phase relation considering feedback flux effect," *Nonlinear Dyn*, vol. 103, no. 2, pp. 1979–1998, 2021, doi: 10.1007/s11071-020-06159-4.

[28] K. Rajagopal *et al.*, "A family of circulant megastable chaotic oscillators, its application for the detection of a feeble signal and PID controller for time-delay systems by using chaotic SCA algorithm," *Chaos Solitons Fractals*, vol. 148, no. May, p. 110992, 2021, doi: 10.1016/j.chaos.2021.110992.

[29] S. A. Celtek and S. Kul, "Parameter Extraction of PV Solar Cells Using Metaheuristic Methods," *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, vol. 12, no. 4, pp. 1041–1053, 2023.

[30] M. Çimen, Z. Garip, E. M, and A. Boz, "Fuzzy Logic PID Design using Genetic Algorithm under Overshoot Constrained Conditions for Heat Exchanger Control," *Journal of the Institute of Science and Technology*, vol. 12, no. 1, pp. 164–181, 2022.

[31] H. Geçmez and H. Deveci, "Optimization of Hybrid Composite Laminates with Various Materials using the GA/GPSA Hybrid Algorithm for Maximum Dimensional Stability," *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, vol. 13, no. 1, pp. 107–133, 2024.

[32] M. E. Çimen and A. F. Boz, "PSO, CS ve FA Algoritmalarıyla Ortak Emiterli BJT'li Yükselteç Tasarımı," *Cumhuriyet Üniversitesi Fen Edebiyat Fakültesi Fen Bilimleri Dergisi*, vol. 38, no. 1, pp. 119–130, 2017.

[33] R. Eberhart and J. Kennedy, "Particle swarm optimization," in *Proceedings of the IEEE international conference on neural networks*, 1995, pp. 1942–1948.

[34] X. S. Yang, "Firefly algorithms for multimodal optimization," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5792 LNCS, pp. 169–178, 2009, doi: 10.1007/978-3-642-04944-6_14.

[35] M. Cimen and Y. Yalçın, "A novel hybrid firefly–whale optimization algorithm and its application to optimization of MPC parameters," *Soft comput*, vol. 26, no. 4, pp. 1845–1872, 2022.

[36] X. S. Yang and S. Deb, "Cuckoo search via Lévy flights," *2009 World Congress on Nature and Biologically Inspired Computing, NABIC 2009 - Proceedings*, pp. 210–214, 2009, doi: 10.1109/NABIC.2009.5393690.

[37] S. Mirjalili, S. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.

[38] M. Çimen, Z. Garip, and A. Boz, "Chaotic flower pollination algorithm based optimal PID controller design for a buck converter," *Analog Integr Circuits Signal Process*, 2021.

[39] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in engineering software*, vol. 95, pp. 51–67, 2016.

[40] M. Juneja and S. K. Nagar, "Particle swarm optimization algorithm and its parameters: A review," in *International Conference on Control, Computing, Communication and Materials (ICCCCM)*, 2016.

[41] R. C. Eberhart and Y. Shi, "Comparing inertia weights and constriction factors in particle swarm optimization," in *Proceedings of the 2000 congress on evolutionary computation. CEC00*, 2000, pp. 84–88.

[42] A. Ratnaweera, S. K. Halgamuge, and H. C. Watson, "Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients," *IEEE Transactions on evolutionary computation*, vol. 8, no. 3, pp. 240–255, 2004.

[43] H. M. Cui and Q. B. Zhu, "Convergence analysis and parameter selection in particle swarm optimization," *Jisuanji Gongcheng yu Yingyong (Computer Engineering and Applications)*, vol. 42, no. 23, pp. 89–91, 2007.

[44] C. Guimin, J. Jianyuan, and H. Qi, "Study on the strategy of decreasing inertia weight in particle swarm optimization algorithm," *Journal-Xian Jiaotong University*, vol. 40, no. 1, p. 53, 2006.

[45] Tanabe R and Fukunaga A, "Success-history based parameter adaptation for differential evolution," in *IEEE Congress on Evolutionary Computation. IEEE*, 2013, pp. 71–78.

[46] Z. Liu and T. Nishi, "Multipopulation ensemble particle swarm optimizer for engineering design problems," *Math Probl Eng*, 2020.

[47] Tatsis VA and Parsopoulos KE, "Grid-based parameter adaptation in particle swarm optimization," in *2th Metaheuristics International Conference (MIC 2017)*, 2017, pp. 217–226.

[48] F. Olivas, F. Valdez, O. Castillo, and P. Melin, "Dynamic parameter adaptation in particle swarm optimization using interval type2 fuzzy logic," *Soft Computing*, vol. 20, no. 3, pp. 1057–1070, 2016.

[49] P. Melin, F. Olivas, O. Castillo, F. Valdez, J. Soria, and M. Valdez, "Optimal design of fuzzy classification systems using pso with dynamic parameter adaptation through fuzzy logic," *Expert Syst Appl*, vol. 40, no. 8, pp. 3196–3206, 2013.

[50] S. Yin *et al.*, "Reinforcement-learning-based parameter adaptation method for particle swarm optimization," *Complex & Intelligent Systems*, vol. 9, no. 5, pp. 5585–5609, 2023.

[51] Y. Xu and D. Pi, "A reinforcement learning-based communication topology in particle swarm optimization," *Neural Comput Appl*, pp. 10007–10032, 2020.

[52] C. Lee and M. Ziegler, "Geometric approach in solving inverse kinematics of PUMA robots," *IEEE Trans Aerosp Electron Syst*, vol. 6, pp. 695–706, 1984.

[53] R. Köker, C. Öz, T. Çakar, and H. Ekiz, "A study of neural network based inverse kinematics solution for a three-joint robot," *Rob Auton Syst*, vol. 49, no. 3–4, pp. 227–234, 2004.

[54] S. Alavandar and M. J. Nigam, "Neuro-fuzzy based approach for inverse kinematics solution of industrial robot manipulators," *International Journal of Computers Communications & Contro*, vol. 3, no. 3, pp. 224–234, 2008.

[55] G. Jin, S. Ma, and Z. Li, "Dynamic simulation modeling of industrial robot kinematics in industry 4.0," *Discrete Dyn Nat Soc*, pp. 1–11, 2022.

[56] Y. Chen, X. Zhang, Y. Huang, Y. Wu, and J. Ota, "Kinematics optimization of a novel 7-DOF redundant manipulator," *Rob Auton Syst*, vol. 163, p. 104377, 2023.

[57] S. Baressi Šegota, N. Anđelić, I. Lorencin, M. Saga, and Z. Car, "Path planning optimization of six-degree-of-freedom robotic manipulators using evolutionary algorithms," *International journal of advanced robotic systems*, vol. 17, no. 2, 2020.

[58] Y. Hou and J. Li, "Learning 6-DoF grasping with dual-agent deep reinforcement learning," *Rob Auton Syst*, vol. 166, 2023.

[59] S. Müftü and B. Gökçe, "Design and Implementation of an Optimized PID Controller for Two-Limb Robot Arm Control," *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, vol. 13, no. 1, pp. 192–204, 2024.

[60] M. Çimen and A. Boz, "Parameter identification of a non-minimum phase second order system with time delay using relay test and PSO, CS, FA algorithms," *Journal of the Faculty of Engineering and Architecture of Gazi University*, vol. 34, no. 1, pp. 461–477, 2019.

[61] A. Angiuli, J. P. Fouque, and M. Laurière, "Unified reinforcement Q-learning for mean field game and control problems," *Mathematics of Control, Signals, and Systems*, vol. 34, no. 2, pp. 217–271, 2022.

[62] C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D., Dissertation, King's College UK, 1989.

[63] A. S. Posamentier and I. Lehmann, *The Glorious Golden Ratio*. Prometheus Books, 2011.

[64] S. Rao, *Engineering optimization: theory and practice*. John Wiley & Sons, 2019.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 969-978 YEAR: 2024
DOI: 10.17798/bitlisfen.1485060

NiO Nanoparticles Via Calcination of Dithiocarbamate Precursors: Characterization and Photocatalytic Activity

Ayşegül ŞENOCAK[1*], Hüseyin AKBAŞ[1], Beyza İŞGÖR[1]

[1]*Tokat Gaziosmanpaşa University, College of Art and Science, Department of Chemistry, Tokat, Türkiye*
ORCID: (0000-0001-9210-4621) (ORCID: 0000-0002-3013-9033) (ORCID: 0009-0000-9530-8750)

Keywords: Dithiocarbamates, NiO nanoparticles, Calcination, Photocatalytic activity, Methylene blue

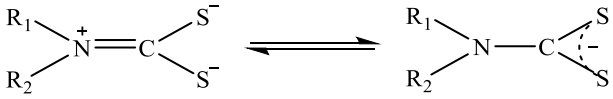
Abstract

By this work, the preparation of NiO nanoparticles (NPs) from a dithiocarbamate complex through calcination, both a cost-effective and practical way for NPs synthesis, was presented. This preparation process involves the decomposition of the precursor to yield highly pure and crystalline NPs. The dithiocarbamate complex was synthesized in a sequential reaction starting from p-anisaldehyde and 3-aminopropanol. Particle size of the NPs, whose formation was confirmed by FT-IR spectroscopy and UV-Vis spectroscopy, were calculated as 10.97 nm using Debye-Scherer equation from X-ray Diffraction Spectroscopy (XRD). In addition, Scanning Electron Microscopy (SEM) revealed generally aggregated and spherical particles, which are stable in colloidal environment according to -15.1 zeta potential value. Besides, the NPs were employed as a photocatalyst for the decomposition of Methylene Blue (MB), one of the most frequently used synthetic dyes in industry and, at the same time, one of the primary water pollutants, resulting in 38% degradation rate.

1. Introduction

Dithiocarbamates, represented by the general formula S_2CNR_2, are a remarkable group of compounds, distinguished by their unique sulfur-sulfur bonds and a variety of intriguing properties. From their origins in the late 19th century to their contemporary members in diverse applications, dithiocarbamates have left an indelible mark on modern chemistry. Dithiocarbamate compounds possess an exceptional ability to easily capture a broad range of metal ions, allowing for the construction of complex architectures, which have brought in great importance to them in coordination chemistry. Synthesis of the dithiocarbamate complexes typically involves the preparation of the dithiocarbamate ligand through the reaction of amine compounds with CS_2 in a basic medium, followed by its coordination with a metal ion [1], [2], [3]. This process allows for convenient regulation of the electrical and structural characteristics of the final dithiocarbamate compound by manipulating the pioneer amine.

The resonance structures of dithiocarbamates allow them to readily coordinate with a wide range of metal ions (Figure 1). The thioureide form functions as a weak-field ligand, providing stability to high-valent metals. On the other hand, the dithiocarbamate form tends to capture low-valent metals as a strong-field ligand. Among the various coordination modes available to dithiocarbamates, the most common is the bidentate chelate, which enhances the stability of the resulting structure [4].



Thioureide Dithiocarbamate

Figure 1. The resonance forms of dithiocarbamates.

*Corresponding author: aysegul.senocak@gop.edu.tr
Received: 16.05.2024, Accepted: 17.12.2024

Due to their strong affinity for metals, numerous dithiocarbamate complexes, demonstrating a broad range of applications including high-pressure lubricants and accelerators in vulcanization [5], anion sensors [6], sensitizers in solar schemes [7], pesticides and fungicides [8], [9], as well as anticancer, antibacterial, and antioxidant agents [10], [11], [12], have been synthesized. Nickel dithiocarbamate complexes, in particular, have been extensively examined for their biological activities, including anticancer, antimicrobial, or antioxidant [13], [14], [15], [16], [17]. Furthermore, numerous studies have focused on the utilization of dithiocarbamate complexes as single-source precursors for nickel NPs [18], [19], [20], [21], [22], [23]. Various methods were employed for nickel NPs production; for instance, in some of these studies, NiO NPs were generated through high-temperature calcination, while in the others, the solvothermal method was employed by adding ethylenediamine for nickel sulfide NPs synthesis [18].

NPs composed of NiO exhibit diverse applications across multiple disciplines, including energy storage [24], rechargeable batteries [25], catalysis [26], bio-active materials [27], and sensors [28]. In recent years, the catalytic activities of NiO NPs have gained great importance, particularly in environmental applications, due to the escalating pollution caused by rapid industrialization and uncontrolled population growth. One of the catalytic applications of NiO NPs pertains to degrading synthetic dyes present in water contaminated by industrial effluents. Numerous methods which can be classified as biological, chemical and physical have been used for remediation of contaminated waters with dyes. Among these methods, photocatalysis has garnered significant attention as it is the most effective and environmentally friendly approach to date [29]. Research in this area revealed that NiO NPs have exhibited successful results as photocatalysts in the breaking down industrial dyes such as MB, Rhodamine B, and Brown HT [30], [31], [32]. The degradation process occurs through the generation of O_2 and OH radicals within the positively charged cavities on the surface of NPs by means of electron excitation and the subsequent interaction of these radicals with synthetic dyes, leading to the decomposition of these dyes [33].

In this study, it was aimed to transform one example of the widely studied dithiocarbamate complexes into nickel NPs, which are utilized in addressing the nowadays pressing issue of environmental pollution. For this aim, a nickel dithiocarbamate complex was prepared through a chain reaction starting with 4-methoxybenzaldehyde and 3-aminopropanol. The resulting complex was then thermally decomposed to produce NiO NPs, which were characterized by using FT-IR, UV-Vis, SEM-EDS, XRD, and Zeta Potential methods. Besides, the photocatalytic capacity of the NPs over the decomposition of MB was evaluated.

2. Material and Method

4-methoxybenzaldehyde (98%), 3-aminopropanol (99%), sodium borohydride (98%), potasium hydroxide (99%), carbon disulfide (99%), nickel(II) sulfate hexahydrate (98%), methylene blue (82%), dicholoromethan, and ethanol were purchased and used without any additional purification procedure. The FT-IR spectra were measured between 400-4000 cm^{-1} range using a Jasco FTIR 4700 spectrometer. By employing a Perkin-Elmer Lambda-35 UV-Vis spectrophotometer, which operates in the range of 200–800 nm, the UV-Vis spectra of the nickel complex and NPs in ethanol (10^{-3} M) were recorded. The XRD pattern was recorded using a PANalytical X-Pert3 Powder X-ray diffractometer with Cu-Kα radiation. The morphological characters of the NPs were investigated using a TESCAN™ MIRA3 XMU SEM. In order to determine the zeta potential of the NPs dispersed in ultra-pure water, the Malvern/Zetasizer Nano ZSP utilized the diffusion light scattering approach.

2.1. Preparation of the Nickel Dithiocarbamate Complex

4-methoxybenzaldehyde and 3-aminopropanol in equimolar amounts were refluxed in methanol for 4 days. Once the transformation into the Schiff base was complete, excess sodium borohydride was added to the reaction mixture in small portions. After the reduction was finished, the methanol was evaporated from the crude product. The resulting oily product was washed twice with a CH_2Cl_2/water mixture and dried over Na_2SO_4, heated in a muffle furnace at 250 ^oC.

The final reduced amine in light yellow, oily form was dissolved in ethanol and cooled to 0 ^oC. To this solution, KOH and CS_2 in ethanol were added dropwise, respectively. After stirring for a day at room temperature, the appearance of the yellow solution indicated the dithiocarbamate formation. Subsequently, an aqueous solution of $NiSO_4.6H_2O$ (0.5 equivalent) was added to form

the complex structure. A green precipitate formed immediately was filtered after stirring for one day, and the residue on the filter paper was allowed to dry and was then isolated.

2.2. Preparation of the NiO NPs

The resulting complex was placed in porcelain crucibles and calcined in a muffle furnace at 500 ºC for 4 hours. Distilled water was added to the resulting substance and mixed at 80 ºC for 20 minutes. Then, the mixture was transferred to centrifuge tubes and centrifuged for 3 minutes. The decanted NPs were centrifuged further 3 minutes after being washed with ethanol. The obtained NiO NPs were left to dry at room temperature and isolated.

2.3. Photocatalytic Performance

A batch process was employed to study the photocatalytic activity and decomposition kinetics of the NiO NPs in the degradation of MB. With this aim, NiO NPs (20 mg), predried at 100 ºC for 2 hours, were added to 10 ppm MB solution (50 mL), and this mixture was stirred magnetically in the dark for an hour to reach thermodynamic equilibrium. Then, the reaction vessel was opened to light and allowed to mix under a daylight lamp (10 watt) positioned 30 cm above the reaction surface. The UV spectrum of the resulting solution was measured by withdrawing 3 mix under a daylight lamp (10 watt) positioned 30 cm above the reaction surface. The UV spectrum of the resulting solution was measured by withdrawing 3 mL samples every hour. The results were interpreted by calculating the amount of MB remaining in the solution by using Eq. 1 where A_0 an A_t represent the initial absorbtion and absorbtion at time *t*, respectively.

$$\boldsymbol{Dye\ Removal} = \left({}^{(\boldsymbol{A_0} - \boldsymbol{A_t})}/_{\boldsymbol{A_0}}\right) * \mathbf{100} \qquad (1)$$

3. Results and Discussion

In this research, a nickel dithiocarbamate complex was obtained by sequential reactions starting with 4-methoxybenzaldehyde and 3-aminopropanol, and the final complex was then subjected to calcination to produce NiO NPs (Scheme 1). A color transformation from green to gray was observed as a consequence of the calcination process (Figure 2). The structure of the resulting NiO NPs was illuminated by a number of spectroscopic techniques, involving FT-IR, UV-Vis, XRD, SEM-EDX, and zeta potential.

Figure 2. Powder forms of the dithiocarbamate complex (green) and the NiO NPs (gray).

As the first step of characterization, the FT-IR and UV-Vis spectra of the dithiocarbamate complex and the NiO NPs were measured and compared to assess the appearance and purity of the NPs. The FT-IR spectra of the dithiocarbamate complex and the final NPs were depicted in Figure 3. One of the key features in the spectrum was the coordinated O-H stretching observed between 3671-3007 cm^{-1}. The splitting of the C-S band (981 and 961 cm^{-1}) gave rise to the hypothesis that the dithiocarbamate group coordinated with the Ni(II) ions in a monodentate manner [34]. Besides, the band at 1501 cm^{-1} emphasized the partial double bond character of the C-N bond, while Ni-S bond formation was confirmed by the peak at 508 cm^{-1} [7]. Considering all these provided spectroscopic evidence, the structural formula for the nickel dithiocarbamate complex was proposed as shown in Scheme 1. The disappearance of all characteristic bands of the pioneer complex in the FT-IR spectrum emphasized the complete conversion of the nickel complex into NPs as a result of calcination. Furthermore, the band at 406 cm^{-1} provided evidence of the existence of NiO NPs in accordance with previous studies [35], [36], [37].

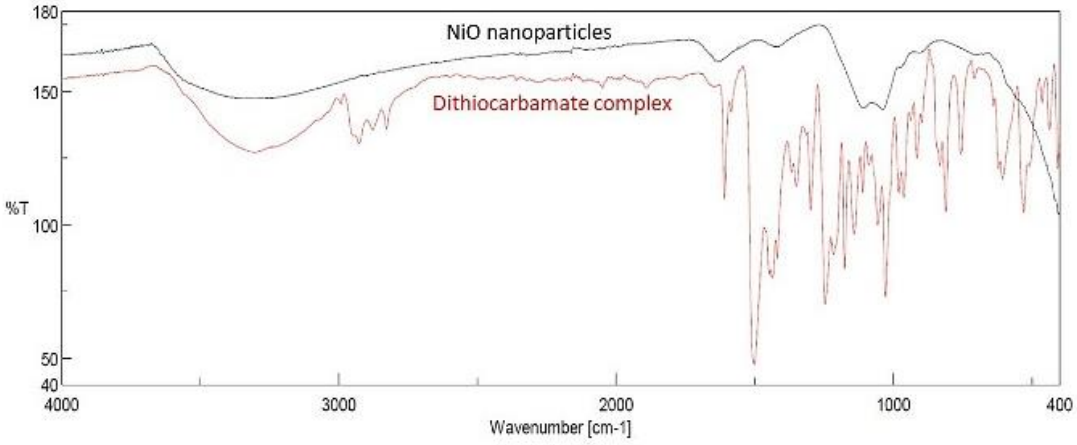
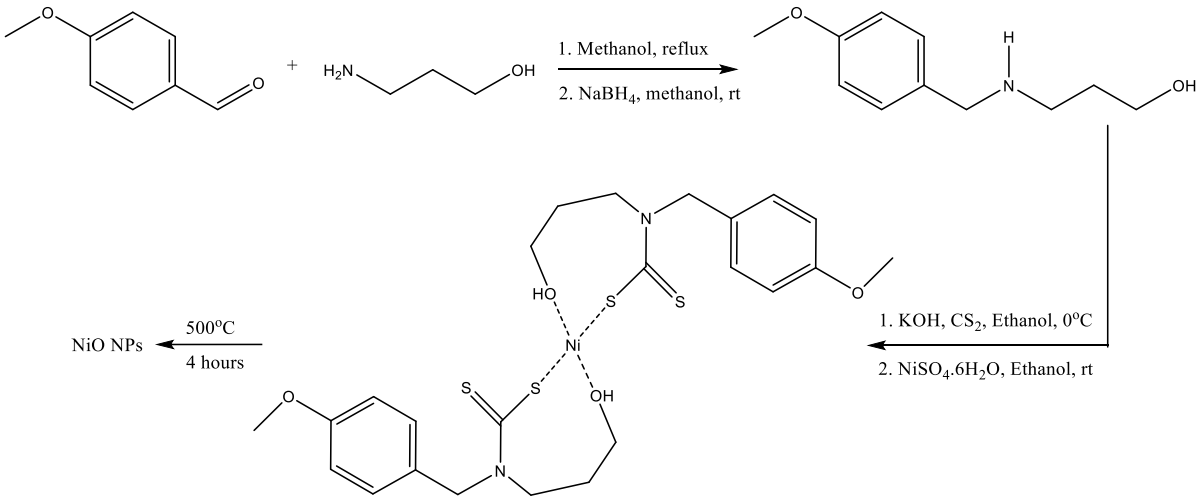


Figure 3. FT-IR spectra of the dithiocarbamate complex and NPs.



Scheme 1. The reaction scheme and the proposed complex structure for the production of NiO NPs.

The UV-Vis spectra of the nickel complex and the NiO NPs dispersed in ethanol by an ultrasonic bath were depicted in Figure 4.

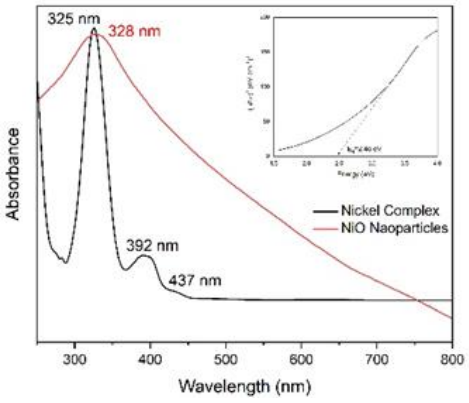


Figure 4. UV-Vis spectra of the complex and the NPs (with $(Ah\upsilon)^2$–$h\upsilon$ plot (inset)).

Two absorbtion maxima at 325 and 392 nm were assigned to metal-ligand charge transfer for square planar complexes. Besides, a d-d transition in the metal center was observed as a shoulder at 437 nm [38]. The UV maxima at 328 nm pointed out the presence of the NiO NPs in line with their counterparts in the previous studies [39], [40], [41]. Furthermore, to calculate the optical band gap of the NPs, a Tauc plot was included in the UV-Vis spectrum. The band gap energy was determined as 2.46 eV from the graph drawn by employing the $(Ah\upsilon)^2 = B(h\upsilon–Eg)$ equation, where A is the absorption coefficient and B is a constant. The band gap energy, an indicator for the ease of electron jump from the valence band to the conduction band, is compatible with the literature [42], [43].

The XRD spectrum of the NPs exhibited diffraction peaks at 37.43 [1 1 1], 43.35 [2 0 0], 63.01 [2 2 0], 75.61 [3 1 1], 79.59 [2 2 2] (Figure 5). XRD analysis confirmed that the nanoparticle crystallized in a Bunsenite structure, identified with COD 98-002-8834, which corresponds to a cubic structure with space group Fm-3m. Besides, the density of the NiO NPs was calculated as 6.84

g/cm^3 based on the collected XRD data. Any impurities originated from the starting compound were not exhibited in the XRD pattern. Besides, the sharpness and intensities of the diffraction peaks pointed out the crystallinity of the NPs. The average radius of the obtained NPs was calculated as 10.97 nm by the Scherer equation, where λ is the wavelength of Cu-Kα, β is full width at the half max of the main peak, and θ is Bragg's angle (Equation 2).

$$\boldsymbol{D = (0.9 * \lambda)/(\beta * cos\theta)} \quad (2)$$

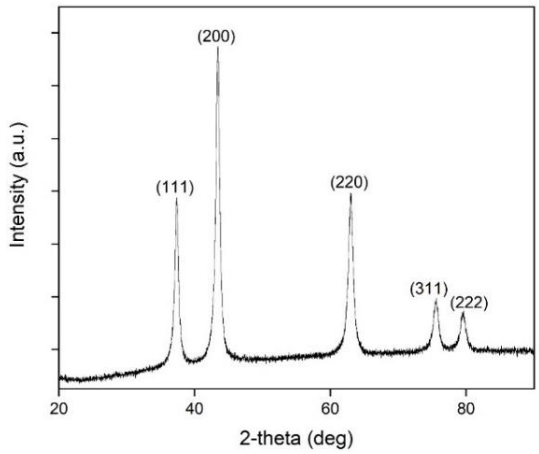


Figure 5. X-ray diffraction pattern for the synthesized NPs.

The images of the NiO NPs captured by the scanning electron microscope were exhibited in Figure 6. The particles were observed to have a generally spherical shape with an average size of 61.14 nm. It was also observed that NiO NPs aggregated as expected since no anti-collapse agent was used to minimize the interaction between the particles.

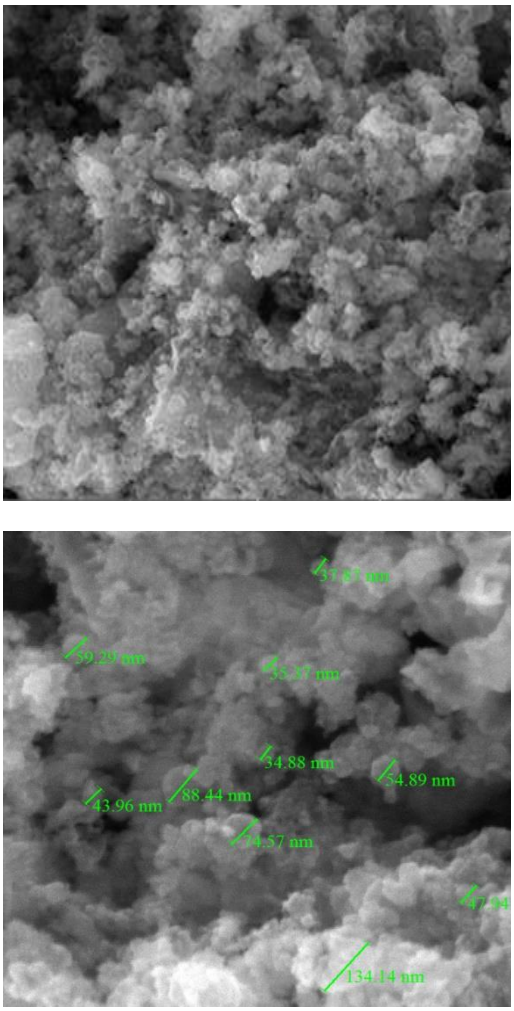
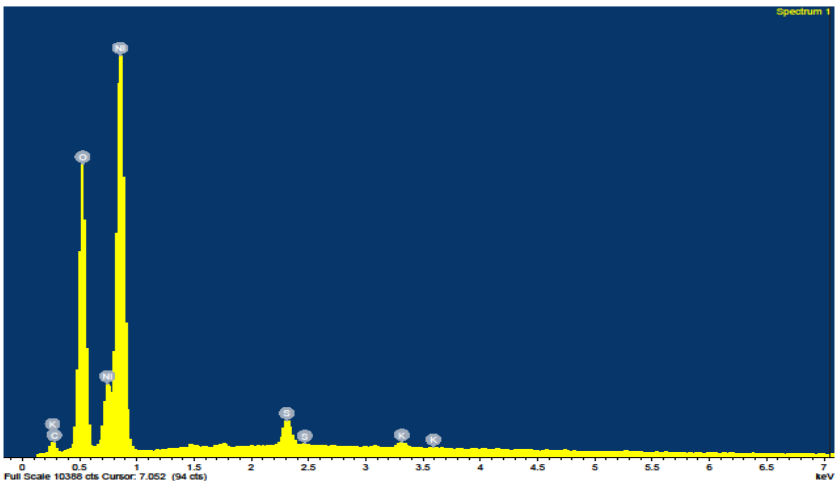


Figure 6. SEM micrograms of the NiO NPs.

EDS analysis revealed that the synthesized NPs were composed of 68.93% Ni and 25.21% O, which closely align with the theoretical ratios (78% Ni, 21% O) for the NiO general formula (Figure 7). The other elements (C, S, and K) observed in negligible amounts in the analysis results were attributed to instrumental impurities.



Element	Weight%	Atomic%
C K	3.97	10.53
O K	25.21	50.23
S K	1.50	1.49
K K	0.39	0.32
Ni K	68.93	37.43
Totals	100.00	

Figure 7. EDS spectrum of the NPs.

Zeta potential is a widely used criterion for interpreting the stability of NPs by providing information regarding the repulsion between similarly charged NPs in a colloidal environment. The more negative zeta potential value means greater stability of the NPs [44]. The -15.1 zeta potential value found in this study indicates that the prepared NiO NPs are stable in the colloidal environment (Figure 8).

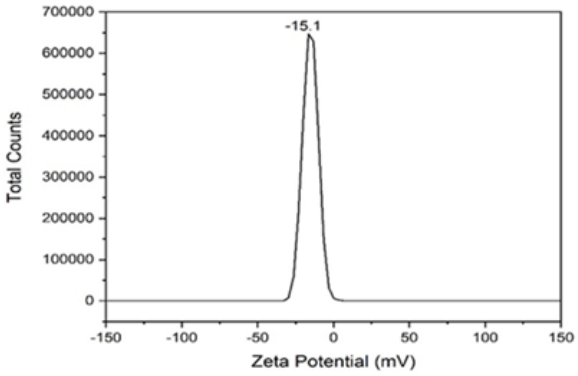


Figure 8. The zeta potential distribution of the NPs.

The photocatalytic activity of the characterized NiO NPs was evaluated for the decomposition of MB synthetic dye. The usage of NiO NPs in catalytic amounts in the presence of light from a daylight lamp ended up decomposition of 38% of MB over six hours (Figure 9). In addition, almost identical "before catalyst" and "0 min" spectra revealed that both light and the catalyst are essential for MB degradation. This moderate activity of the NPs emerged from the formation of positive holes based on electron motion between the energy bands of the NiO NPs by reason of exposure to sunlight and the transformation of methylene blue into small harmless molecules via the radicals forming in these cavities.

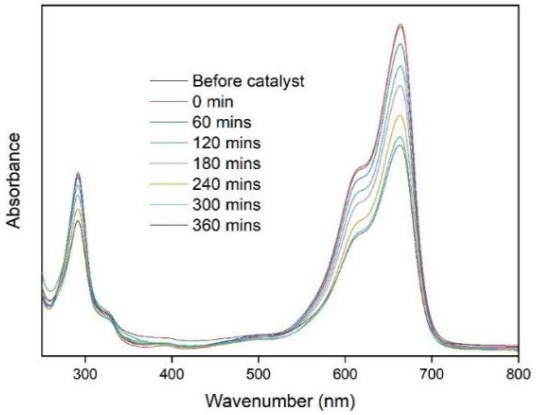


Figure 9. The NiO catalyzed photocatalytic degradation of MB over time.

By using Eq. 3, where C_t, C_0, and t represented the concentration at time t, the initial concentration, and time, the pseudo-first-order decomposition kinetics were examined for MB. The rate constant for the decomposition was calculated as 0.0014 min^{-1} (R^2= 0.982) which is lower than those in the previous study (Fig 10) [45].

$$ln\,{}^{C_t}/_{C_0} = -k.t \quad (3)$$

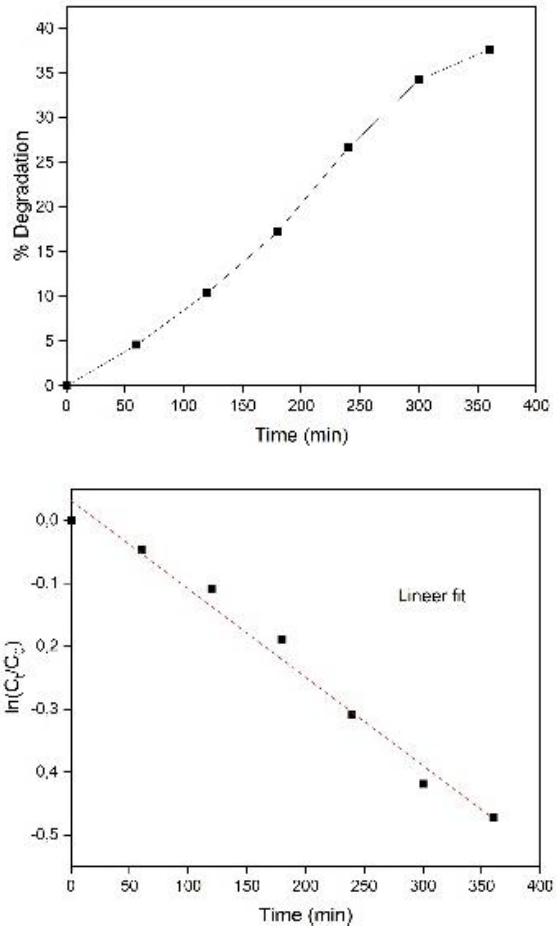


Figure 10. (a) The NiO catalyzed percentage degradation of MB **(b)** Pseudo-first-order kinetics for the MB decomposition.

4. Conclusion

In this study, NiO NPs were produced using the thermal decomposition (calcination) method from a nickel complex synthesized through a series of chemical reactions. The characterization of the synthesized NPs was carried out by FT-IR, UV-Vis, SEM, EDS, XRD, and Zeta Potential analyses. The values of 405 cm^{-1} and 328 nm observed in the FT-IR and UV-Vis spectra, respectively, are compatible with NiO NPs synthesized in the literature. Besides, XRD and EDS spectra further

confirmed the formation of NiO NPs. By utilizing the XRD spectrum, the dimensions of the synthesized particles were calculated as 10.97 nm and, the sharpness of the XRD peaks was referred to the high crystallinity of the NPs. The SEM micrograph revealed non-uniform shaped and highly aggregated particles, which have an average size of 61.14 nm. The particle size from SEM micrograph was observed greater from those calculated by Debye-Scherer equation. The reason for this discrepancy may spring from the fact that XRD analysis determines the average size of the crystalline domains, whereas SEM spectroscopy measures the size of the individual particles. Besides, SEM spectroscopy can detect agglomerates, which can lead to an overestimation of the grain size. In contrast, the Scherer equation considers a perfect crystalline structure not agglomerates. The zeta potential value of -15.1 indicated that the synthesized NPs remained stable in the colloidal environment. Finally, the catalytic activity of the NiO NPs was evaluated for the degradation of MB under daylight lamp where the NPs exhibited moderate (38 %) photocatalytic activity.

Contributions of the authors

Ayşegül Şenocak contributed to the experimental studies, data interpretation, and the preparation of the manuscript.
Hüseyin Akbaş contributed to the experimental studies and the preparation of the manuscript.
Beyza İşgör contributed to the experimental studies.

Acknowledgment

The authors would like to thank to TUBİTAK (the Scientific and Technological Research Council of Türkiye) for supporting this study through the 2209-A Research Project Support Programme for Undergraduate Students (Project Number: 1919B012206215).

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] S. D. Oladipo, B. Omondi, and C. Mocktar, “Synthesis and structural studies of nickel(II)- and copper(II)-N,N′-diarylformamidine dithiocarbamate complexes as antimicrobial and antioxidant agents”, *Polyhedron*, vol. 170, pp. 712-722, 2019.

[2] G. Hogarth, “Metal-dithiocarbamate complexes: chemistry and biological activity”, *Mini-Rev. Med. Chem.*, vol. 12, no. 12, pp. 1202-1215, 2012.

[3] M. Shahid, T. Rüffer, H. Lang, S. A. Awan, and S. Ahmad, “Synthesis and crystal structure of a dinuclear zinc(II)-dithiocarbamate complex, bis{[(μ^2-pyrrolidinedithiocar-bamato-S,S’)(pyrrolidinedithio- carbamato-S,S’)zinc(II)]}”, *J. Coord. Chem.*, vol. 62, no. 3, pp. 440-445, 2009.

[4] J. C. Sarker, and G. Hogarth, “Dithiocarbamate Complexes as Single Source Precursors to Nanoscale Binary, Ternary and Quaternary Metal Sulfides”, *Chem. Rev.*, vol. 121, no. 10, pp. 6057-6123, 2021.

[5] T. A. Saiyed, J. O. Adeyemi, and D. C. Onwudi, “The structural chemistry of zinc(II) and nickel(II) dithiocarbamate complexes”, *Open Chem.*, vol. 19, pp. 974-986, 2021.

[6] P. D. Beer, A. G. Cheetham, M. G. B. Drew, O. D. Fox, E. J. Hayes, and T. D. Rolls, “Pyrrole-based metallo-macrocycles and cryptands”, *Dalton T.*, pp. 603-61, 2003.

[7] E. Sathiyaraj, G. Gurumoorthy, and S. Thirumaran, “Nickel(II) dithiocarbamate complexes containing the pyrrole moiety for sensing anions and synthesis of nickel sulfide and nickel oxide nanoparticles”, *New J. Chem.*, vol. 39, pp. 5336-5349, 2015.

[8] P. Ferreira, G. M. Lima, E. B. Paniago, J. A. Takahashi, and C. B. Pinheiro, "Synthesis, characterization and antifungal activity of new dithiocarbamate-based complexes of Ni(II), Pd(II) and Pt(II)", *Inorg. Chim. Acta*, vol. 423(A), pp. 443-449, 2014.

[9] V. A. Fitsanakis, V. Amarnath, J. T. Moore, K. S. Montine, J. Zhang, and T. J. Montine, "Catalysis of catechol oxidation by metal-dithiocarbamate complexes in pesticides", *Free Radical Bio. and Med*., vol. 33, no. 12, pp. 1714-1723, 2002.

[10] D. C. Onwudiwe, and A. C. Ekennia, "Synthesis, characterization, thermal, antimicrobial and antioxidant studies of some transition metal dithiocarbamates", *Res. Chem. Intermediat*., vol. 43, pp.1465-1485, 2017.

[11] S. M. Mambaa, A. K. Mishraa, B. B. Mambaa, P. B. Njobehb, M. F. Duttonb, and E. Fosso-Kankeu, "Spectral, thermal and in vitro antimicrobial studies of cyclohexylamine-N-dithiocarbamate transition metal complexes", *Spectrochim. Acta A*, vol. 77, pp. 579-587, 2010.

[12] I. Ali, W. A. Wani, K. Saleem, and M. -F. Hseih, "Design and synthesis of thalidomide based dithiocarbamate Cu(II), Ni(II) and Ru(III) complexes as anticancer agents", *Polyhedron*, vol. 56, pp. 134-143, 2013.

[13] S. Balakrishnan, S. Duraisamy, M. Kasi, S. Kandasamy, R. Sarkar, and A. Kumarasamy, "Syntheses, physicochemical characterization, antibacterial studies on potassium morpholine dithiocarbamate nickel (II), copper (II) metal complexes and their ligands", *Heliyon*, vol. 5, no. 5, pp. e01687, 2019.

[14] A. C. Ekennia, D. C. Onwudiwe, and A. A. Osowole, "Spectral, thermal stability and antibacterial studies of copper, nickel and cobalt complexes of N-methyl-N-phenyl dithiocarbamate", *J. Sulfur Chem*., vol. 36, no. 1, pp. 96-104, 2015.

[15] A. Pastrana-Dávila, A. Amaya-Flórez, C. Aranaga, J. Ellena, M. Macías, E. Flórez-López, and R. F. D'Vries, "Synthesis, characterization, and antibacterial activity of dibenzildithiocarbamate derivates and Ni(II)–Cu(II) coordination compounds", *J. Molec. Struct*., vol. 1245, pp. 131109, 2021.

[16] A. Singh, L. B. Prasad, K. Shiv, R. Kumar, and S. Garai, "Synthesis, characterization, and in vitro antibacterial and cytotoxic study of Co(II), Ni(II), Cu(II), and Zn(II) complexes of N-(4-methoxybenzyl) N-(phenylethyl) dithiocarbamate ligand", *J. Molec. Struct*., vol. 1288, pp. 135835, 2023.

[17] M. M. Hrubaru, E. Bartha, A. C. Ekennia, S. N. Okafor, C. D. Badiceanu, D. A. Udu, D. C. Onwudiwe, S. Shova, and C. Draghici, "Ni(II), Pd(II) and Pt(II) complexes of N,N-bis(3,3-dimethyl-allyl)-dithiocarbamate: Synthesis, spectroscopic characterization, antimicrobial and molecular docking studies", *J. Molec. Struct*., vol. 1250, no. 1, pp. 131649, 2022.

[18] A. Nqombolo, and P. A. Ajibade, "Synthesis and Spectral Studies of Ni(II) Dithiocarbamate Complexes and Their Use as Precursors for Nickel Sulphides Nanocrystals", *J. Chem.-NY*, vol. 2016, pp. 1293790, 2016.

[19] D. C. Onwudiwe, N. H. Seheri, L. Hlungwani, H. Ferjani, and R. Rikhotso-Mbungela, "NiO nanoparticles by thermal decomposition of complex and evaluation of the structural, morphological, and optical properties", *J. Mol. Struct*., vol 1317, pp. 139084, 2024.

[20] V. K. Maurya, L. B. Prasad, A. Singh, K. Shiv, and A. Prasad, "Synthesis, spectroscopic characterization, biological activity, and conducting properties of functionalized Ni(II) dithiocarbamate complexes with solvent extraction studies of the ligands" *J. Sulphur Chem.*, vol 44 no.3, pp 336–353, 2022.

[21] E. Sathiyaraj, S. Thirumaran, S. Ciattini, and S. Selvanayagam, "Synthesis and characterization of Ni(II) complexes with functionalized dithiocarbamates: New single source precursors for nickel sulfide and nickel-iron sulfide nanoparticles", *Inorg. Chim. Acta*, vol. 498, pp. 119162, 2019.

[22] F. F. Bobinihi, O. E. Fayemi, and D. C. Onwudiwe, "Synthesis, characterization, and cyclic voltammetry of nickel sulphide and nickel oxide nanoparticles obtained from Ni(II) dithiocarbamate", *Mater. Sci. Semicond. Process*., vol. 121, pp. 105315, 2021.

[23] P. Lakshmanan, R. Arulmozhi, S. Thirumaran, and S. Ciattini, "Ni(II) dithiocarbamate: Synthesis, crystal structures, DFT studies and applications as precursors for nickel sulfide and nickel oxide nanoparticles", *Polyhedron*, vol. 218, pp. 115766, 2022.

[24] H. Wu, Y. Wang, C. Zheng, J. Zhu, G. Wu, and X. Li, "Multi-shelled NiO hollow spheres: Easy hydrothermal synthesis and lithium storage performances", *J. Alloy. Comp*., vol. 685, pp. 8-14, 2016.

[25] K. N. Patel, M. P. Deshpande, V. P. Gujarati, S. Pandya, V. Sathe, and S. H. Chaki, "Structural and optical analysis of Fe doped NiO nanoparticles synthesized by chemical precipitation route", *Mater. Res. Bull*., vol. 106, pp. 187-196, 2018.

[26] G. Grivani, M. Vakili, A. D. Khalaji, G. Bruno, H. A. Rudbari, M. Taghavi, and V. Tahmasebi, "Synthesis, characterization, crystal structure determination, computational study, and thermal decomposition into NiO nano-particles of a new $Ni^{II}L_2$ Schiff base complex (*L* = 2-{(E)-[2-chloroethyl)imino] methyl phenolate)", *J. Molec. Struct*., vol. 1072, pp. 77-83, 2014.

[27] K. Karthik, S. Dhanuskodi, C. Gobinath, S. Prabukumar, and S. Sivaramakrishnan, "Multifunctional properties of microwave assisted CdO–NiO–ZnO mixed metal oxide nanocomposite: enhanced photocatalytic and antibacterial activities", *J. Mater. Sci.-Mater. El.*, vol. 29, pp. 5459-5471, 2018.

[28] A. Salehirad, "Synthesis of high surface area NiO nanoparticles through thermal decomposition of mixed ligand Ni(II) Complex, [Ni(binol)(bpy)].CH_3OH", *Russ. J. Appl. Chem*., vol. 89, pp. 63-69, 2016.

[29] H. Kumari, Sonia, Suman, R. Ranga, S. Chahal, S. Devi, S. Sharma, S. Kumar, P. Kumar, S. Kumar, A. Kumar, and R. Parmar, "A Review on Photocatalysis Used For Wastewater Treatment: Dye Degradation", *Water Air Soil Pollut.*, vol. 234, no.6, pp. 349, 2023.

[30] M. Y. Nassar, H. M. Aly, E. A. Abdelrahman, and M. E. Moustafa, "Synthesis, characterization, and biological activity of some novel Schiff bases and their Co(II) and Ni(II) complexes: A new route for Co3O4 and NiO nanoparticles for photocatalytic degradation of methylene blue dye", *J. Mol. Struct.*, vol. 1143, pp. 462-471, 2017.

[31] Z. Sabouri, A. Akbari, H. A. Hosseini, M. Khatami, and M. Darroudi, "Egg white-mediated green synthesis of NiO nanoparticles and study of their cytotoxicity and photocatalytic activity", *Polyhedron*, vol. 178, pp. 114351, 2020.

[32] M. Darbandi, M. Eynollahi, N. Badri, M. F. Mohajer, and Z. A. Li, "NiO nanoparticles with superior sonophotocatalytic performance in organic pollutant degradation", *J. Alloy. Comp*., vol. 889, pp. 161706, 2021.

[33] Z. Sabouri, A. Akbari, H. A. Hosseini, A. Hashemzadeh, and M. Darroudi, "Bio-based synthesized NiO nanoparticles and evaluation of their cellular toxicity and wastewater treatment effects", *J. Molec. Struct.*, vol. 1191, pp. 101-109, 2019.

[34] E. Sathiyaraj, P. Valarmathi, S. Thirumaran, S. Ciattini, V. K. Gupta, and R. Kant, "Effect of N-Bound Organic Moiety in Dithiocarbamate (R2NCS−2) and trans Influence of Triphenylphosphine on NiS2PN Chromophore", *Phosphorus, Sulfur Relat. Elem*., vol. 190, no. 7, pp. 1127-1137, 2015.

[35] M. Abdur Rahman, R. Radhakrishnan, and R. Gopalakrishnan, "Structural, optical, magnetic and antibacterial properties of Nd doped NiO nanoparticles prepared by co-precipitation method", *J. Alloy. Comp.*, vol. 742, pp. 421-429, 2018.

[36] M. A. J. Kouhbanani, Y. Sadeghipour, M. Sarani, E. Sefidgar, S. Ilkhani, A. M. Amani, and N. Beheshtkhoo, "The inhibitory role of synthesized Nickel oxide nanoparticles against Hep-G2, MCF-7, and HT-29 cell lines: the inhibitory role of NiO NPs against Hep-G2, MCF-7, and HT-29 cell lines", *Green Chem. Lett. Rev.*, vol. 14, no. 3, pp. 444-454, 2021.

[37] E. Gobinath, M. Dhatchinamoorthy, P. Saran, D. Vishnu, R. Indumathy, and G. Kalaiarasi, "Synthesis and characterization of NiO nanoparticles using Sesbania grandiflora flower to evaluate cytotoxicity", *Results Chem*., vol. 6, pp. 101043, 2023.

[38] F. F. Bobinihi, D. C. Onwudiwe, and E. C. Hosten, "Synthesis and characterization of homoleptic group 10 dithiocarbamate complexes and heteroleptic Ni(II) complexes, and the use of the homoleptic Ni(II) for the preparation of nickel sulphide nanoparticles", *J. Molec. Struct*., vol. 1164, pp. 475-485, 2018.

[39] A. A. Barzinjy, S. M. Hamad, S. Aydın, M. H. Ahmed, and F. H. S. Hussain, "Green and eco-friendly synthesis of Nickel oxide nanoparticles and its photocatalytic activity for methyl orange degradation", *J. Mater. Sci.-Mater. El.*, vol. 31, pp. 11303-11316, 2020.

[40] R. S. Kumar, S. J. Jeyakumar, M. Jothibas, I. K. Punithavathy, and J. P. Richard, "Influence of molar concentration on structural, optical and magnetic properties of NiO nanoparticles", J. Mater. Sci.-Mater. El., vol. 28, pp. 15668-15675, 2017.

[41] M. El-Kemary, N. Nagy, and I. El-Mehasseb, "Nickel oxide nanoparticles: Synthesis and spectral studies of interactions with glucose", *Mater. Sci. Semicond. Process.*, vol. 16, pp. 1747-1752, 2013.

[42] A. K. Ramasami, M. V. Reddy, and G. R. Balakrishna, "Combustion synthesis and characterization of NiO nanoparticles", *Mater. Sci. Semicond. Process.*, vol. 40, pp. 194-202, 2015.

[43] F. Ascencio, A. Bobadilla, and R. Escudero, "Study of NiO nanoparticles, structural and magnetic characteristics", *Appl. Phys. A*, vol. 125, pp. 279, 2019.

[44] A. B. García, A. Cuesta, M. A. Montes-Morán, A. Martínez-Alonso, and J. M. D. Tascón, "Zeta Potential as a Tool to Characterize Plasma Oxidation of Carbon Fibers", *J. Colloid. Interf. Sci*., vol. 192, pp. 363-367, 1997.

[45] K. Motene, L. Mahlaule-Glory, N. Ngoepe, M. Mathipa, and N. Hintsho-Mbita, "Photocatalytic degradation of dyes and removal of bacteria using biosynthesised flowerlike NiO nanoparticles", *Int. J. Environ. Anal. Chem.*, vol. 103, no. 5, pp. 1107-1122, 2021.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 979-987 YEAR: 2024
DOI: 10.17798/bitlisfen.1487072

Synthesis and Characterization of Several Mannich Bases Derived from 2-(4-methylpiperazin-1-yl)acetohydrazide

Yıldız UYGUN CEBECİ[1*]
[1]*Kırklareli University, Faculty of Science, Chemistry Department, Kırklareli, Türkiye*
(ORCID: 0000-0001-7949-0329)

Keywords: Methylpiperazin, fluoroquinolone, mannich base, 1,2,4-triazole.

Abstract
This study explores the transformation of 2-(4-methylpiperazin-1-yl)acetohydrazide (1) through its reaction with phenylisothiocyanate (2), resulting in the formation of a novel urea derivative (3). The synthesis involves an intramolecular ring closure, where the hydrazide functionality plays a crucial role, leading to the formation of a 1,2,4-triazole structure. The journey continues as the 1,2,4-triazole derivative undergoes a Mannich reaction, leveraging its active methylene group. By reacting Compound 3 with formaldehyde and a selection of primary or secondary amines, a β-aminocarbonyl compound is synthesized, showcasing a significant molecular transformation.
The structural elucidation of the synthesized compounds is carried out using a range of sophisticated analytical techniques, including mass spectrometry, infrared (IR) spectroscopy, and nuclear magnetic resonance (NMR) spectroscopy for both ^{1}H and ^{13}C nuclei. These methods provide deep insights into the molecular architecture, enabling the determination of atomic connectivity and the identification of functional groups, thus confirming the integrity and structure of the synthesized molecules.

1. Introduction

Heterocyclic compounds are integral to the field of chemistry, particularly in pharmaceutical development, due to their unique ability to incorporate heteroatoms like nitrogen, oxygen, or sulfur into their ring structures. These compounds exhibit a broad spectrum of biological activities, which makes them critical targets in medicinal chemistry [1]. The structural diversity of heterocycles has led to the discovery of new therapeutic agents, including antimicrobials, anti-inflammatory drugs, anticancer agents, and antioxidants [2,3]. Among these, five-membered heterocyclic rings such as triazoles and diazoles have gained significant attention due to their potent biological properties [4,5].
Triazoles, in particular, are heterocyclic compounds characterized by the presence of three contiguous nitrogen atoms [6]. Their exceptional stability and ability to form hydrogen bonds enhance their solubility and binding affinity for biomolecular targets [7]. These characteristics make triazoles attractive scaffolds for the development of bioactive molecules [8]. Triazole-containing drugs such as fluconazole, voriconazole, and ravuconazole are widely used antifungal agents in clinical practice, demonstrating the therapeutic potential of this functional group [9,10]. Beyond antifungal activity, triazoles have also shown antibacterial, antiviral, anticancer, and anti-inflammatory effects, making them versatile candidates in drug discovery [11,12].
Fluoroquinolones are another class of compounds with significant antibacterial efficacy. These drugs inhibit DNA replication and transcription in bacteria by stabilizing the complex between DNA and topoisomerases, specifically targeting enzymes such as DNA gyrase and topoisomerase IV [13,14]. While early quinolones like nalidixic acid exhibited limited activity due to their narrow spectrum [15], the introduction of fluorine atoms at the C6 position in fluoroquinolones led to a significant enhancement in antibacterial potency and pharmacokinetics [16]. This structural modification, along with further advancements in fluoroquinolone chemistry, has expanded their clinical utility against a broad range of bacterial infections (Figure 1) [17].

* Corresponding author: yildizuygun41@hotmail.com
Received: 20.05.2024, Accepted: 18.10.2024

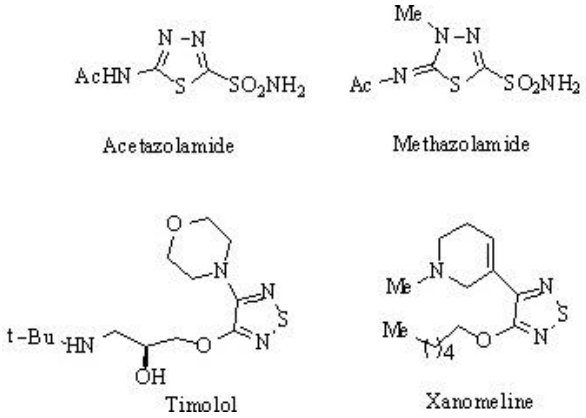


Figure 1: Some Selected Thiadiazole-Based Drugs

Research into bioactive heterocyclic compounds has consistently demonstrated the potential of these molecules in various therapeutic areas [18]. Triazoles and thiadiazoles, in particular, have emerged as critical components in the design of new drugs due to their diverse biological activities [19,20]. Compounds containing 1,2,4-triazole and 1,3,4-thiadiazole moieties have been reported to exhibit anticancer, antibacterial, antifungal, and antioxidant properties [21,22]. These heterocyclic structures are commonly incorporated into pharmaceuticals to enhance their biological efficacy [23]. Clinically used drugs like timolol, methazolamide, and acetazolamide highlight the medical importance of triazole and thiadiazole-containing compounds [24]. Their ability to interact with biological targets through hydrogen bonding and other interactions underpins their wide therapeutic applicability [25,26].

Triazoles are particularly noteworthy for their role as linker molecules in the design of hybrid drugs. By connecting two pharmacophores through a triazole moiety, researchers have been able to develop bifunctional drugs that target multiple biological pathways simultaneously, thereby enhancing therapeutic outcomes [27,28]. This strategy is exemplified by antifungal drugs like fluconazole and voriconazole, where the triazole group plays a critical role in improving drug-target interactions (Figure 2) [29]. Additionally, the triazole scaffold has been employed in the design of compounds with antibacterial, antiviral, and anticancer activities, broadening its application in drug development [30].

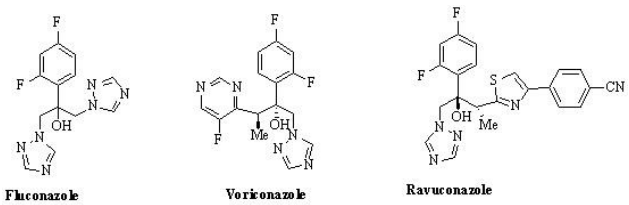


Figure 2: Some Selected Triazole-Based Drugs

Fluoroquinolones have also undergone substantial development over the years, particularly in addressing their limitations against Gram-positive bacteria [31]. Structural modifications, such as the introduction of nitrogen-containing heterocyclic rings at the C7 position, have expanded the antibacterial spectrum of fluoroquinolones and improved their pharmacokinetic profiles [32,33]. These modifications have resulted in the synthesis of widely used drugs like ciprofloxacin, norfloxacin, and ofloxacin, which exhibit broad-spectrum activity against a variety of bacterial pathogens [34]. Fluoroquinolones work by inhibiting DNA gyrase and topoisomerase IV, enzymes crucial for bacterial DNA replication [35]. This mode of action, combined with their ability to target both Gram-positive and Gram-negative bacteria, makes fluoroquinolones invaluable in clinical practice [36].

Recent studies have focused on further optimizing triazole and fluoroquinolone derivatives to enhance their biological activity and combat resistance mechanisms in pathogens [37]. For instance, fluoroquinolones have shown effectiveness against bacteria resistant to beta-lactam and aminoglycoside antibiotics, making them critical in treating drug-resistant infections [38]. Ongoing research in this area aims to develop new compounds that retain efficacy against resistant strains, thereby addressing one of the most pressing challenges in modern medicine [39,40]. Figure 3 showcases a selection of notable fluoroquinolone-containing compounds, highlighting their diverse applications and structural innovations.

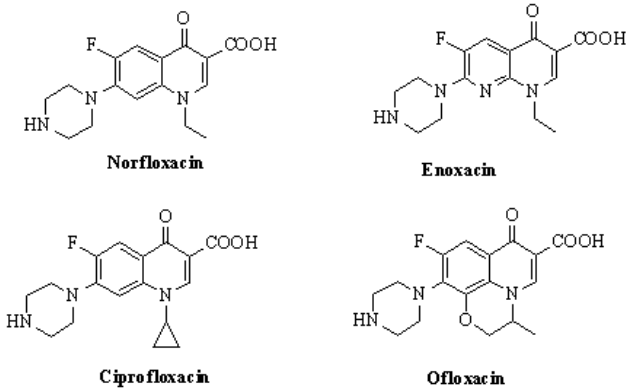


Figure 3: Chemical Structures of Selected Fluoroquinolones

Regarding their mechanism of action, FQs prevent bacteria from replicating and transcribing DNA by maintaining the stability of the complex formed between DNA and topoisomerases. Quinolones function by modifying the enzymes they specifically target, gyrase and topoisomerase IV, causing them to become detrimental and disrupt the

structure of the bacterial chromosome. Resistance related to the target is a widely observed and highly significant form of resistance in clinical settings. These mutations weaken the interactions between quinolones and enzymes like gyrase and topoisomerase IV, leading to the main cause of the issue [41]. Additional resistance mechanisms involve safeguarding the target, deactivating the drug through metabolism, and controlling drug entry and exit [42–44]. By developing novel quinolones that remain effective against these modified enzymes, the potential of this drug class could be significantly enhanced in clinical applications.

2. Material and Method

2.1. Chemistry

In this study, commercially available reagents were used without further purification. Melting points of the synthesized compounds were measured using a Büchi B-540 melting point apparatus. Thin-layer chromatography (TLC) was performed on silica gel 60 F254 plates, using ethyl acetate and diethyl ether in a 1:2 ratio as the mobile phase, and the reaction progress was monitored under UV light. Infrared (IR) spectra were recorded on a Perkin Elmer 1600 series FTIR spectrometer. Nuclear magnetic resonance (NMR) spectra were obtained using a BRUKER AVENE II 400 MHz spectrometer, with DMSO-d6 as the solvent. Chemical shifts (δ) were reported in ppm relative to tetramethylsilane (TMS), and coupling constants (J) were given in Hz.

2-(4-Methylpiperazin-1-yl)Acetohydrazide (1)

Hydrazine hydrate (25 mmol) was added dropwise to a solution of the precursor compound (10 mmol) in ethanol. The reaction mixture was refluxed for 8 hours. After cooling, the crude product was purified by column chromatography using silica gel and a hexane/ethyl acetate mixture (7:3). Yield: 85%, m.p.: 134-136°C. FT IR (υ_{max}, cm^{-1}): 1734 (C=O). ^{1}H NMR (DMSO-d_6, δ ppm): 2. 14 (3H, s, CH_3), 2.30 (2H, s, CH_2), 2.40 (4H, s, $2CH_2$), 2.88 (4H, s, $2CH_2$), 3.75 (2H, brs, NH_2), 8.84 (1H, s, CH_2). ^{13}C NMR (DMSO-d_6, δ ppm): 46.20 (CH_3), 53.21 ($2CH_2$), 55.01 ($2CH_2$), 60.37 (CH_2), 168.69 (C=O). EI MS *m/z* (%): 173.89 ($[M+1]^+$, 100), 195.75 ($[M+Na]^+$, 75).

***N*-Phenyl-2-[(4-Methylpiperazin-1-yl)Acetyl]Hydrazinecarbothioamide (2)**

A solution of phenylisothiocyanate (20 mmol) was added to the hydrazide compound 1 (10 mmol) in dichloromethane. The reaction was stirred at room temperature for 24 hours. The resulting solid was filtered and crystallized from ethanol to yield the pure product. Yield: 91%, m.p.: 152-154°C. FT IR (υ_{max}, cm^{-1}): 3217 (NH), 3254 (NH), 1703 (C=O), 2145 (C=S). ^{1}H NMR (DMSO-d_6, δ ppm): 1.89 (3H, s, CH_3), 2.42 (2H, s, CH_2), 2.62 (2H, s, CH_2), 2.72 (2H, s, CH_2), 3.11 (2H, s, CH_2), 3.11 (2H, s, CH_2), 5.76 (2H, s, CH_2), 7.16-7.65 (5H, m, arH), 9.56 (2H, s, 2NH), 9.87 (1H, s, NH). ^{13}C NMR (DMSO-d_6, δ ppm): 44.81 (CH_3), 51.81 ($2CH_2$), 54.22 ($2CH_2$), 59.68 (CH_2), arC: [117.35 (CH), 121.48 (CH), 128.61 (CH), 129.43 (CH), 129.52 (CH), 141.66 (C)] 156.16 (C=O), 181.05 (C=S). EI MS *m/z* (%): 308.26 ($[M+1]^+$, 100), 330.15 ($[M+Na]^+$, 72), 203.19 (51), 113.58 (30)

5-((4-Methylpiperazin-1-yl)Methyl)-4-Phenyl-4H-1,2,4-Triazole-3-Thiol (3)

The carboxamide derivative (2) was treated with sodium hydroxide in ethanol and refluxed for 8 hours. The reaction mixture was then acidified to pH 4-5 using hydrochloric acid, and the resulting precipitate was purified by filtration and recrystallized from ethyl acetate. Yield: 75%, m.p.: 162-164°C. FT IR (υ_{max}, cm^{-1}): 3067 (aromatic CH), 2558 (SH). ^{1}H NMR (DMSO-d_6, δ ppm): 2.10 (3H, s, CH_3), 3.35 (10H, s, $5CH_2$), 6.91-7.60 (5H, m, arH),13.35 (1H, s, SH). ^{13}C NMR (DMSO-d_6, δ ppm): 25.52 (CH_3), 53.89 ($2CH_2$), 54.85 ($2CH_2$), 57.23 (CH_2), arC: [129.09 (CH), 129.42 (CH), 129.87 (CH), 129.92 (CH), 133.15 (CH), 140.62 (C)], 148.72 (triazole C-3), 165.82 (triazole C-5). EI MS *m/z* (%): 290.36 ($[M+1]^+$, 100), 169.78 (69), 134.15 (37), 111.23 (20).

General Method for The Synthesis of Compounds 4a-d: Mannich bases were synthesized by reacting compound 3 (10 mmol) with formaldehyde (30 mmol) and various amines in dimethylamine at room temperature for 24 hours. The precipitate was filtered, washed with water, and recrystallized from a dimethylsulfoxide (1:1) mixture to yield the desired compounds.

5-((4-Methylpiperazin-1-yl)Methyl)-2-(Morpholinomethyl)-4-Phenyl-2H-1,2,4-Triazole-3(4H)-Thione (4a)

Yield: % 85, m.p: 178-180°C. FT IR (υ_{max}, cm^{-1}): 3088 (aromatic CH). ^{1}H NMR (DMSO-d_6, δ ppm): 2.15 (3H, s, CH_3), 3.40 (16H, s, $8CH_2$), 4.23 (2H, d, *J*= 4.0 Hz, CH_2), 5.46 (2H, s, CH_2), 7.28 (2H, d, *J*= 4.0 arH), 7.65 (1H, s, arH), 7.98 (2H, d, *J*= 8.0 Hz, arH). ^{13}C NMR (DMSO-d_6, δ ppm): 28.36 (CH_3), 46.12 ($2CH_2$), 47.63 ($2CH_2$), 48.69 ($2CH_2$), 49.79 ($2CH_2$), 50.10 (CH_2), 52.63 (CH_2), arC: [121.45 (CH), 122.87 (CH), 123.96 (CH), 124.60 (CH), 126.70 (CH), 139.50 (C)], 151.52 (triazole C-3), 159.63 (triazole C-5). EI MS *m/z* (%): 110.12 (100), 389.89 ($[M+1]^+$, 92), 214.63 (69), 143.56 (29).

5-((4-Methylpiperazin-1-yl)Methyl)-4-Phenyl-2-((4-Phenylpiperazin-1-yl)Methyl)-2H-1,2,4-Triazole-3(4H)-Thione (4b)

Yield: % 80, m.p: 181-183°C. FT IR (υ_{max}, cm^{-1}): 3075 (aromatic CH), 3089 (aromatic CH). ^{1}H NMR (DMSO-d_6, δ ppm): 1.24 (3H, s, CH_3), 2.01 (2H, s, CH_2), 2.11 (2H, s, CH_2), 3.40 (10H, s, $5CH_2$), 4.37 (2H, s, CH_2), 4.70 (2H, s, CH_2), 5.76 (2H, s, CH_2), 6.94-7.27 (7H, m, arH), 7.49 (2H, d, *J*= 8.0 Hz, arH), 8.07 (1H, s, arH). ^{13}C NMR (DMSO-d_6, δ ppm): 21.14 (CH_3), 47.71 ($2CH_2$), 48.88 ($2CH_2$), 50.12 ($2CH_2$), 51.98 ($2CH_2$), 53.43 (CH_2), 55.79 (CH_2), arC: [112.85 (CH), 113.97 (CH), 114.96 (CH), 115.37 (CH), 118.93 (CH), 119.63 (CH), 120.12 (CH), 121.53 (CH), 122.38 (CH), 123.49 (CH), 139.12 (C), 140.46 (C)], 153.96 (triazole C-3), 157.83 (triazole C-5). EI MS *m/z* (%): 464.89 ($[M+1]^+$, 100), 300.19 (78), 178.12 (55), 112.09 (31).

1-Ethyl-6-Fluoro-7-(4-((3-((4-Methylpiperazin-1-yl)Methyl)-4-Phenyl-5-Thioxo-4,5-Dihydro-1,2,4-Triazol-1-yl)Methyl)Piperazin-1-yl)-4-Oxo-1,4-Dihydroquinoline-3-Carboxylic Acid (4c)

Yield: % 89, m.p: 252-254°C. FT IR (υ_{max}, cm^{-1}): 3312 (OH), 3083 (aromatic CH), 1698 (C=O), 1712 (C=O). ^{1}H NMR (DMSO-d_6, δ ppm): 1.17 (3H, s, CH_3), 1.42 (3H, s, CH_3), 2.73 (2H, s, CH_2), 2.82 (2H, s, CH_2), 2.89 (2H, s, CH_2), 3.39 (14H, s, CH_2), 6.95-7.27 (3H, m, arH), 7.49-7.56 (3H, m, arH), 7.87-7.95 (1H, m, arH), 8.93 (1H, s, CH), 15.35 (1H, s, OH). ^{13}C NMR (DMSO-d_6, δ ppm): 17.23 (CH_3), 21.25 (CH_3), 48.52 ($2CH_2$), 49.12 ($2CH_2$), 51.10 ($2CH_2$), 52.23 ($2CH_2$), 53.96 (CH_2), 54.30 (CH_2), 57.23 (CH_2), arC: [112.52 (CH), 113.12 (CH), 117.60 (CH), 119.63 (CH), 124.89 (CH), 125.43 (CH), 126.30 (CH), 138.12 (C), 139.50 (C), 140.12 (C), 141.23 (C), 143.98 (C), 144.07 (C)], 148.12 (quinolone CH), 155.13 (triazole C-3), 159.96 (triazole C-5), 171.12 (C=O), 173.20 (C=O).EI MS *m/z* (%): 389.12 (100), 621.89 ($[M+1]^+$, 90), 643.30 ($[M+Na]^+$, 71), 187.12 (36).

1-Cyclopropyl-6-Fluoro-7-(4-((3-((4-Methyl piperazin-1-yl)Methyl)-4-Phenyl-5-Thioxo-4,5-Dihydro-1,2,4-Triazol-1-yl)Methyl)Piperazin-1-yl)-4-Oxo-1,4-Dihydroquinoline-3-Carboxylic Acid (4d)

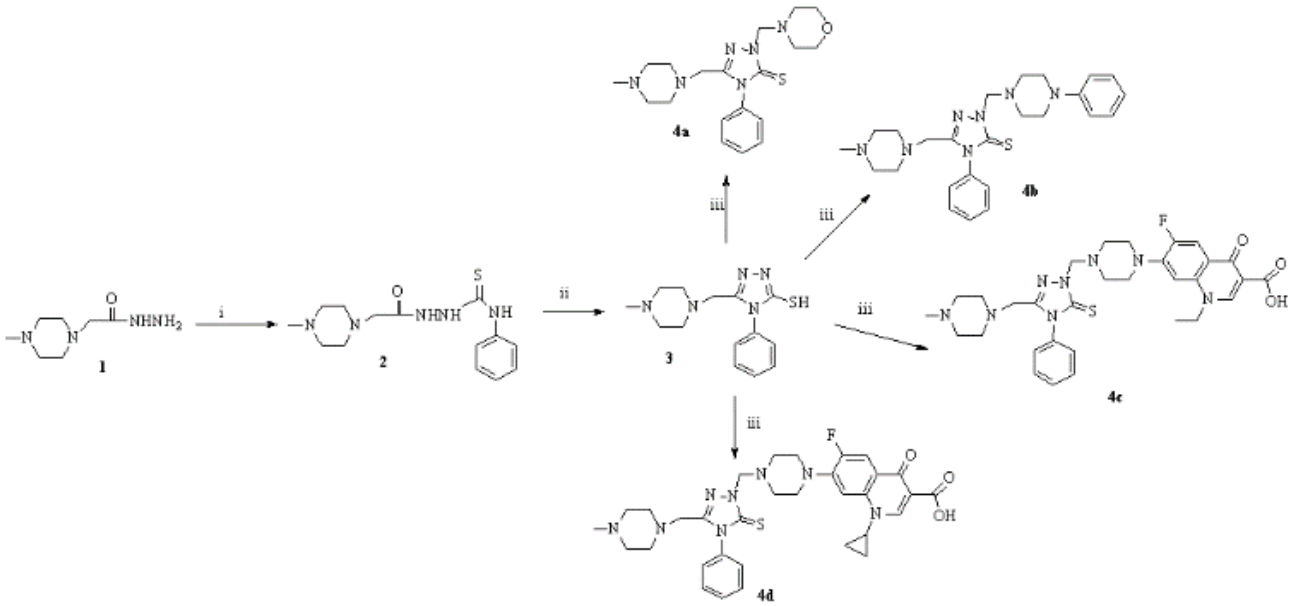
Yield: % 91, m.p: 250-252 °C. FT IR (υ_{max}, cm^{-1}): 3323 (OH), 3091 (aromatic CH), 1702 (C=O), 1723 (C=O). ^{1}H NMR (DMSO-d_6, δ ppm): 1.13 (3H, s, CH_2), 1.29 (3H, s, CH_2), 2.73 (3H, s, CH_3), 2.82 (2H, s, CH_2), 2.89 (2H, s, CH_2), 3.29 (2H, s, CH_2), 3.35 (8H, s, $4CH_2$), 4.33 (2H, d, *J*= 8.0 Hz, CH_2), 5.00 (2H, s, CH_2), 5.24 (2H, s, CH_2), 7.21-7.52 (7H, s, arH), 7.80 (1H, d, *J*= 12.0 Hz, CH), 8.65 (1H, s, CH), 15.23 (1H, s, OH). ^{13}C NMR (DMSO-d_6, δ ppm): 12.12 (CH_2), 14.78 (CH_2), 34.79 (CH_3), 37.98 (CH_2), 39.12 (CH_2), 40.65 (CH_2), 41.01 (CH_2), 43.50 (CH_2), 44.97 (CH_2), 45.32 (CH_2), 46.89 (CH_2), 47.81 (CH_2), 50.03 (CH_2), 103.12 (CH), arC: [115.12 (CH), 116.89 (CH), 121.20 (CH), 122.85 (CH), 126.50 (CH), 128.90 (CH), 130.41 (CH), 133.56 (C), 134.52 (C), 136.12 (C), 137.20 (C), 138.43 (C), 139.61 (C)], 148.34 (quinolone CH), 157.10 (triazole C-3), 160.12 (triazole C-5), 171.77 (C=O), 174.09 (C=O). EI MS *m/z* (%): 632.12 ($[M+1]^+$, 100), 498.12 (88), 346.98 (71), 219.52 (55).

3. Results and Discussion

3.1. Chemistry

This work involved the conversion of a carbox(thio)amide derivative (1) through an intramolecular cyclization reaction to produce 1,2,4-triazoles (2) as shown in Scheme 1. The molecule was identified by the presence of a signal at 13.35 ppm in the 1H NMR data, which was a singlet that could be exchanged with D2O. This confirmed the presence of a -SH group. The stretching band originating from these groups was observed at a wavenumber of 2558 cm-1 in the FT-IR data of this molecule.

Mannich bases are bioactive compounds [48–50]. Mannich bases (4a–d) were synthesized by reacting compound (3) with formaldehyde and several physiologically active amines (Scheme 1) [51, 52]. The hydroxyl (OH) peaks of molecules 4c and 4d, which belong to the fluoroquinolone ring, were observed with chemical shifts of 15.35 and 15.23 ppm, respectively, in the 1H NMR spectrum. The carbon-hydrogen (CH) peaks of molecules 4c and 4d, which belong to the fluoroquinolone ring, exhibited signals at 148.12 and 148.34 ppm in the 13C NMR spectrum.



Scheme 1: i: RNCS, EtOH, ii. NaOH, EtOH, H_2O; iii: dimethyl formamide, seconder amine, room temperature, 24

3. Results and Discussion

3.1. Chemistry

This work involved the conversion of a carbox(thio)amide derivative (1) through an intramolecular cyclization reaction to produce 1,2,4-triazoles (2) as shown in Scheme 1. The molecule was identified by the presence of a signal at 13.35 ppm in the 1H NMR data, which was a singlet that could be exchanged with D2O. This confirmed the presence of a -SH group. The stretching band originating from these groups was observed at a wavenumber of 2558 cm-1 in the FT-IR data of this molecule.

Mannich bases are bioactive compounds [48–50]. Mannich bases (4a–d) were synthesized by reacting compound (3) with formaldehyde and several physiologically active amines (Scheme 1) [51, 52]. The hydroxyl (OH) peaks of molecules 4c and 4d, which belong to the fluoroquinolone ring, were observed with chemical shifts of 15.35 and 15.23 ppm, respectively, in the 1H NMR spectrum. The carbon-hydrogen (CH) peaks of molecules 4c and 4d, which belong to the fluoroquinolone ring, exhibited signals at 148.12 and 148.34 ppm in the 13C NMR spectrum.

4. Conclusion and Suggestions

In this study, the synthesis of a 1,2,4-triazole derivative was accomplished, starting from the precursor 2-(4-methylpiperazin-1-yl)acetohydrazide (1). Following the initial synthesis, a series of novel Mannich bases (4a-d) were obtained through a Mannich reaction with various amines, introducing structural diversity and potential for enhanced biological activity. The synthesized compounds are of significant interest due to their relevance in medicinal chemistry and widespread applications reported in the literature. Comprehensive structural characterization of these derivatives was carried out using a combination of advanced spectroscopic techniques, including infrared (IR) spectroscopy, proton nuclear magnetic resonance (^{1}H NMR), carbon-13 nuclear magnetic resonance (^{13}C NMR), and mass spectrometry, confirming their molecular frameworks and functional integrity. The successful synthesis and detailed structural validation of these compounds contribute valuable insights to the field, potentially paving the way for future studies on their pharmacological properties.

Statement of Research and Publication Ethics.

The study is complied with research and publication ethics

References

[1] L. I. Belen'kii, G. A. Gazieva, Y. B. Evdokimenkova, and N. O. Soboleva, "The literature of heterocyclic chemistry, part XVIII," in *Adv. Heterocycl. Chem.*, E. F. V. Scriven and C. A. Ramsden, Eds. Elsevier, 2020, pp. 385–468.

[2] P. T. Acharya, Z. A. Bhavsar, D. J. Jethava, D. B. Patel, and H. D. Patel, "A review on development of bio-active thiosemicarbazide derivatives: Recent advances," *J. Mol. Struct.*, vol. 1226, p. 129268, 2021.

[3] P. Martins *et al.*, "Heterocyclic anticancer compounds: Recent advances and the paradigm shift towards the use of nanomedicine's toolbox," *Molecules*, vol. 20, pp. 16852–16891, 2015.

[4] H. Huo, G. Li, B. Shi, and J. Li, "Recent advances on synthesis and biological activities of C-17 aza-heterocycle derived steroids," *Bioorg. Med. Chem.*, vol. 116882, 2022.

[5] M. M. Heravi and S. Sadjadi, "Recent advances in the application of the Sonogashira method in the synthesis of heterocyclic compounds," *Tetrahedron*, vol. 65, pp. 7761–7775, 2009.

[6] M. Bashir, A. Bano, A. S. Ijaz, and B. A. Chaudhary, "Recent developments and biological activities of N-substituted carbazole derivatives: A review," *Molecules*, vol. 20, pp. 13496–13517, 2015.

[7] A. R. Shaikh, M. Farooqui, R. Satpute, and S. Abed, "Overview on nitrogen containing compounds and their assessment based on 'International Regulatory Standards'," *J. Drug Deliv. Ther.*, vol. 8, pp. 424–428, 2018.

[8] T. Qadir, A. Amin, P. K. Sharma, I. Jeelani, and H. Abe, "A review on medicinally important heterocyclic compounds," *The Open Med. Chem. J.*, vol. 16, 2022.

[9] E. Kabir and M. Uzzaman, "A review on biological and medicinal impact of heterocyclic compounds," *Results Chem.*, vol. 100606, 2022.

[10] T. A. Farghaly *et al.*, "Hydrazonoyl halides as precursors for synthesis of bioactive heterocyclic compounds via reaction with aryl- and hetaryl-amine," *Curr. Org. Chem.*, vol. 27, no. 15, 2022.

[11] A. Y. Khormi *et al.*, "Novel thiazole derivatives incorporating phenyl sulphonyl moiety as potent BRAFV600E kinase inhibitors targeting melanoma," *RSC Adv.*, vol. 12, no. 42, pp. 27355–27369, 2022.

[12] M. S. Saini, A. Kumar, J. Dwivedi, and R. Singh, "A review: Biological significances of heterocyclic compounds," *Int. J. Pharm. Sci. Res.*, vol. 4, pp. 66–77, 2013.

[13] P. V. Thanikachalam, R. K. Maurya, V. Garg, and V. Monga, "An insight into the medicinal perspective of synthetic analogs of indole: A review," *Eur. J. Med. Chem.*, vol. 180, pp. 562–612, 2019.

[14] T. A. Farghaly, W. A. Al-Hasani, M. H. Ibrahim, M. H. Abdellattif, and Z. A. Abdallah, "Design, synthesis, anticancer activity and docking studies of thiazole-linked phenylsulfone moiety as cyclin-dependent kinase 2 (CDK2) inhibitors," *Polycycl. Aromat. Compd.*, in press, 2022.

[15] M. S. Vasava, M. N. Bhoi, S. K. Rathwa, D. J. Jethava, P. T. Acharya, and D. B. Patel, "Benzimidazole: A milestone in the field of medicinal chemistry," *Mini Rev. Med. Chem.*, vol. 20, pp. 532–565, 2020.

[16] M. G. Zayda, A.-A.-H. Abdel-Rahman, and F. A. El-Essawy, "Synthesis and antibacterial activities of different five-membered heterocyclic rings incorporated with pyridothienopyrimidine," *ACS Omega*, vol. 5, pp. 6163–6168, 2020.

[17] F. Plescia, B. Maggio, G. Daidone, and D. Raffa, "4-(3H)-quinazolinones N-3 substituted with a five-membered heterocycle: A promising scaffold towards bioactive molecules," *Eur. J. Med. Chem.*, vol. 213, p. 113070, 2021.

[18] S. M. Gomha, S. M. Riyadh, T. A. Farghaly, and R. Haggam, "Synthetic utility of bis-aminomercapto[1,2,4] triazoles in the preparation of bis-fused triazoles and macrocycles," *Polycycl. Aromat. Compd.*, in press, 2022.

[19] I. D. H. Elnaggar *et al.*, "Antiproliferative activity of some newly synthesized substituted nicotinamides candidates using pyridine-2(1H) thione derivatives as synthon," *ACS Omega*, vol. 7, no. 12, pp. 10304–10316, 2022.

[20] M. M. Burbuliene, V. Sakociute, and P. Vainilavicius, "Synthesis and characterization of new pyrimidine-based 1,3,4-oxa (thia) diazole, 1,2,4-triazole and 4-thiazolidinones," *Arkivoc*, vol. 12, pp. 281–289, 2009.

[21] I. Çapan, S. Servi, I. Yıldırım, and Y. Sert, "Synthesis, DFT Study, Molecular Docking and Drug-Likeness Analysis of the New Hydrazine-1-Carbothioamide, Triazole and Thiadiazole Derivatives: Potential Inhibitors of HSP90," *ChemistrySelect*, vol. 6, pp. 5838–5846, 2021.

[22] Y. Li, J. Geng, Y. Liu, S. Yu, and G. Zhao, "Thiadiazole – A promising structure in medicinal chemistry," *ChemMedChem*, vol. 8, pp. 27–41, 2013.

[23] I. Khan *et al.*, "Synthesis, antioxidant activities and urease inhibition of some new 1,2,4-triazole and 1,3,4-thiadiazole derivatives," *Eur. J. Med. Chem.*, vol. 45, pp. 5200–5207, 2010.

[24] S. Sathish Kumar and H. P. Kavitha, "Synthesis and biological applications of triazole derivatives–a review," *Mini-Rev. Org. Chem.*, vol. 10, pp. 40–65, 2013.

[25] O. M. Soltan *et al.*, "Molecular hybrids: A five-year survey on structures of multiple targeted hybrids of protein kinase inhibitors for cancer therapy," *Eur. J. Med. Chem.*, vol. 225, p. 113768, 2021.

[26] E. N. da Silva Júnior *et al.*, "Synthesis of quinones with highlighted biological applications: A critical update on the strategies towards bioactive compounds with emphasis on lapachones," *Eur. J. Med. Chem.*, vol. 179, pp. 863–915, 2019.

[27] R. Aggarwal and G. Sumran, "An insight on medicinal attributes of 1,2,4-triazoles," *Eur. J. Med. Chem.*, vol. 205, p. 112652, 2020.

[28] A. H. Malani, A. H. Makwana, and H. R. Makwana, "A brief review article: Various synthesis and therapeutic importance of 1,2,4-triazole and its derivatives," *Moroccan J. Chem.*, vol. 5, pp. 41–58, 2017.

[29] R. V. Veloso *et al.*, "Antioxidant and anti-sickling activity of glucal-based triazoles compounds–An in vitro and in silico study," *Bioorg. Chem.*, vol. 109, p. 104709, 2021.

[30] C. M. Oliphant and G. M. Green, "Quinolones: a comprehensive review," *Am. Fam. Physician*, vol. 65, pp. 455–464, 2002.

[31] C. Mugnaini, S. Pasquini, and F. Corelli, "The 4-quinolone-3-carboxylic acid motif as a multivalent scaffold in medicinal chemistry," *Curr. Med. Chem.*, vol. 16, pp. 1746–1767, 2009, doi: 10.2174/092986709788186156.

[32] J. M. Blondeau, "Expanded activity and utility of the new fluoroquinolones: a review," *Clin. Ther.*, vol. 21, pp. 3–40, 1999, doi: 10.1016/S0149-2918(00)88266-1.

[33] G. L. Patrick, *Antibacterial agents in An Introduction to Medicinal Chemistry*. Oxford University Press, Oxford, New York, 2003.

[34] P. C. Appelbaum and P. A. Hunter, "The fluoroquinolone antibacterials: past, present, and future perspectives," *Int. J. Antimicrob. Agents*, vol. 16, pp. 5–15, 2000, doi: 10.1016/s0924-8579(00)00192-8.

[35] P. C. Sharma, A. Jain, and S. Jain, "Fluoroquinolone antibacterials: a review on chemistry, microbiology, and therapeutic prospects," *Acta Pol. Pharm.*, vol. 66, pp. 587–604, 2009.

[36] M. V. De Almeida *et al.*, "Synthesis and antitubercular activity of lipophilic moxifloxacin and gatifloxacin derivatives," *Bioorg. Med. Chem. Lett.*, vol. 17, pp. 5661–5664, 2007, doi: 10.1016/j.bmcl.2007.07.073.

[37] G. Anquetin *et al.*, "Design, synthesis, and activity against *Toxoplasma gondii, Plasmodium spp.*, and *Mycobacterium tuberculosis* of new 6-fluoroquinolones," *Eur. J. Med. Chem.*, vol. 41, pp. 1478–1493, 2006, doi: 10.1016/j.ejmech.2006.07.003.

[38] B. A. Cunha, "The fluoroquinolones for urinary tract infections: a review," *Adv. Ther.*, vol. 11, pp. 277–296, 1994.

[39] R. N. Jones, D. E. Low, and M. A. Pfaller, "Epidemiologic trends in nosocomial and community-acquired infections due to antibiotic-resistant gram-positive bacteria: the role of streptogramins and other newer compounds," *Diagn. Microbiol. Infect Dis.*, vol. 33, pp. 101–112, 1999, doi: 10.1016/s0732-8893(98)00108-4.

[40] D. C. Hooper, "Mechanisms of action and resistance of older and newer fluoroquinolones," *Clin. Infect. Dis.*, vol. 31, pp. S24–S28, 2000, doi: 10.1086/314056.

[41] D. M. Campoli-Richards *et al.*, "Ciprofloxacin: a review of its antibacterial activity, pharmacokinetic properties, and therapeutic use," *Drugs*, vol. 35, pp. 373–447, 1988, doi: 10.2165/00003495-198835040-00003.

[42] K. Drlica *et al.*, "Quinolones: action and resistance updated," *Curr. Top. Med. Chem.*, vol. 9, pp. 981–998, 2009, doi: 10.2174/156802609789630947.

[43] D. C. Hooper, "Emerging mechanisms of fluoroquinolone resistance," *Emerg. Infect. Dis.*, vol. 7, pp. 337–341, 2001, doi: 10.3201/eid0702.010239.

[44] X. Guan *et al.*, "Plasmid-mediated quinolone resistance–current knowledge and future perspectives," *J. Int. Med. Res.*, vol. 41, pp. 20–30, 2013, doi: 10.1177/0300060513475965.

[45] V. Lorian, *Antibiotics in Laboratory Medicine*, Lippincott Williams & Wilkins, 2005.

[46] P. A. Wayne, *Performance Standards for Antimicrobial Susceptibility Testing: Nineteenth Informational Supplement, Clin. Lab. Standards Inst.*, M100-S19, 2009.

[47] P. A. Wayne, *Reference Method for Broth Dilution Antifungal Susceptibility Testing of Yeasts, Approved Standard*, vol. 3, pp. 6–12, 2008.

[48] Y. J. Zhao, W. Wei, Z. G. Su, and G. H. Ma, "Poly(ethylene glycol) prodrug for anthracyclines via N-Mannich base linker: design, synthesis and biological evaluation," *Int. J. Pharm.*, vol. 379, pp. 90–99, 2009.

[49] O. G. Roman, "Mannich bases in medicinal chemistry and drug design," *Eur. J. Med. Chem.*, vol. 89, pp. 743–816, 2015.

[50] K. Ganiyat, I. E. Willie, and O. Oluwakemi, "Synthesis of Mannich bases: 2-(3-Phenylaminopropionyloxy)-benzoic acid and 3-phenylamino-1-(2, 4, 6-trimethoxy-phenyl)propan-1-one, their toxicity, ionization constant, antimicrobial and antioxidant activities," *Food Chem.*, vol. 165, pp. 515–521, 2014.

[51] T. Fekner *et al.*, "Synthesis of (6S)-cephalosporins from 6-aminopenicillanic acid," *Tetrahedron*, vol. 56, pp. 6053–6074, 2000.

[52] A. Dassonville-Klimpt *et al.*, "Synthesis and antibacterial activity of catecholate-ciprofloxacin conjugates," *Bioorg. Med. Chem.*, vol. 22, pp. 4049–4060, 2014.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 988-998 YEAR: 2024
DOI: 10.17798/bitlisfen.1490703

Experimental Investigation of Rejuvenated Asphalt Mixtures Using Bio-Oils from Different Biomass Sources

Beyza FURTANA YALÇIN[1*], Mehmet YILMAZ[2], Hamdi Soner ALTUNDOĞAN[3]

[1]*Munzur University, Civil Eng. Department, Tunceli, Türkiye*
[2]*Fırat University, Civil Eng. Department, Elazığ, Türkiye*
[3]*Fırat University, Bioengineering. Department, Elazığ, Türkiye*
(ORCID: 0000-0003-4565-7324) (ORCID: 0000-0002-2761-2598) (ORCID: 0000-0003-3328-9667)

Keywords: RAP, Rejuvenator, Bio-oil, Biomass, Pyrolysis.

Abstract

Asphalt mixes containing Recycled Asphalt Pavement (RAP) materials are generally harder than traditional mixes. Different rejuvenators are used to soften recycled asphalt pavement by reducing this hardness. In this study, recycled asphalt pavement was added to bituminous hot mixtures at 3 different rates (25%, 50% and 75%). To rejuvenate bituminous hot mixtures containing RAP, three different rejuvenators (pine cone, olive pomace and wheat straw) were added to the mixtures. This study investigated the effect of asphalt mixtures containing RAP and rejuvenator on the mechanical properties. Marshall test and moisture damage test were performed to examine the mechanical properties of asphalt mixture samples. As the bio-oil content in asphalt mixtures increases, the stability values of the mixtures decrease. As the bio-oil ratio in asphalt mixtures increases, tensile strength (TS) values decrease in both unconditioned and conditioned mixtures. While the increase in bio-oil reduces the tensile strength ratio, the shrinkage ratio of mixtures prepared with RAP increases. The bio-oil derived from pomace has been identified as the most effective rejuvenator among those tested.

1. Introduction

The cost of bitumen derived from petroleum has significantly increased as a result of its use in road building and maintenance [1-4]. Furthermore, the upkeep of ecologically appropriate pavement has become popular [2, 5-8]. Research is being done on alternate binders as a substitute for bitumen, which is often used in traditional processes. One efficient method of lowering bitumen use is to use Recycled Asphalt Pavement (RAP) material [9]. An important problem in infrastructure engineering is highlighted by the aging of asphalt binders throughout construction and service life [10–12]. Oxidation and volatile component loss are the main aging mechanisms of asphalt binders. As a result, asphalt binders become more viscous and tougher than when they were first created [13, 14]. To address these issues, various types of regenerating RAP mixes have been utilized [9,15–19]. Because of its many sources, high efficiency, and low cost, bio-oil is becoming more and more popular among other renewable energy [2, 20, 21]. The paragraph indentation should be used at the start of any subsequent paragraphs after the first.

Using pyrolysis techniques, biomass is the main source of bio-oil. Slow pyrolysis and rapid pyrolysis are the two categories of pyrolysis techniques. The residence time during the pyrolysis process is what distinguishes slow pyrolysis from quick pyrolysis. Fast pyrolysis has a residence period of less than 10 seconds, whereas slow pyrolysis has a residence time of 5 to 12 hours [22]. Three primary products of the pyrolysis process are produced: liquids, gases, and charcoal. The substance is regarded as bio-oil [23–25]. The use of bio-oils to alter or partially replace asphalt binders has grown [2, 26, 27].

According to earlier research, bio-oils can improve the low-temperature performance of asphalt

*Corresponding author: beyzafurtana@munzur.edu.tr
Received: 27.05.2024, Accepted: 25.10.2024

binders by softening them [21-22, 28]. Furthermore, the possibility of restoring reclaimed asphalt binder to its initial condition has been investigated by certain researchers. As modifiers, a variety of bio-oils can be employed, including distilled tall oil, organic oil, waste vegetable oil, and others. Depending on the type of source, different rejuvenating effects apply. Specifically, it has been demonstrated that rejuvenation with waste vegetable oil at a concentration of 12% by weight can increase the aged asphalt's resistance to cracking, lowering its performance grade from PG 94-12 to PG 72-33 [17]. By adding 1.75-2% by weight of bio-oil, a bio-oil rejuvenator made from biodiesel residue makes up for the light components lost in aged binders and improves crack resistance at low temperatures [29]. The physical and rheological properties of 40/50 penetration grade bitumen can be enhanced with 3-4% by weight of waste cooking oil, approaching the characteristics of 80/100 penetration grade bitumen [30]. Rzek and associates used the pyrolysis of used tires to create a rejuvenator for their investigation. They supplemented three distinct ratios of recycled asphalt mixtures with these rejuvenators. They consequently discovered that recovered asphalt mixtures were used at a higher rate when the recently created alternative rejuvenator was added at a 20% rate. They discovered that adding this alternative rejuvenator might raise the percentage of reclaimed asphalt to 60% by the application of routine mechanical and rheological tests to the asphalt mixtures [27]. Avsenik and his colleagues looked at how aged bitumen was affected by bio-oils made from burning discarded tires. The bitumen, which was laboratory-aged and had undergone both short- and long-term aging operations, was 50/70 penetration grade. They evaluated the effect of rejuvenator addition on aged and unaged bitumen using standard mechanical and rheological tests at four different rates (3%, 5%, 10%, and 20%). The rejuvenator was found to be appropriate for altering aged bitumen by the researchers based on their mechanical and rheological testing [28]. In their studies, Oruc et al. prepared mixtures containing 10%, 25% and 40% RAP. Marshall tests were applied to these mixture samples, density and void analyses were performed, and the findings were compared with samples without recycling and core samples taken from the field [31]. In their study, Kumandas et al. added 2-8 wt% waste cooking oils (WCO) to bitumen modified with 1.5% reactive ethylene terpolymers (RET) and 0.2% polyphosphoric acid (PPA) as catalysts, in increments of 2%. It was observed that WCO addition increased the workability of RETMBs and their resistance to fatigue and low-temperature cracking [32].

In general, the properties of bio-rejuvenators are dependent on the biomass source. There is no study available on rejuvenating RAP binders using a bio-oil rejuvenator produced from wheat straw, pine cones, and rice husks. The rejuvenators used in the study were obtained through the pyrolysis method. In the pyrolysis method, biomass is burned at high temperatures in an oxygen-free environment to produce bio-oils. The use of this method also enhances the uniqueness of the study. Furthermore, utilizing bio-oil as a rejuvenator to recycle aged binders supports the repurposing of waste biomass and construction materials, promoting environmental sustainability and development. This study aims to offer an overview of RAP recyclability, focusing on the proportions of added RAP material, as well as performance, economic, and environmental considerations. The main goal is to promote the use of this material in similar applications and thereby maximize its use, contributing simultaneously to an environmentally and economically sustainable circular economy. Therefore, in the study, recycled asphalt pavements were added to bituminous hot mixtures at different percentages, and their mechanical properties were investigated. Additionally, three different bio-oils produced in the study were used to rejuvenate recycled asphalt mixtures. Subsequently, asphalt mixtures containing RAP and rejuvenator were compared with bituminous hot mixtures to investigate the effect of RAP and rejuvenator on mixtures.

2. Material and Method

2.1. Unmodified bitumen and Recycled Asphalt Pavement

The study used B 50/70 class bitumen from the TUPRAS\Batman refinery. Table 1 presents the general characteristics of the bitumen employed in the research. In this investigation, only one source of recycled asphalt pavement (RAP) material obtained from the Elazig Municipality in Turkey was utilized. To minimize its impact on the environment, the obtained RAP material was kept in impermeable containers. The original bitumen of the RAP material in the construction phase is B 50/70 class bitumen. The purpose of this study is to add rejuvenators to the RAP binder to raise its hardness to that of the bitumen utilized during the first building phase. 50/70 was therefore chosen as the pure bitumen. The penetration, softening point and viscosity test results at 135°C of the aged binder obtained from RAP are 10, 79.3°C, and 6338 cP, respectively.

Table 1. General properties of bitumen

Specification	Unit	Standard	Results
Penetration	mm^{-1}	EN 1426	56
Softening point	˚C	EN 1427	53.3
Flash point	˚C	EN ISO 2719	245
Specific weight	g/cm^3	ASTM D70 - 18a	1.015
Elastic recovery	%	EN 13398	30
Viscosity 135°C /165°C	cP	ASTM D4402	737.5/225
Mixing temperature	˚C	-	159-165
Compression temperature	˚C	-	145-151

2.2. Biomass

The biomass materials for the study, namely ground wheat straw, ground pine cones, and bran, were procured from Cambaylar Co. Ltd. situated in the province of Elazig. For use in this investigation, the biomass materials were filtered through a No 30 sieve to guarantee a consistent particle size distribution. This was done to ensure that the experimental results would not be impacted by the biomass's size.

2.3. Properties of the Reclaimed Asphalt Pavement (RAP) Material

The EN 932-1 standard was followed while sampling from the RAP material [33]. To ascertain the gradation of the recycled asphalt pavement material, an extraction test was conducted (Figure 1a). The amount of bitumen contained in the RAP material was ascertained after the extraction test. The recycled asphalt mixture was found to contain 4.6% bitumen. Following the TS EN 12697-3+A standard [34], the aged binder extracted and separated from the recycled asphalt binder was recovered utilizing a rotary evaporator (Figure 1b). To find the bitumen content in this investigation, 50 extraction tests were carried out. The entire investigation was conducted with the recycled and old binder.

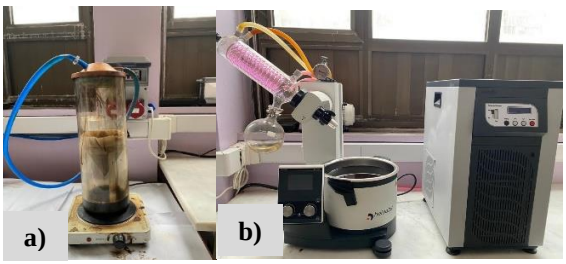


Figure 1. a) Extraction apparatus b) Rotary evaporator

2.4. Bio-oil Production

In this investigation, three distinct rejuvenators will be employed. These rejuvenators come from agricultural materials and are bio-based. The project will make use of bran bio-oils, pine cones, and wheat straw. In this investigation, materials that had been through a No. 30 sieve were employed to guarantee that the biomass produced had a consistent particle size distribution. This was done to ensure that the experimental results would not be impacted by the biomass's size. The biomass was subjected to a slow pyrolysis process using a device. The laboratory slow pyrolysis experimental setup comprises several components: a water cooling system, a programming device box for temperature adjustment during the experiment, a high-temperature-resistant cylindrical container with dimensions of 150 mm inner diameter and 240 mm height, and a tank where the condensed biogas is collected as oil after cooling. Figure 2 depicts the general design of the slow pyrolysis experimental setup.

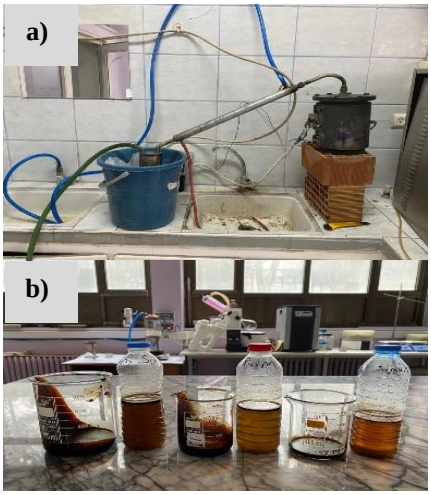


Figure 2. a) General view of the laboratory slow pyrolysis experimental setup, b) samples of bio-oil and water obtained as liquid products after pyrolysis

Before commencing the pyrolysis process, 1000 grams of biomass were consistently introduced into the chamber of the device to mitigate the influence of biomass quantity on carbonization. As a result, a fixed-bed slow pyrolysis system operating at 500°C was used to pyrolyze 1 kg of dry biomass sample, as shown in Figure 2. Based on research published in the literature, this temperature was selected. The system's pyrolysis vapors were burned in a chimney inside the system outlet, while the uncondensed pyrolysis gases were condensed into liquid pyrolysis products (bio-oil) by a condenser cooled with water. The pyrolysis process kept going until no more gas was released. To calculate the

quantities of biochar, the liquid pyrolysis products and the pyrolysis residue were weighed after cooling. The non-condensed gas product quantity was then determined by deducting the total weight of these products from the dry biomass amount that was initially used. Table 2 displays the results that were achieved. To extract water and organic acids from the oil phase, the produced liquid products (Figure 2), which contained a high percentage of water, were vacuum-extracted at 80°C and 200 mmHg pressure using a Heidolph rotary evaporator. The goal of this procedure was to remove water and organic acids like acetic acid from high-viscosity oil products. Consequently, oil products with weight ratios of 12.75%, 11.56%, and 10.95% of water and organic acids eliminated were obtained for pine cones, bran, and waste wheat straw, respectively, in comparison to the amount of dry biomass.

Table 2. The results of biomass pyrolysis include yields of liquid bio-oil, biochar, and non-condensable gas

Biomass	Bio-oil yield (%)	Bio-char yield (%)	Gas yield (%)
Pine cone	37.92	32.09	29.99
Olive pomace	31.83	32.52	35.65
Wheat straw	27.13	31.57	41.30

Table 3 shows the analysis results determined by EDS analysis in rejuvenators. However, the elemental compositions determined by EDS analysis are not accepted as quantitative results because the presence of single-electron hydrogen cannot be determined by this technique. In other words, it is not possible to determine the hydrogen content due to the presence of hydrocarbons in the structure of all these materials. On the other hand, it can be considered as an idea about the carbon and oxygen composition other than hydrogen. Another importance of EDS analyses is that they provide information about the presence of elements other than C, H and O.

Table 3. Elemental analysis of rejuvenators

Element	Weight (%)		
	K	P	S
C	56.41	75.86	57.82
O	35.57	24.14	34.81
Al	0.07	-	1.21
N	3.60	-	-
S	0.25	-	-
Fe	4.10	-	-
Ca	-	-	1.06
K	-	-	0.13
Si	-	-	2.20
Mg	-	-	0.49

2.5. Preparation of Hot Mix Asphalt Samples

For combinations made with pure bitumen, the Marshall technique was used to determine the ideal bitumen content (Table 4). Using the Marshall method, 75 blows were delivered to each side of the specimens to create mixture samples. The limestone came from Elazig's Karayazi area. The ideal bitumen concentration was established for mixture samples made entirely of bitumen. Initially, bituminous hot mixes were treated with RAP material at three different percentages (25%, 50%, and 75%) of the overall mixture. They are named 2R, 5R and 7R, respectively, according to the percentages they contain. The same bitumen content was used to prepare modified mixes. For mixtures generated with 4.9% bitumen content, both with pure and modified mixtures, Table 4 gives the bulk specific gravities (Gmb), air voids (Va), voids filled with asphalt (VFA), voids between aggregates (VMA), and mixing-compaction temperatures. Using a rotational viscometer on the binder, the mixing-compaction temperatures of the mixtures were found to correspond to viscosity values of 170 ± 20 and 280 ± 30 cP [35]. In terms of volumetric qualities, it was found that the mixtures satisfied the specification requirement (KTS).

Table 4. Volumetric properties of mixture specimens

Mixture	Bitumen (%)	Mixing temperature (ºC)	Compaction temperature (ºC)	Va (%)	VFA (%)	VMA (%)	Gmb
Unmodified bitumen	5	171	158	3.78	74.3	14.7	2.292

Equation 1 was used to determine the bitumen contents for mixtures containing RAP. The optimum bitumen content determined for the reference bitumen will be used for all mixtures. Table 5 provides the optimum bitumen contents. The purpose of using this ratio is to compare the experimental results accurately.

$$Pr = Pc-(Pa*Pp) \quad (1)$$

Here, Pr represents the percentage of unmodified bitumen to be added to the mixture containing RAP, Pa denotes the percentage of RAP binder in the mixture, Pc indicates the total binder percentage in the mixture, and Pp signifies the percentage of RAP in the mixture.

In the second stage, three different percentages of RAP will be added to HMA samples, and for each of these mixtures, three different rejuvenators will be added during mixing. Different

asphalt mixture specimens (pure/pure+RAP+rejuvenator) have been prepared. The rejuvenator ratio was chosen as 20% according to the study. In the figures, "K" indicates the use of pine cone bio-oil, "P" indicates the use of olive pomace bio-oil, "S" indicates the use of waste wheat straw bio-oil and finally "pure" indicates the use of unaged 50/70 penetration pure binder.

Table 5. Optimum binder contents

RAP content (%)	Pc (%)	Pa (%)	Pr (%)
25	4.9	4.6	3.75
50	4.9	4.6	2.6
75	4.9	4.6	1.45

2.6. Marshall Stability and Flow Test

The Marshall Stability and Flow Test applied to HMA samples is conducted to determine the maximum resistance to deformation (stability) and the vertical deformation (flow) that occurs in the sample when the maximum load is reached. Before starting the test, the heights of the compacted and cooled samples are measured and recorded. Then, the samples are kept in a water bath at a temperature of 60±1°C for 40-60 minutes. At the end of this period, the sample is removed from the water and placed in a manner that aligns the breaking jaws, then loaded at a rate of 50±2 mm/min. Samples subjected to the Marshall Stability and Flow Test are shown in Figure 3.

Figure 3. HMA specimens subjected to Marshall stability and flow test

In the test, the maximum load and the deformation values at the time of reaching this load are recorded. The test should be completed within 40 seconds after the sample is removed from the water. It is assumed that the standard sample height is 63.5 mm in the test. Correction factors for stability are used for samples with different heights. At the end of the test, the average stability and flow values obtained are calculated. Samples with stability values deviating more than 15% from the average stability and flow values deviating more than 20% from the average flow are excluded from the evaluation. The remaining samples are averaged again, and if there are samples with the same deviation (15% for stability, 20% for flow), this sample series is canceled, and the test is repeated on a new set of samples. The Marshall quotient (MQ), serving as an indicator of the stiffness and resistance to deformation of bituminous hot mixtures, is calculated by dividing the Marshall stability value by the flow value.

2.7. Resistance to Moisture Damage Test

Various tests have been devised to assess the resistance of asphalt mixtures to moisture damage. Although these developed tests do not fully reflect real field conditions, they provide us with close information. The adhesion of bitumen to aggregate is one of the significant factors affecting the performance of the pavement over time. Water penetrating between bitumen and aggregate over time reduces the effectiveness of adhesion, leading to premature failure in asphalt mixture overlays. Factors affecting the resistance of asphalt mixtures to moisture damage can be listed as follows:

a) An effective factor in moisture sensitivity is the thickness of the asphalt film. Mixtures with thin asphalt film thickness are more sensitive to the negative effects caused by water compared to mixtures with thicker asphalt film thickness.

b) Another factor affecting the moisture damage of the mixture is the refining process used in the production of asphalt cement and crude oil. However, many asphalt cements do not have a significant impact on moisture damage.

c) Another factor affecting the stripping potential of the asphalt mixture is traffic and environmental indicators.

d) Another factor is the characteristic of the asphalt concrete mixture. Qualities such as permeability, porosity, asphalt cement, and aggregate gradation are important factors because they control the saturation level and drainage of water. Mixtures with more than 6% void content are more prone to moisture damage.

e) The type of coarse and fine aggregate significantly affects the potential moisture damage in the mixture.

In this method, 6 Marshall specimens with an air void content in the range of 6-8% are used. These specimens are divided into conditioned and unconditioned groups. Conditioned specimens are placed in a pycnometer, and a vacuum process is applied until approximately 70-80% of the air voids in the specimen are filled with water. The water-filling ratio in the voids should not be less than 70%. If it exceeds 80%, those specimens should not be used.

The degree of saturation of the specimens is determined using the following formulas:

$$J = B' - B \quad (2)$$
$$I = Va * V / 100 \quad (3)$$
$$S' = J / I * 100 \quad (4)$$

J represents the absorbed water volume in cubic centimeters, B' is the saturated surface-dry weight of the specimen after the vacuum process in grams, B indicates the dry weight of the specimen before the vacuum process in grams, I signifies the volume of air voids in cubic centimeters, Va refers to the air voids percentage, and V represents the volume of the specimen in cubic centimeters.

After the vacuum process, the cylindrical specimens are tightly wrapped with stretch film. These wrapped specimens are initially kept in a water bath at -18°C for 16 hours, followed by another period of 60°C (Figure 4). Once the waiting period is over, the stretch films are removed. Conditioned and unconditioned specimens are then immersed in water at 25°C for 2 hours before being subjected to axial loading for fracture.

Figure 4. Wrapping and incubation of specimens with stretch film

3. Results and Discussion

3.1. Marshall stability and flow test results

Marshall stability and flow test specimens were prepared following EN 12697-34 and submerged in water at 60°C for 40 minutes before undergoing the test. Each asphalt mixture had a constant 20% bio-oil addition. Moreover, Figure 5 shows how the use of bio-oil causes fluctuations in Marshall stability values.

The use of bio-oil lowers the Marshall stability values, as seen in Figure 5. It was discovered that the combinations made with 7R 75% RAP had the highest stability values. The 7P asphalt combination has the lowest stability value. It was found that the bio-oil produced by pyrolyzing olive pomace was the most efficient. In comparison to the pure mixture, the stability values of the 2R, 5R, and 7R mixtures increased by 1.12, 1.17, and 1.57 times, respectively. These findings suggest that the stability values can be raised by including recycled asphalt mixtures in the pure mixture at specific percentages. The stability values of the asphalt mixtures drop and get closer to the stability value of pure mixtures when three distinct rejuvenators are introduced to HMA samples with three different amounts of RAP. The stability ratings of the asphalt mixtures decline as the amount of bio-oil in the mixtures increases. In comparison to the stability value of the 2R asphalt mixture, the Marshall stability values of the 2P, 2K, and 2S mixes drop by 7.83%, 9.60%, and 1.26%, respectively. The rejuvenator added to the 2R mixture is responsible for the observed decline. The Marshall stability values of the 5P, 5K, and 5S asphalt mixtures drop by 22.33%, 13.99%, and 10.50%, respectively, in relation to the stability value of the 5R mixture. In a similar vein, the Marshall stability values of the 7P, 7K, and 7S asphalt mixtures drop by 44.26%, 37.23%, and 43.74%, respectively, with respect to the stability value of the 7R asphalt mixture. It has been determined that P bio-oil is the most useful bio-oil for affecting the stability values of the asphalt mixtures.

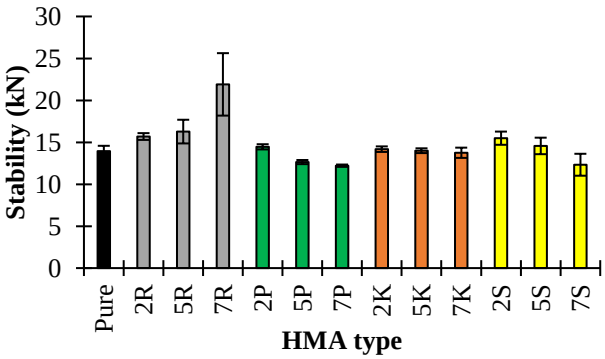


Figure 5. Variation in the stability values of asphalt mixtures with the use of bio-oil

When the flow values of the mixtures given in Figure 6 are examined, it is observed that as the proportion of bio-oil added to the aged binder increases, the flow values of the asphalt mixtures decrease. Among the mixtures, the lowest flow value is obtained from the 7R mixture, while the highest flow value is obtained from the 2R mixture. The mixtures with S bio-oil have lower flow values compared to other mixtures containing bio-oil. Adding bio-oil as a rejuvenator to RAP-containing asphalt mixtures decreases the flow values of these mixtures. The Marshall ratio values of the mixtures are shown in Figure 7.

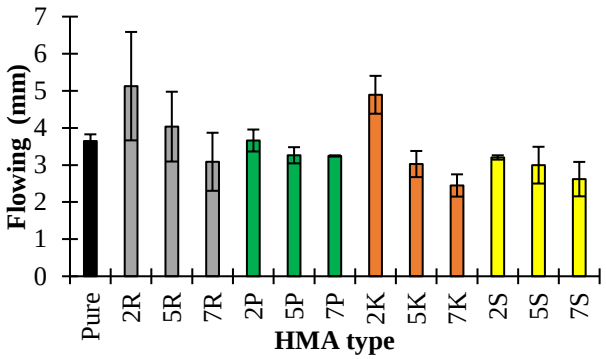


Figure 6. Variation in the flowing values of asphalt mixtures with the use of bio-oil.

When the Marshall quotient (MQ) values given in Figure 7 are examined, it is observed that, except for the K bio-oil, the use of bio-oil generally increases the MQ values. As the proportion of RAP mixtures added to pure mixtures increases, the MQ values also increase. The addition of P and S rejuvenators to the asphalt mixtures decreases the MQ values. Adding rejuvenators to RAP-containing mixtures generally causes a decrease in the MQ values of the mixtures. Evaluating the results of the Marshall stability and flow tests, it is determined that the use of bio-oil generally decreases the stability and MQ values while significantly reducing the flow values.

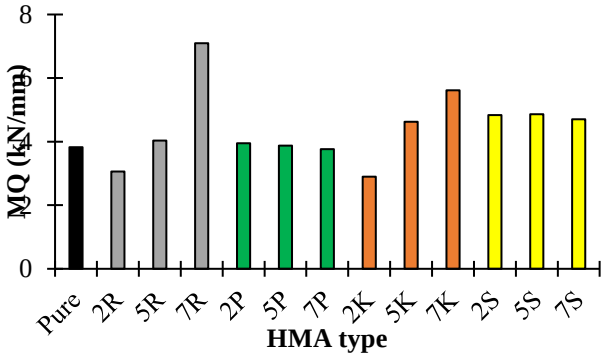


Figure 7. Variation in the MQ values of mixtures with the use of bio-oil

3.2. Tensile strength ratio test results

Tensile strength ratio tests were conducted on mixture samples prepared with rejuvenator and RAP following the AASHTO T 283 standard to determine their resistance to moisture damage. Six samples were prepared for each type of mixture, each having a void content of 7 ± 0.5%. During the preparation of the mixture samples to be used for determining resistance to moisture damage according to the standard, compaction was applied with a gyratory compactor for 40 revolutions. In the short-term aging of the mixtures, the samples were kept in an oven at 60ºC for 16 hours, followed by two hours at the compaction temperature. The samples were then left to cool at room temperature for 24 hours, and their volumetric properties were determined to group them with similar void contents. Three samples from each mixture were subjected to a vacuum to allow them to be filled with water to a saturation level of 70-80%. To achieve this saturation level, all samples underwent conditioning and were subjected to a vacuum for 10 minutes. Samples that did not reach the desired saturation level were subjected to vacuum again. Saturation levels greater than 80% in any sample resulted in their removal from the test and replacement with fresh samples. After being vacuum-treated, the samples were wrapped in cling film, and put into bags with 10 milliliters of water, and the bags were sealed tightly.

The samples went through a particular process: they were submerged in water at 60°C for 24 hours after spending 16 hours in a freezer at -18°C. The samples were then removed from the cling film and bags, and they were immersed in water at 25°C for two hours before breaking. In the case of unconditioned materials, indirect tensile strength jaws were used to fracture them in the Marshall testing apparatus after they had been soaked in water at 25°C for two hours. The samples were loaded at a constant pace of 2 inches per minute (50.8 mm/min) during the breaking process. A selection of the test-submitted samples is shown in Figure 8.

Figure 8. Appearance of some of the samples subjected to the moisture damage test

The tensile strength (TS) values of the mixture samples were assessed using the maximum load values derived from the tensile strength ratio tests performed on both unconditioned and conditioned asphalt hot mix samples, illustrated in Figure 9. As depicted in the figure, the tensile strength

(TS) values of both unconditioned and conditioned mixtures exhibited an increase with the addition of recycled asphalt compared to the pure (reference) mixture. Incorporating 20% bio-oil into these mixtures led to a reduction in the tensile strength (TS) values of both unconditioned and conditioned mixtures in comparison to the RAP mixture. With a higher proportion of bio-oil in the asphalt mixtures, there is a noticeable decrease in the tensile strength (TS) values observed in both unconditioned and conditioned mixtures. As the content of bio-oil increases, the tensile strength ratio diminishes, whereas mixtures prepared with RAP demonstrate an enhancement in the tensile ratio. This observation indicates that in unconditioned mixtures, the 7R mixture has the highest tensile strength ratio, whereas the 7K mixture has the lowest tensile strength ratio. Tensile strength (TS) values of the 2R, 5R, and 7R mixtures were found to rise 1.02, 1.07, and 1.22 times, respectively, above the pure mixture's TS values prior to conditioning. In contrast, when compared to the unconditioned TS value of the 2R asphalt mixture, the unconditioned TS values of the 2P, 2K, and 2S mixtures dropped by 8.22%, 14.14%, and 11.91%, respectively. The rejuvenator that was added to the 2R mixture is responsible for the reported decrease. In particular, when compared to the unconditioned TS value of the 5R asphalt combination, the unconditioned TS values of the 5P, 5K, and 5S mixtures dropped by 20.08%, 33.28%, and 23.39%, respectively. Comparing the unconditioned TS value of the 7R asphalt mixture to that of the 7P, 7K, and 7S mixtures, respectively, showed a decrease of 34.94%, 39.43%, and 31.96%. Following conditioning, it was found that, in comparison to the conditioned TS values of the pure mixture, the TS values of the 2R, 5R, and 7R mixes increased by 1.024, 1.04, and 1.34 times, respectively. In comparison to the conditioned TS value of the 2R asphalt mixture, the unconditioned TS values of the 2P, 2K, and 2S mixtures dropped by 1.32%, 9.80%, and 7.16%, respectively. In addition, when compared to the conditioned TS value of the 5R asphalt combination, the conditioned TS values of the 5P, 5K, and 5S mixtures dropped by 14.63%, 20.60%, and 18.75%, respectively. Comparing the conditioned TS value of the 7R asphalt mixture to that of the 7P, 7K, and 7S mixtures, respectively, showed a decrease of 36.02%, 44.56%, and 39.71%. Figure 10 shows the tensile strength ratios (TSR) of the mixture samples using samples of conditioned and unconditioned asphalt hot mix.

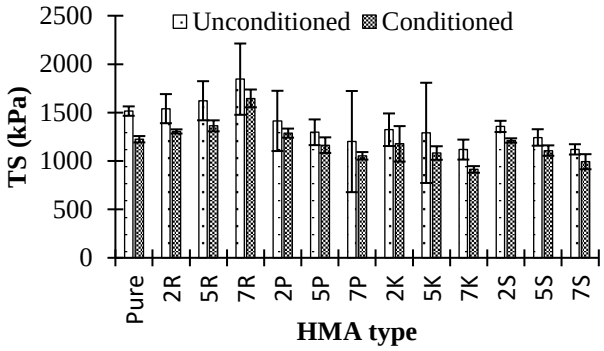


Figure 9. Tensile strength values of mixtures before and after conditioning

The addition of RAP to the mixes improved their resistance to moisture damage, as evidenced by the tensile strength ratios (Figure 10), which act as markers of resistance to moisture damage. Increasing the amount of RAP in the combinations often strengthened their resistance to moisture degradation. However, increasing the percentage of each of the three bio-oils in the mixtures led to a reduction in their ability to withstand moisture degradation. Tensile strength ratio (TSR) values of hot bituminous mixtures should preferably be above 80%, as per the Superpave method, to protect against moisture damage. The TSR values of all the created mixes were found to be either near or above 80% when the acquired values were examined.

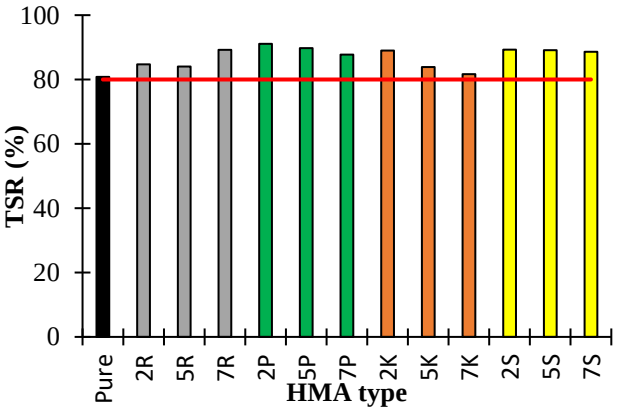


Figure 10. Variation of the tensile strength ratios of the mixtures

4. Conclusion

In this study, to meet the properties of the selected reference bitumen (B50/70-grade pure bitumen was chosen as the reference binder for the rejuvenation of recycled asphalt (RAP) binder), three different rejuvenators were added in varying proportions to the aged bitumen obtained from RAP, and the physical and rheological properties of the binders were examined. Physical and mechanical tests were

conducted on the prepared HMA (Hot Mix Asphalt) samples to assess the impacts of the rejuvenators and RAP (Recycled Asphalt Pavement) material on the physical and mechanical properties of the reference mixtures.

• It was determined that the aged bitumen obtained from the recycled bituminous mixture softened as a result of the addition of rejuvenators.

• It has been determined that adding recycled asphalt mixtures to the pure mixture at certain percentages increases the stability values. When three different rejuvenators are added during the mixing process to HMA samples containing three different proportions of RAP, a decrease in the stability values of the asphalt mixtures occurs, approaching the stability value of the pure mixtures. As the bio-oil content in the asphalt mixtures increases, the stability values of the mixtures decrease.

• Adding recycled asphalt to hot bituminous mixtures led to an increase in the tensile strength (TS) values of both unconditioned and conditioned mixtures compared to the pure (reference) mixture. However, when 20% bio-oil was introduced to these mixtures, the tensile strength (TS) values of both unconditioned and conditioned mixtures decreased in comparison to the RAP mixture. As the proportion of bio-oil in the asphalt mixtures increased, the tensile strength (TS) values decreased in both unconditioned and conditioned mixtures. Additionally, while the increase in bio-oil content reduced the tensile strength ratio, the tensile strength of the mixtures prepared with RAP increased.

Consequently, it may be said that using bio-oil lessens stiffness. Because of this, choosing the right dosage needs to be done carefully to prevent rutting resistance from being compromised and eventual failure. According to the findings of this study, bio-oils derived from pine cones, olive pomace, and wheat straw are viable rejuvenators. Among the rejuvenators investigated, the bio-oil obtained from olive pomace was identified as the most effective.

Contributions of the authors

M.Y and B.F.Y. designed the study, performed the experiments and wrote the article. H.S.A. performed the calculations, checked the language and contributed to the writing of the manuscript.

Acknowledgment

This study was carried out within the scope of the Scientific and Technological Research Council of Turkey (TUBITAK). We gratefully acknowledge the financial support provided by TUBITAK to Research Project 122M042.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study complies with research and publication ethics.

References

[1] E. Chailleux, M. Audo, B. Bujoli, C. Queffelec, J. Legrand, and O. Lepine, “Workshop alternative binders for sustainable asphalt pavements,” in *Alternative Binder from Microalgae: Algoroute Project*, 2012, pp. 7–14.

[2] R. Zhang, H. Wang, J. Gao, Z. You, and X. Yang, “High temperature performance of SBS modified bio-asphalt,” *Constr. Build. Mater.*, vol. 144, pp. 99–105, 2017.

[3] S. Lv, C. Liu, H. Yao, and J. Zheng, “Comparisons of synchronous measurement methods on various moduli of asphalt mixtures,” *Constr. Build. Mater.*, vol. 158, pp. 1035–1045, 2018.

[4] K. W. Lee, T. E. Brayton, M. Mueller, and A. Singh, “Rational mix-design procedure for cold in-place recycling asphalt mixtures and performance prediction,” *J. Mater. Civ. Eng.*, vol. 28, no. 6, p. 04016008, 2016.

[5] A. Onochie, E. Fini, X. Yang, J. Mills-Beale, and Z. You, “Rheological Characterization of Nano-particle based Biomodified Binder,” in *Transportation research board 92nd annual meeting*, Washington DC, 2013.

[6] R. Zhang, H. Wang, J. Gao, X. Yang, and Z. You, "Comprehensive performance evaluation and cost analysis of SBS-modified bioasphalt binders and mixtures," *J. Mater. Civ. Eng.*, vol. 29, no. 12, p. 04017232, 2017.

[7] D. Ge, K. Yan, Z. You, and H. Xu, "Modification mechanism of asphalt binder with waste tire rubber and recycled polyethylene," *Constr. Build. Mater.*, vol. 126, pp. 66–76, 2016.

[8] R. Zhang, H. Wang, X. Jiang, Z. You, X. Yang, and M. Ye, "Thermal storage stability of bio-oil modified asphalt," *J. Mater. Civ. Eng.*, vol. 30, no. 4, p. 04018054, 2018.

[9] H. Wang, R. Zhang, Y. Chen, Z. You, and J. Fang, "Study on microstructure of rubberized recycled hot mix asphalt based X-ray CT technology," *Constr. Build. Mater.*, vol. 121, pp. 177–184, 2016.

[10] C. A. Bell, A. J. Wieder, and M. J. Fellin, *Laboratory aging of asphalt-aggregate mixtures: Field validation. No. SHRP-A-390*. 1994.

[11] S. Said, "Aging effect on mechanical characteristics of bituminous mixtures," *Transp. Res. Rec.*, vol. 1901, pp. 1–9, 2005.

[12] K. Z. Ramadan and A. A. Saad, "Effect of Superpave short-term aging on binder and asphalt mixture rheology," *Period. Polytech. Transp. Eng.*, vol. 45, no. 4, p. 196, 2017.

[13] G. Holleran, T. Wieringa, and T. Tailby, "Rejuvenation treatments for aged pavements," *Transit New Zealand and New Zealand Institute of Highway Technology*, 2006.

[14] R. H. Zhang, T. S. Zhao, P. Tan, M. C. Wu, and H. R. Jiang, "Ruthenium dioxide-decorated carbonized tubular polypyrrole as a bifunctional catalyst for non-aqueous lithium-oxygen batteries," *Electrochim. Acta*, vol. 257, pp. 281–289, 2017.

[15] S. N. Nahar, "Turning back time: rheological and microstructural assessment of rejuvenated bitumen," *Transportation Research Record*, vol. 2444, pp. 52–62, 2014.

[16] X. Yu, M. Zaumanis, S. dos Santos, and L. D. Poulikakos, "Rheological, microscopic, and chemical characterization of the rejuvenating effect on asphalt binders," *Fuel (Lond.)*, vol. 135, pp. 162–171, 2014.

[17] M. Zaumanis, R. B. Mallick, L. Poulikakos, and R. Frank, "Influence of six rejuvenators on the performance properties of Reclaimed Asphalt Pavement (RAP) binder and 100% recycled asphalt mixtures," *Constr. Build. Mater.*, vol. 71, pp. 538–550, 2014.

[18] Z. Xie, N. Tran, G. Julian, A. Taylor, and L. D. Blackburn, "Performance of asphalt mixtures with high recycled contents using rejuvenators and warm-mix additive: Field and lab experiments," *J. Mater. Civ. Eng.*, vol. 29, no. 10, p. 04017190, 2017.

[19] N. Tran *et al.*, "Effect of a recycling agent on the performance of high-RAP and high-RAS mixtures: Field and lab experiments," *J. Mater. Civ. Eng.*, vol. 29, no. 1, p. 04016178, 2017.

[20] Z. Lei, H. Bahia, and T. Yi-qiu, "Effect of bio-based and refined waste oil modifiers on low temperature performance of asphalt binders," *Constr. Build. Mater.*, vol. 86, pp. 95–100, 2015.

[21] X. Yang, Z. You, and Q. Dai, "Performance evaluation of asphalt binder modified by bio-oil generated from waste wood resources," *International Journal of Pavement Research and Technology,* vol. 6, pp. 431–439, 2013.

[22] Y. Li, B. Xing, Y. Ding, X. Han, and S. Wang, "A critical review of the production and advanced utilization of biochar via selective pyrolysis of lignocellulosic biomass," *Bioresour. Technol.*, vol. 312, no. 123614, p. 123614, 2020.

[23] S. López, D. R. Davies, F. J. Giráldez, M. S. Dhanoa, J. Dijkstra, and J. France, "Assessment of nutritive value of cereal and legume straws based on chemical composition and in vitro digestibility," *Journal of the Science of Food and Agriculture,* vol. 85, no. 9, pp. 1550–1557, 2005.

[24] M. Wang, N. S. Hettiarachchy, M. Qi, W. Burks, and T. Siebenmorgen, "Preparation and functional properties of rice bran protein isolate," *Journal of agricultural and food chemistry*, vol. 47, no. 2, pp. 411–416, 1999.

[25] J. Wang, B. Sun, Y. Cao, and C. Wang, "In vitro fermentation of xylooligosaccharides from wheat bran insoluble dietary fiber by Bifidobacteria," *Carbohydr. Polym.*, vol. 82, no. 2, pp. 419–423, 2010.

[26] I. Ortiz de Zárate, A. Ezcurra, J. P. Lacaux, P. Van Dinh, and J. D. de Argandoña, "Pollution by cereal waste burning in Spain," *Atmos. Res.*, vol. 73, no. 1–2, pp. 161–170, 2005.

[27] L. Rzˇek, M.R. Turk, M. Tušar, "Increasing the rate of reclaimed asphalt in asphalt mixture by using alternative rejuvenator produced by tire pyrolysis," *Constr. and Build. Mater.*, 232, 117177, 2020.

[28] L. Avsenik, D. Klinar, M. Tušar, L.S. Perše, "Use of modified slow tire pyrolysis product as a rejuvenator for aged bitumen," *Constr. and Build. Mater.*, 120, 605–616, 2016.

[29] J. Akhtar, S. K. Kuang, and N. S. Amin, "Liquefaction of empty palm fruit bunch (EPFB) in alkaline hot compressed water," *Renew. Energy,* vol. 35, no. 6, pp. 1220–1227, 2010.

[30] A. Bridgwater, "Fast pyrolysis processes for biomass," *Renew. Sustain. Energy Rev.*, vol. 4, no. 1, pp. 1–73, 2000.

[31] Ş. Oruç, B. Yılmaz, and M. S. Mazlum, "Geri kazanılan asfalt kaplamaların sıcak asfalt karışımlarda yeniden kullanabilirliğinin araştırılması," *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, vol. 30, pp. 87–93, 2018.

[32] A. Kumandaş, E. Çavdar, Ş. Oruç, E. B. Pancar, and B. V. Kök, "Effect of WCO addition on high and low-temperature performance of RET modified bitumen," *Constr. Build. Mater.*, vol. 323, no. 126561, p. 126561, 2022.

[33] ″Agregaların genel özellikleri için deneyler-Kısım 1 numune alma metotları″. TS EN 932-1. 1997. (In Turkish)

[34] "Bitumlu karisimlar - Deney yontemleri - Bolum 3: Bitum geri kazanilmasi: Doner buharlastirici yontemi," TS EN 12697-3+A1., 2019. (In Turkish)

[35] B. F. Yalçın and M. Yilmaz, "Investigation of the performance of bio-oils from three different agricultural wastes as rejuvenators for recycled asphalt," *Turkish Journal of Civil Engineering*, vol. 35, no. 3, pp. 95–123, 2024.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 999-1012 YEAR: 2024
DOI: 10.17798/bitlisfen.1490837

Development of a Functionalized SiO_2 Supported Ni Nanoparticles Based Non-Enzymatic Amperometric Dopamine Sensor

Yunus Emre YILDIRIM[1], Muhammet GÜLER[1*]

[1]*Van Yüzüncü Yıl University, Faculty of Science, Department of Chemistry, Van, Turkiye (ORCID: 0009-0004-5813-8640) (ORCID: 0000-0002-1040-8988)*

Keywords: Sensor, Functionalized SiO_2, Ni nanoparticles, Dopamine

Abstract

In the present work, a novel electrochemical dopamine (DA) sensor depending on Nickel (Ni) nanoparticles decorated (3-aminopropyl)triethoxysilane (APTES) modifed silica (SiO_2) was fabricated. Hence, Ni@SiO_2-APTES was synthesized by the conventional wet-impregnation method. The structure of the compozite was evaluated using Fourier transform infrared spectroscopy (FTIR), Scanning electron microscopy-energy dispersive X-ray (SEM-EDX), and X-ray diffraction (XRD). The synthesized Ni@SiO_2-APTES was loaded on glassy carbon working electrode (GCE). Also, Nafion (Nf) was drop-casted on Ni@SiO_2-APTES/GCE to stabilize the electrode. The fabricated Nf/Ni@SiO_2-APTES/GCE working electrode was electrochemicaly evaluated using cyclic voltammetry (CV), electrochemical impedance spectroscopy (EIS) and amperometry. CV and EIS results indicated that Ni nanoparticles increased both the conductivity and sensitivity of the working electrode. The linear detection range for DA was found to be 0.2 – 252 µM with limit of detection (LOD) was 0.07 µM depending on S/N of 3. The sensitivity was found to be 578.26 µA mM^{-1} cm^{-2} depending on the active surface area of the modified working electrode. The sensor exhibited excellent selectivity in the electrolyte solution including ascorbic acid, glucose, fructose, sucrose, mannose, uric acid, and phenylalanine. The sensor had satisfactory repeatability and reproduciblity. It was observed that the sensor showed an electrocatalytic response of 95.33% after 28 days. According to this result, it was concluded that the sensor was extremely stable within the studied time period. The applicability of Nf/Ni@SiO_2-APTES/GCE was tested using dopamine HCl injection (200 mg/5 mL).

1. Introduction

Enzyme-based electrochemical biosensors appear to be ideal materials for the determination of target analytes due to their high selectivity. However, in the in vitro environment, factors such as temperature, pH and binding of the biomolecule to the surface of the electrode can shorten the working life of the biosensor. Further studies are needed to increase the stability of such biosensors, and new fabrication techniques should be developed for easier preparation of biosensors. Contrary to enzyme-based electrochemical sensors, non-enzymatic sensors have high stability, satisfactory selectivity and are relatively easy to prepare. Non-enzymatic electrochemical sensors have been shown to remain electrochemically active in undiluted blood for more than 30 days, which is not possible with enzyme-based sensors. However, there are fundamental difficulties such as the use of alkaline working environments especially for the determination of analytes such as glucose and the expensiveness of the materials used [1].

Dopamine, known as [4-(2-aminoethyl) benzene-1,2-diol], is a significant catecholamine substance widely secreted in the cerebrospinal fluid of mammals and the main member of the neurotransmitter's family. Dopamine has a significant impact in cardiovascular functions, hormonal, renal and central nervous systems. High or low dopamine levels can cause diseases such as tourette syndrome [2], schizophrenia [3], hyperactivity, attention deficit [4] and tumor formation [5]. It has been concluded that insufficient dopamine levels in the brain can lead

*Corresponding author: mguler@yyu.edu.tr
Received: 27.05.2024, Accepted: 11.12.2024

to involuntary muscle activity in Parkinson's disease [6], while excessive amounts can lead to Huntington's disease [7]. Given the critical function of this substance in the metabolism, a lot of efforts have been made to identify this biogenic amine. Dopamine has been determined in biological fluids and pharmaceutical drugs using a variety of methods, including electrochemical, UV-visible spectrophotometry, high-performance liquid chromatography (HPLC), liquid chromatography/mass spectrometry (LC/MS), fluorometry, and capillary zone electrophoresis [8], [9]. Electrochemical sensors/biosensors are one of these techniques with numerous benefits. For example, they are cheap, easy to prepare, not complex systems, and exhibit rapid responses [10]. Electrochemical biosensors are not complicated to install and costly to fabricate compared to other measurement techniques such as HPLC and LC/MS. The reason for this can be considered as the production of relatively low-cost microelectronic circuits and the fabrication of interfaces that can easily respond to the target analyte. In addition, other advantages of electrochemical sensors include their robustness, easy miniaturization, excellent response to even small analyte concentrations, and the ability to provide this response within seconds [11], [12].

For the design of electrochemical sensors, nanocomposites have been prepared by deposition of nanoparticles such as metal and metal oxide on supports including carbon nanotube [13], graphene oxide, reduced graphene oxide [14], and SiO_2 [15].

By modifying the working electrodes with these nanocomposites, the determination of electrochemically active substances like hydrogen peroxide, dopamine, ascorbic acid, folic acid and paracetamol was achieved [16], [17]. However, the performance factors such as LOD, linear determination range, especially repeatability, reproducibility, and storage stability of these sensors need to be improved.

Silica (SiO_2), which has a large surface area, was used as a support material, allowing metal and metal oxide nanoparticles to be dispersed uniformly on the surface [15], [18]. Surface-modified silica can enhance the electrocatalytic performance of metal or metal oxide nanoparticles. For example, in a study performed by Bayram and Guler, an electrochemical hydrogen peroxide sensor depending on Au nanoparticles deposited on SiO_2-APTES was designed. The SiO_2-APTES support provided an excellent surface for the deposition of Au nanoparticles on the surface. Because both low diameter of Au nanoparticles (mean particle size of 3.05 ± 0.14 nm) were deposited on the surface and it was determined that the nanoparticles were homogeneously distributed on the support. A very high sensitivity for hydrogen peroxide (2514.6 μA mM^{-1} cm^{-2} between 0.014 and 0.18 mM and 894.2 μA mM^{-1} cm^{-2} between 0.18 and 7.15 mM) was achieved depending on the conductivity and size of the Au nanoparticles deposited on the support [17].

The purpose of the present work is to prepare an electrochemical DA sensor using Ni@SiO_2-APTES modified glassy carbon electrode (GCE). The specific aims are (i) to synthesize Ni@SiO_2-APTES, (ii) to modify GCE using Ni@SiO_2-APTES and, (iii) to detect performance factors such as linear range, LOD, and selectivity for DA using amperometry and cyclic voltammetry.

2. Material and Method

Silica (SiO_2, 240-400 mesh), ethanol, anhydrous ethanol, and sodium borohydride ($NaBH_4$) were purchased from Merck Company. (3-Aminopropyl)triethoxysilane (APTES), dopamine hydrochloride (DA), nickel (II) nitrate hexahydrate $Ni(NO_3)_2.6H_2O$, nafion (Nf) solution (5% in a mixture of aliphatic alcohol and water), potassium chloride (KCl) , glucose (Glu), sucrose (Suc), fructose (Fruc), Monnose (Man), ascorbic acid (AA), uric acid (UA), toluene, boric acid (H_3BO_3), acetic acid (CH_3COOH), phosphoric acid (H_3PO_4), acetone, sodium hydroxide (NaOH), hydrochloric acid (HCl) were acquired from Sigma-Aldiric Company and All other chemicals used were purchared from Merck Company. In the electrochemical studies, glassy carbon electrode (GCE of 3 mm diameter) was used as the working electrode, Ag/AgCl (3 M KCl) as the reference electrode, and platinum wire as the counter electrode was purchased from BASİ company. For the electrochemical studies, Autolab brand PGSTAT128N model potentiostat/galvanostat device containing the FRA 32M module was used.

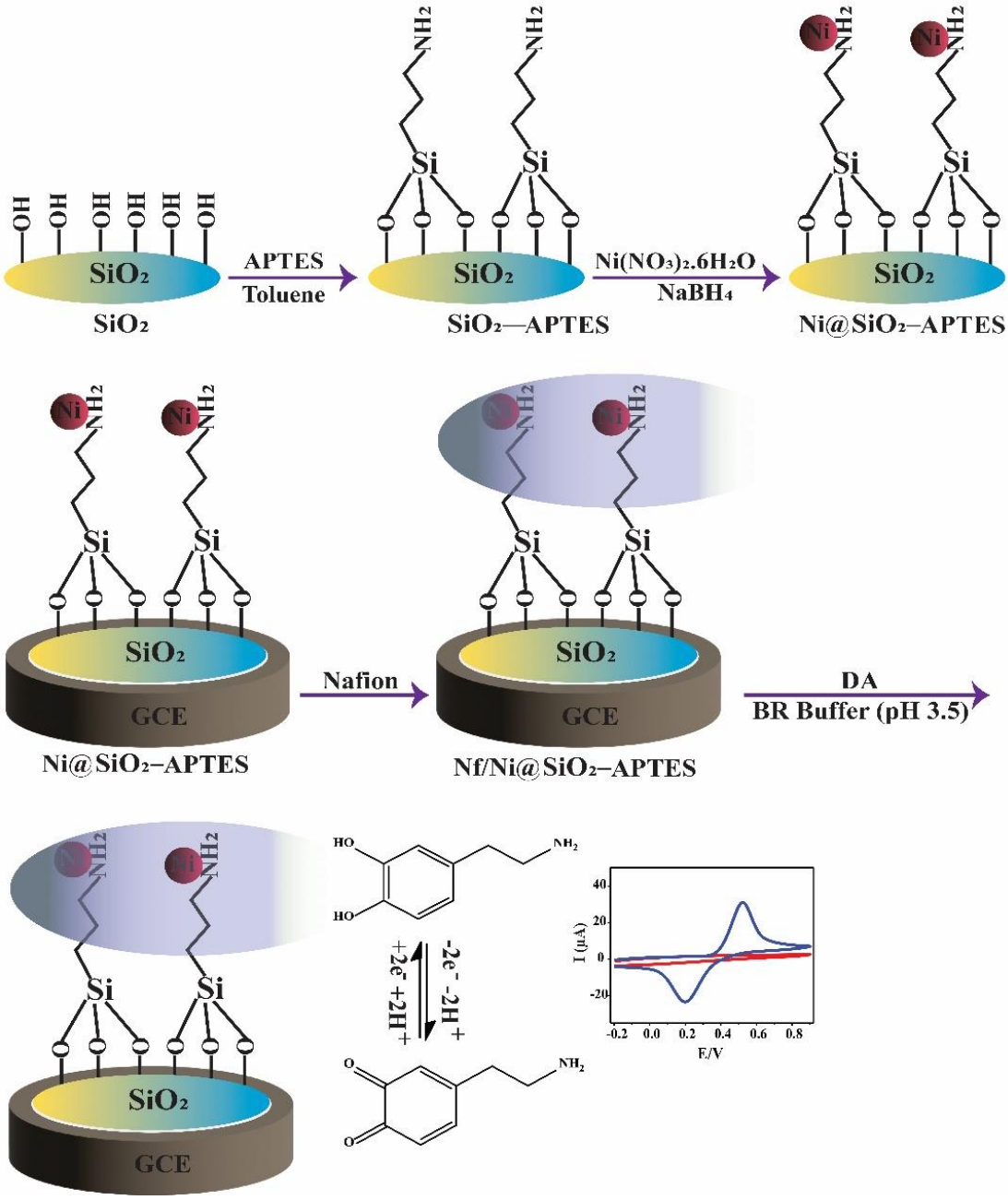
The characterization of the samples was monitored by Field emission scanning electron microscope (FE-SEM; Zeiss sigma 300) and Fourier transform infrared spectroscopy (FTIR) at Van Yüzüncü Yıl University Science Application and Research Center. X-ray diffraction (XRD) was performed at Selçuk University Advanced Technology Research and Application Center.

2.1. Synthesis of SiO_2-APTES and Ni@SiO_2-APTES composites

Before attached APTES to SiO_2, SiO_2 was dried in an oven (JP Selecta S. A., Spain**)** at 100 °C. 2 g of dried SiO_2 was included in 30 mL of pure toluene and

stirred for a while. Then, 5.4 mL APTES was included in the mixture and it was stirred for 24 hours. After the reaction was completed, the mixture was filtered and cleaned down using toluene (5x20 mL). The washed SiO_2-APTES was dried in the oven at 100 °C for 1 hour [19].

Ni@SiO_2-APTES was synthesized by the conventional wet-impregnation method. For this, 0.1 g support was put in 5 mL ultra-pure water and mixed in an ultrasonic cleaner for 10 min. Then, 26.06 mg $Ni(NO_3)_2x6H_2O$ (5% Ni) was put into the above mixture and stirred for 3 hours using a magnetic stirrer. 50.84 mg $NaBH_4$ (1.344mmol) was included in the mixture and it was stirred until the reduction was completed. The resulting Ni@SiO_2-APTES nano composite was filtrated and then washed with ethanol (3x20 mL) and water (3x20 mL), respectively, and dried in the oven at 100 °C. To achieve the best sensor response, Ni@SiO_2-APTES was obtained by using different ratio of Ni salt (2, 5, 10, 15 and 20%) [20]. The fabrication of Nf/Ni@SiO2-APTES/GCE working electrode and its electrochemical response to DA is given step by step in Scheme 1.



Scheme 1. Step by step fabrication of Nf/Ni@SiO_2-APTES/GCE working electrode and its electrochemical response to DA.

2.2. Fabrication of working electrodes

In the study, the working electrode, the GCE, was polished using Al_2O_3 slurry before being modified. Afterwards, it was cleaned using nitric acid/pure water prepared in the ratio of (1:1) and acetone in an ultrasonic bath. 5 µL of 2 mg/mL Ni@SiO_2-APTES mixture was dropped onto the GCE. The fabricated electrode was dried in the laboratory temperature. Then, 3 µL of 0.3% Nf solution was loaded on the electrode and the electrode was left to dry [17]. The obtained Nf/Ni@SiO_2-APTES/GCE was used for further electrochemical studies.

3. Results and Discussion

The structural properties of the obtained SiO_2-APTES were evaluated using FTIR. Figure 1 shows the FTIR spectrum of SiO_2 and SiO_2-APTES. Figure 1a shows the general FTIR peaks of the SiO_2 compound. The peaks at 794, 972 and 1056 cm^{-1} show the Si-O symmetric bending vibration peak, Si-O-(H-H_2O) bending vibration peak and Si-O-Si asymmetric vibration peak, respectively. Figure 1b shows the FTIR spectrum of SiO_2-APTES composite. As seen in the figure, the peaks occurring at 696, 790, 1062, 1392, 1537, 2902 and 2978 cm^{-1} show the characteristic vibration peak of the Si-CH_2 group, symmetric bending vibration peak of Si-O, asymmetric vibration peak of Si-O-Si, the stretching vibration peak of C-N, the bending vibration peak of the N-H group, and the stretching vibration peak of the C-H groups, respectively. Additionally, it can be seen that the peak at 972 cm^{-1} disappears after APTES is attached to the SiO_2. This is due to the reaction that occurs between the ethoxy groups in the APTES and Si-OH groups [21], [22].

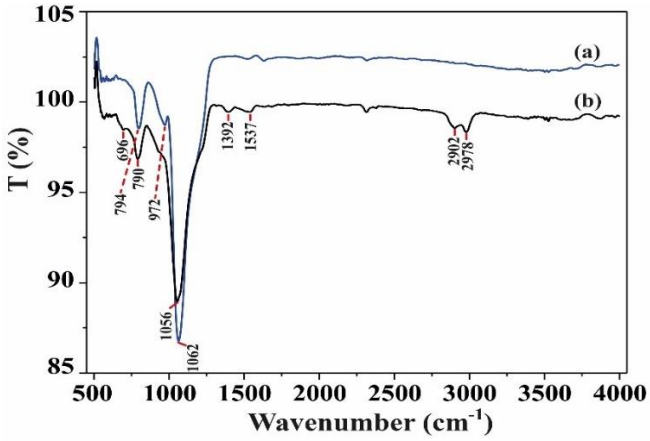


Figure 1. FTIR spectrum of (a) SiO_2 and (b)SiO_2-APTES.

The crystal structure of SiO_2-APTES and Ni@SiO_2-APTES composites was investigated using XRD (Figure 2). Figure 2a and Figure 2b show the XRD spectra of SiO_2-APTES and Ni@SiO_2-APTES structures, respectively. As seen in Figure 2b, even though 5% nickel salt was used, a weak peak was formed at about 2θ = 45°. In this study, $NaBH_4$, a good reducing agent, was used to reduce Ni nanoparticles on SiO_2-APTES support. This shows that the Ni-B amorphous structure forms by sodium borohydride with Ni at room temperature [23], [24].

SEM-EDX was used to investigate the surface properties of SiO_2-APTES and Ni@SiO_2-APTES composites. Figures 3A and 3B show SEM pictures of the synthesized SiO_2-APTES support obtained at 30 µm and 200 nm. Figures 3C and 3D display SEM images of the Ni@SiO_2-APTES obtained at 400 and 200 nm. Figure 4A exhibits the SEM image of the Ni@SiO_2-APTES composite, and the energy dispersion X-ray (EDX) spectra of the selected region is seen in Figure 4B. The weight of the Ni obtained in the spectrum is 6.78%, and this value is close to the 5% value, which is the weight percentage of Ni in the Ni salt used to synthesize the Ni@SiO_2-APTES composite in the experimental section. The presence of carbon and nitrogen in the EDX graph shows that the APTES compound was bound on SiO_2 support. Figures 4C/D/E/F show the elemental mapping images of the composite.

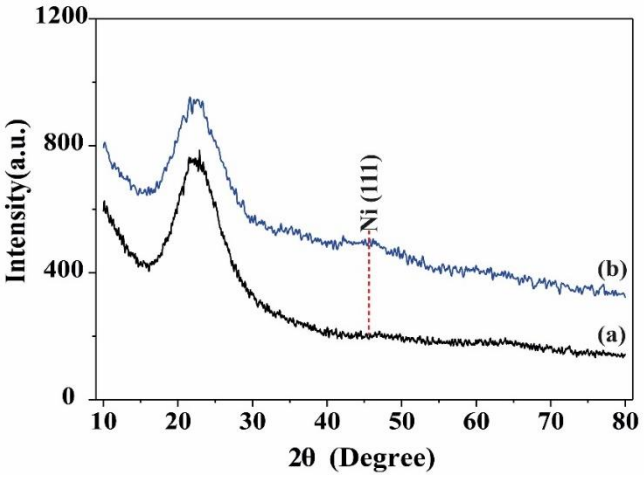


Figure 2. XRD graph of (a) SiO_2-APTES and (b) Ni@SiO_2-APTES.

CV and EIS methods were performed to evaluate the electrocatalytic properties of the working electrodes and evaluate their electrocatalytic responses to DA. For this purpose, firstly, CVs at GCE, Nf/SiO_2-APTES/GCE and Nf/Ni@SiO_2-APTES/GCE working electrodes were obtained in BR buffer (pH 3.5) in the absence and presence

of DA at the scan rate of 50 mV/s and the potential range of -0.2 to 0.8 V (Figure 5A and 5B). As can be seen in Table 1, it was displayed that the anodic peak current of the Nf/Ni@SiO_2-APTES/GCE was approximately 1.5 times the anodic peak current of the Nf/SiO_2-APTES/GCE. The rate of oxidation and reduction peak currents of the Nf/Ni@SiO_2-APTES/GCE was calculated as I_{pa}/I_{pc} = 1.076. This shows that a quasi-reversible electrode process occurs on the Nf/Ni@SiO_2-APTES/GCE electrode. In additon, as illustrated in Figure 6C, an increase in scan rate results in an increase in the redox peak potential difference (ΔEp). The increment in the peak potential difference depending on the scan rate indicates a quasi-reversible reaction process [9].

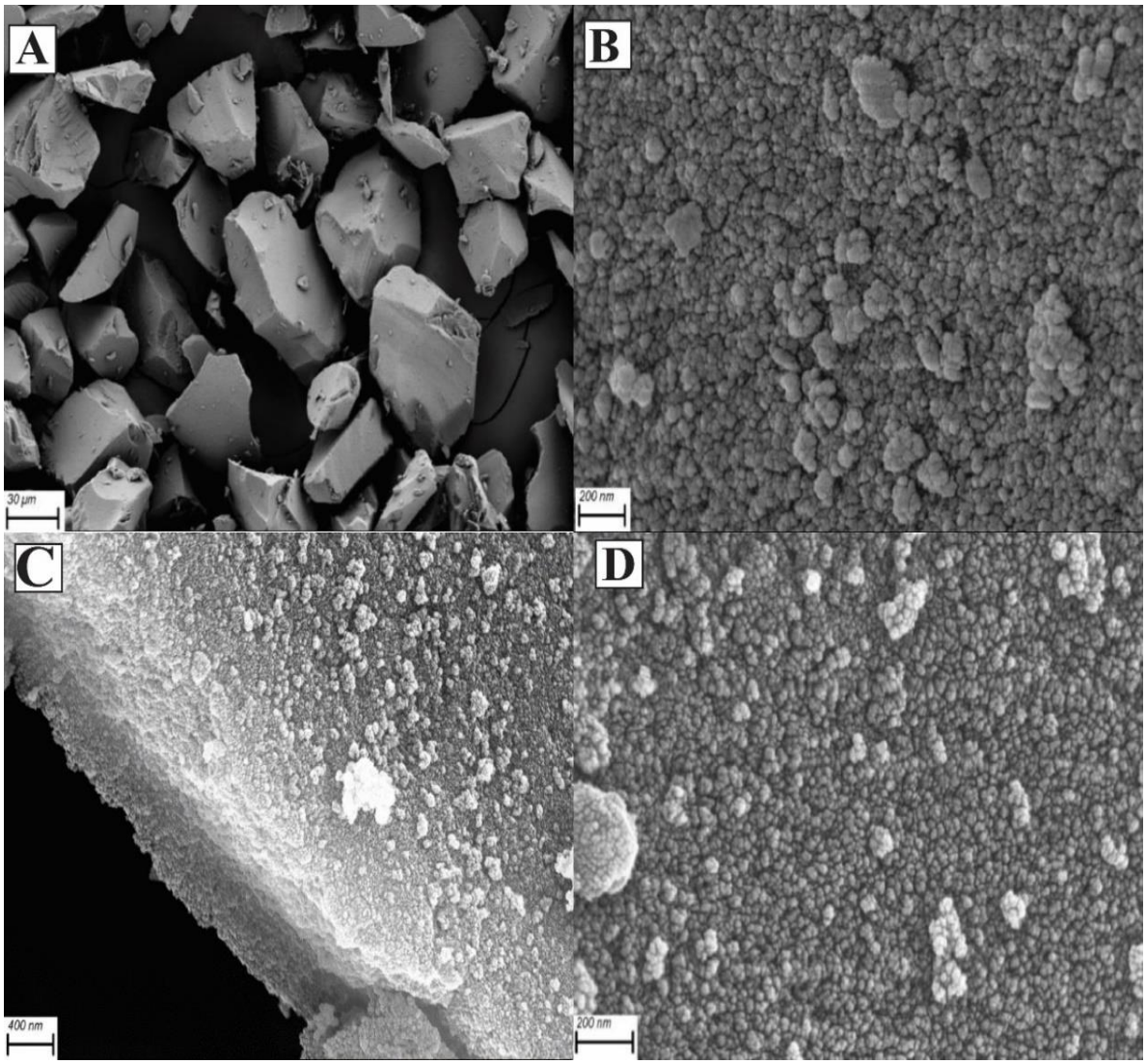


Figure 3. SEM images of (A, B) SiO_2-APTES and (C, D) Ni@SiO_2-APTES.

Table 1. Some electrochemical response data of working electrodes to DA.

Electrode	i_{pa}/µA	i_{pc}/µA	E_{pa}/V	E_{pc}/V	i_{pa}/i_{pc}	ΔE_p
GCE	2.53	-2.26	0.564	0.235	1.119	0.329
SiO_2-APTES	6.74	-6.59	0.474	0.247	1.023	0.227
Ni@SiO_2APTES	9.69	-9.01	0.466	0.252	1.076	0.214

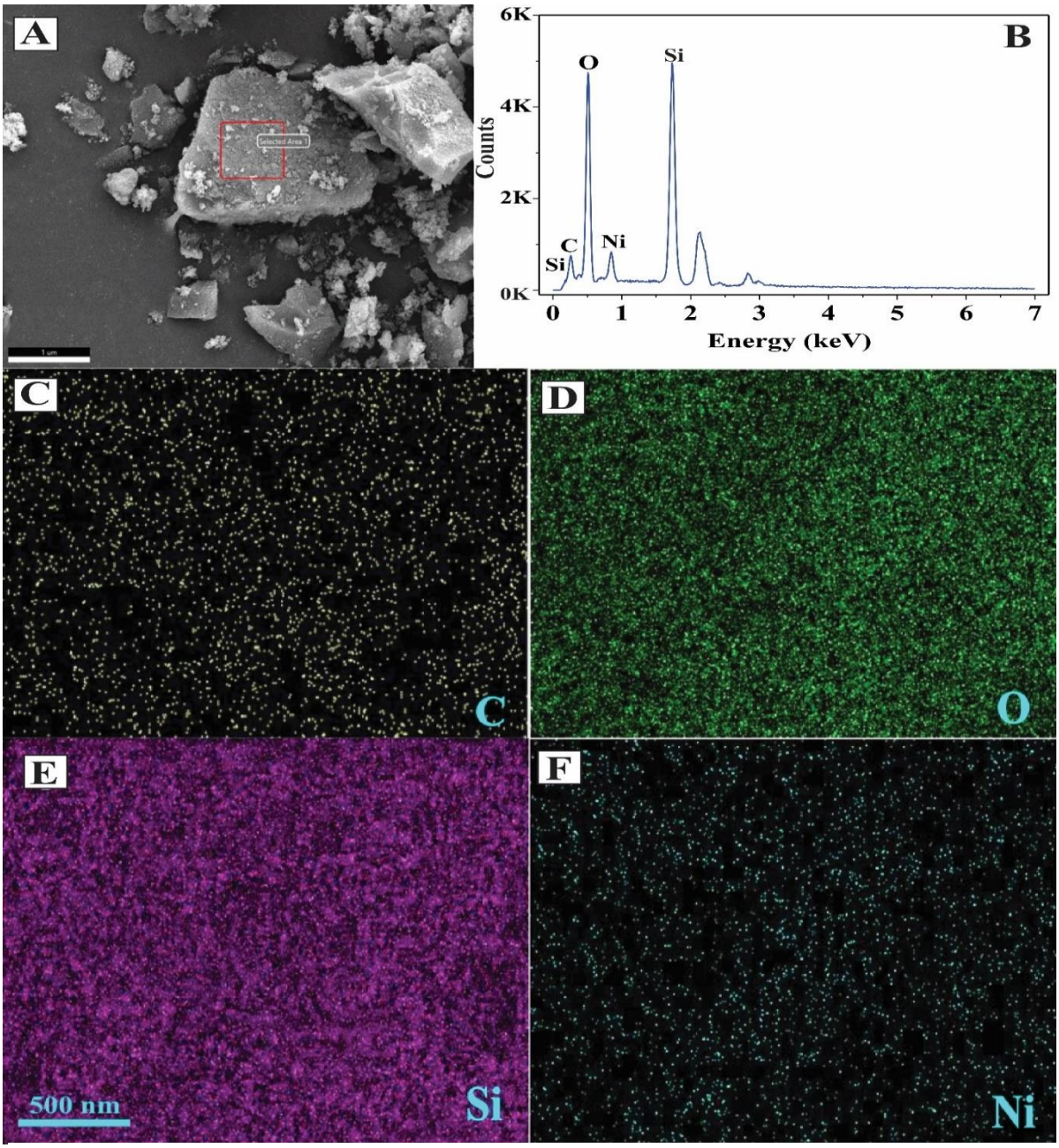


Figure 4. SEM image of (A) Ni@SiO_2-APTES, EDX of Ni@SiO_2-APTES, elemental mapping pictures of carbon (C), oxygen (D), silisium (E), and nickel (F).

Figure 5C shows the typical EIS Nyquist plot for GCE, Nf/SiO_2-APTES/GCE, and Nf/Ni@SiO_2-APTES/GCE electrodes. Nyquist curves were obtained in a solution containing 5 mM $[Fe(CN)_6]^{3-}$/$[Fe(CN)_6]^{4-}$ and 0.1M KCl. The used frequency interval and potential value were 0.1 Hz-100 kHz and 0.2 V, respectively. As shown in the figure, the Nyquist plot contains a semicircular region at high frequency and a linear domain at low frequency. In the resulting graph, the charge transfer resistance is equal to the semicircle's diameter (R_{ct}) [25]. By using these plots, R_{ct} data were calculated as 178 Ω for GCE, 25.5 kΩ for Nf/SiO_2-APTES/GCE and 7.55 kΩ for Nf/Ni@SiO_2-APTES/GCE. According to these values, it is seen that the charge transfer resistance of the Nf/Ni@SiO_2-APTES/GCE decreased after Ni nanoparticles was deposited on the support.

Ni@SiO_2-APTES nanocomposites containing different amounts of Ni (2, 5, 10, 15, and 20%) were synthesized to determine the Ni percentage which exhibits the best electrochemical response to DA. Nf/Ni@SiO_2-APTES/GCE were prepared using these nanocomposites. Then, CVs were obtained in BR (pH 3.5) buffer containing 0.2 mM DA (Figure 5D). The anodic peak current values

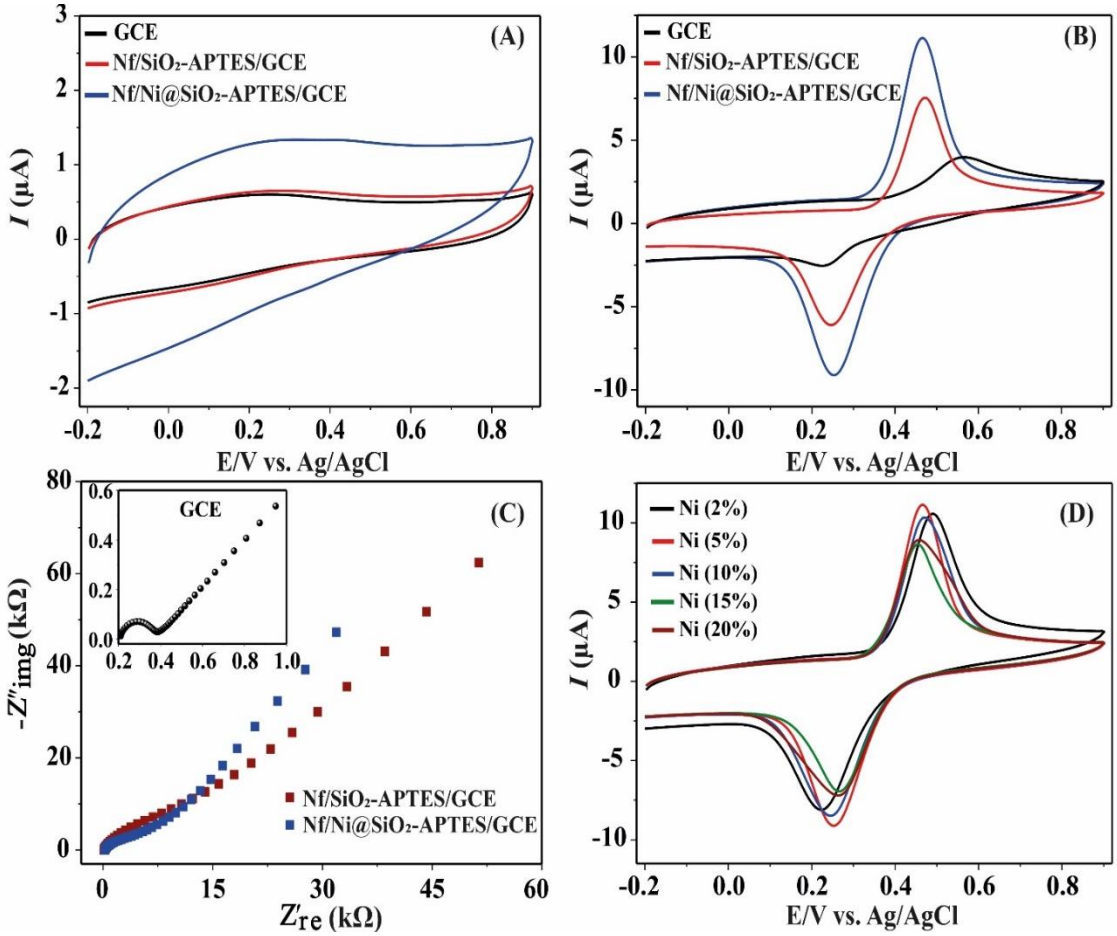


Figure 5. (A) CVs of GCE, Nf/SiO_2-APTES/GCE, and Nf/Ni@SiO_2-APTES in BR buffer (pH 3.5) in the absence of DA. (B) CVs of GCE, Nf/SiO_2-APTES/GCE, and Nf/Ni@SiO_2-APTES in BR buffer (pH 3.5) in the presence of DA at the scan rate of 50 mV/s. (C) Nyquist graph of GCE, Nf/SiO_2-APTES/GCE, and Nf/Ni@SiO_2-APTES obtained in 5 mM $[Fe(CN)_6]^{3-}$/$[Fe(CN)_6]^{4-}$ and 0.1M KCl. (D) Effect of Ni ratio on the redox reaction of DA.

of DA of these voltammograms were compared. The oxidation peak currents of the sensors were 8.89 µA for 2% Ni, 9.77 µA for 5% Ni, 8.98 µA for 10% Ni, 7.29 µA for 15% Ni and 7.56 µA for 20% Ni. According to these results, Nf/Ni@SiO_2-APTES (5% Ni) exhibited the highest oxidation peak current and electrochemical measurements were performed using this ratio.

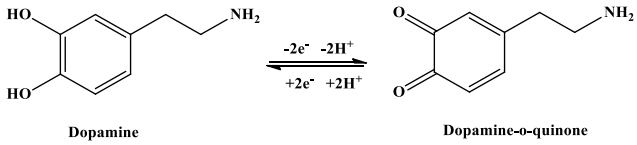
In our study, BR buffer, which has a wide buffering range (2-12), was used to investigate the influence of the pH of the electrolyte on the redox behavior of DA. Figure 6A shows CVs obtained in BR buffer with different pH values (3-8). Current-pH and potential-pH graphs were drawn by means of these voltammograms (Figure 6B). The linear equation obtained for the potential-pH graph was E_{pa} = - 0.0462pH + 0.605 (R^2 = 0.9875). The equation's slope was computed to be 46.2 mV/pH, which is comparable to the Nernts equation's slope of 59 mV/pH. In some previous studies, it is thought that two proton ($2H^+$) and two electron ($2e^-$) transfers occur in the anodic reaction of DA to form dopamine-o-quinone [26], [27], [28]. As can be seen from the graph, the anodic peak current of DA increases as the pH of the electrolyte decreases. BR buffer with pH 3.5 was used in electrochemical studies by taking into account the electrochemical performance of the working electrode. In addition, the oxidation peak potential shifted to the negative direction as the pH value incremented. This means that protons participate in the electrochemical reaction [29].

The following Laviron equations were used to determine the number of electrons getting in on the electrochemical reaction and the charge transfer coefficient (1 eq. and 2 eq.) [30].

$$E_{pc} = E^{\circ} - \frac{RT}{\alpha nF} lnv \quad (1)$$

$$E_{pa} = E^{\circ} + \frac{RT}{(1-\alpha)nF} lnv \quad (2)$$

In the above equations, *n* is the number of electrons taking part in the reaction, *α* is the charge transfer coefficient, *υ* is the scan rate, *R* is the gas constant (8.314 j/mol K), *T* is the temperature in kelvin, and F is the Faraday constant (96485 C/mol). For this purpose, CVs were achieved at various scan rates (0.02-0.4 V/s) in BR (pH 3.5) buffer containing 0.2 mM DA (Figure 6C). The relationship between the reduction and oxidation peak potentials and *lnυ* is seen in Figure 6D. The linear equations were found to be E_{pa} = 0.0298*lnυ* + 0.5335 (R^2 = 0.9978) and E_{pa} = -0.0333*lnυ* + 0.1356 (R^2 = 0.9983). The slopes of these equations are equal to the slopes in the Laviron equations (1 eq. and 2 eq.), and n and α were computed to be 1.63 ~ 2 and 0.47, respectively. These values were found to be consistent with previous studies. For example, in a study by Gong and his colleagues, an electrochemical sensor was prepared using poly-tryptophan-functionalized graphene. In the study, α and n values were calculated as 0.67 and 1.65, respectively [31]. The electrochemical oxidation and reduction reaction of DA can be written as given in Scheme 2.



Scheme 2. The electrochemical oxidation and reduction reaction of DA

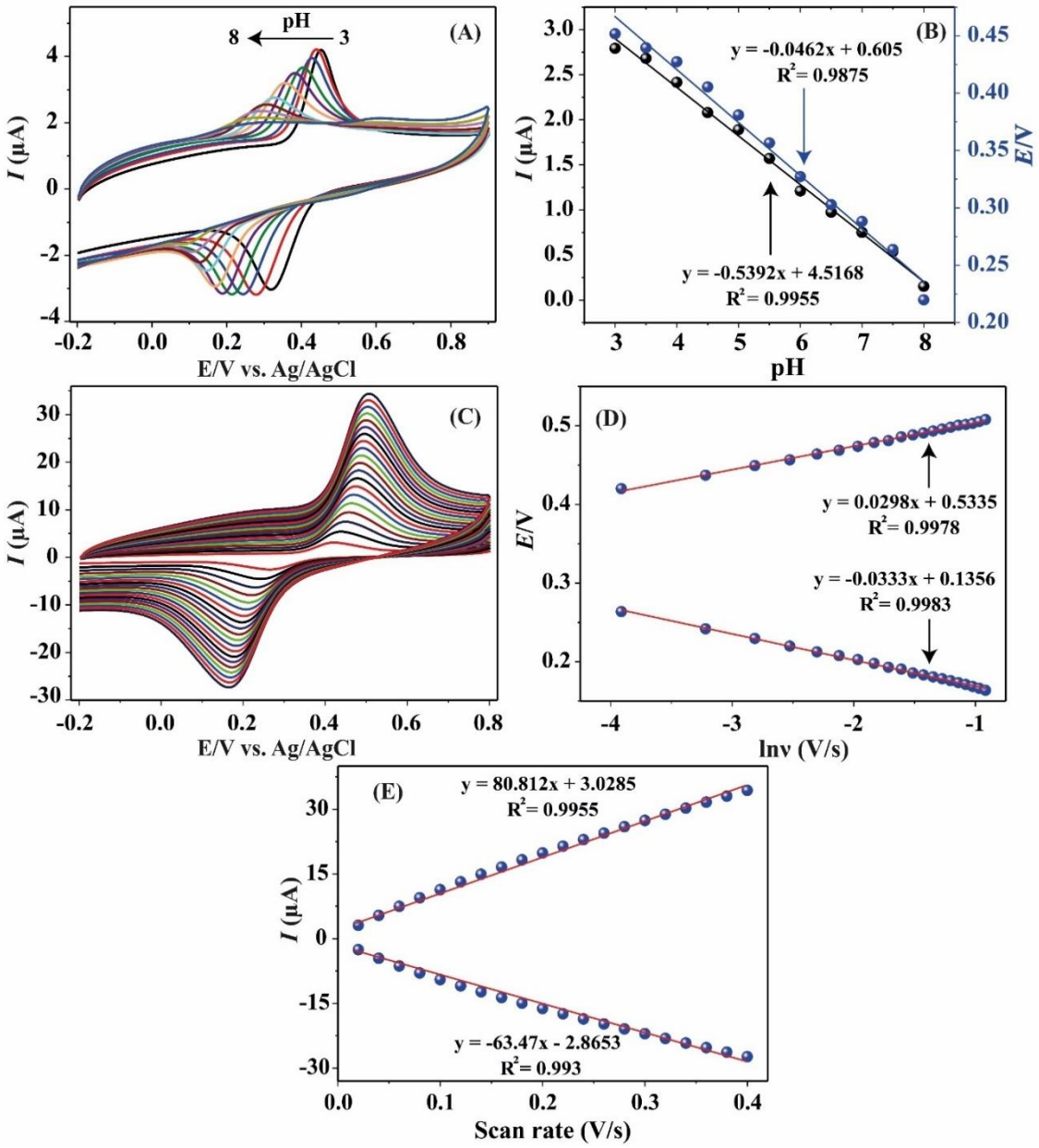


Figure 6. (A) CVs obtained by Nf/Ni@SiO_2-APTES in BR buffer with various pH (3, 4, 5, 6, 7, and 8) PBS solution. (B) *I* (μA) *vs.* pH graph. (C) CVs obtained at various scan rates (from 0.02 to 0.4 V/s). (D) *E* (V) *vs.* lnν graph. (E) *I* (μA) vs. scan rate (V/s) graph.

In order to detect the type of the reaction process of DA on the working electrode, the oxidation and reduction peak currents versus scan rate graph was obtained using CVs obtained at different scan rates (Figure 6E). According to the graph, it was concluded that both oxidation and reduction peak currents were linear against the scan rate. These results show that the electrode reaction of DA is an adsorption-controlled process [32].

In our study, DA was determined using both cyclic voltammetry and amperometric methods. For this purpose, the voltammograms were obtained by adding different concentrations of DA to BR (pH 3.5) buffer solution (Figure 7A). Using the oxidation peak current values of these voltammograms, the graph of concentration of DA versus peak current value was obtained. As seen in Figure 7B, two linear determination ranges were obtained. The first linear equation is I (μA) = 48.991[DA] (mM) + 2.2531 (R^2 = 0.9916) and the second linear equation is I (μA) = 27.681[DA] (mM) + 10.684 (R^2 = 0.9952). The sensitivities were computed to be 2130.04 μA mM^{-1} cm^{-2} and 1203.52 μA mM^{-1} cm^{-2}, respectively. The LOD was detected to be 0.36 μM depending on 3xSD/Slope. The linear detection range, sensitivity and limit of detection (LOD) values of the sensor for DA were determined using amperometry. Figure 7C shows the current-time graph obtained at 0.44 V potential using the amperometric method. Figure 7D exhibits the current-DA concentration graph. The linear equation of the graph was determined as I (μA) = 0.0133[DA] (μM) + 0.0596 (R^2 = 0.9974). The linear detection range for DA was $2x10^{-7}$ – $2.52x10^{-4}$ M. The sensitivity of the sensor was calculated to be 578.26 μA mM^{-1} cm^{-2} depending on the active surface area of the working electrode. The active surface area of each working electrode was achieved using The Randles-Sevcik equation [33]. The active surface area was detected to be 0.092 cm^2 for GCE, 0.004 cm^2 for Nf/SiO_2-APTES/GCE, and 0.023 cm^2 for Nf/Ni@SiO_2-APTES. Due to the low conductivity of the SiO_2-APTES, it showed a lower active surface area than the bare electrode. However, after Ni was deposited on the surface of the support, the active surface area of the working electrode increased by 5.75 times. The LOD value of the sensor was 0.07 μM depending on the signal-to-noise ratio of 3. Due to both the working environment and the high sensitivity of the device used, high noise was obtained at the high analyte concentrations of DA when using amperometric measurement. For this purpose, one linear determination range was determined by the amperometric method when compared with the CV method. In addition, CV is mostly used to evaluate the reduction/oxidation process of the target analyte.

To date many studies have been carried out to determine electrochemically DA. For example, in one study performed by Gong et al., the glassy carbon electrode was modified using polytryptophan-functionalized graphene. DA was determined in the range of 0.2-100 μM using the differential pulse voltammetric method and the LOD value of the sensor was found to be 0.06 μM. DA was successfully determined in dopamine injections using the sensor [34]. In another study, flower-like gold nanoparticles were deposited on the gold electrode using the potentiostatic electrochemical deposition method. The resulting nanostructures had large surface areas, and ascorbic acid (AA) and DA were determined selectively and sensitively. DA was determined in the linear range of 1-150 μM and the LOD value of the sensor was found to be 0.2 μM [35]. In one study, the working electrode was modified using platinum–silver graphene (Pt–Ag/Gr) nanocomposite for the determination of DA. CV and differential pulse voltammetry (DPV) were used in electrochemical studies for DA determination. CV results indicated that Pt–Ag/Gr/GCE increased the electrocatalytic response to DA electro-oxidation because of the synergistic influence between platinum-silver nanoparticles and graphene. The sensor displayed a linear detection range of 0.1-60 μM with a detection limit of 0.012 μM for DA using DPV method [36]. In a study reported by Li et al., Ag nanoparticles deposited on CuO porous nanobelts (Ag/CuO PNPs) were prepared for DA detection. DA electrochemical sensor was developed using Ag/CuO. The linear detection range was 0.04-10 μM with LOD of 7 nM [37]. In a study by Xue and colleagues, DA was determined amperometrically using gold nanoparticles (AuNPs) doped molecularly imprinted polymers (MIPs). The electrode exhibited a linear range from 0.02 to 0.54 μM and a LOD of 7.8 nM [38]. In our study the constructed Nf/Ni@SiO_2-APTES/GCE electrochemical sensor displayed a linear range between 0.2 and 252 μM with LOD of 0.07 μM and sensitivity of 578.26 μA mM^{-1} cm^{-2}. Considering these values obtained, the sensor showed a satisfactory linear determination range and detection limit.

So as to evaluate the selectivity of the Nf/Ni@SiO_2-APTES/GCE electrode, ascorbic acid (AA), uric acid (UA), phenylalanine (Phe), glucose (Glu), fructose (Fruc), mannose (Man) and sucrose (Suc) were used. Figure 8A shows the current-time graph obtained using various concentrations of DA (0.4, 1.0, 4.0 and 10.0 μM) and 25 μM AA, UA, Phe, Glu, Fruc, Man and Suc. As seen from the figure, the sensor did not show a remarkable electrocatalytic response to these compounds. This shows that the

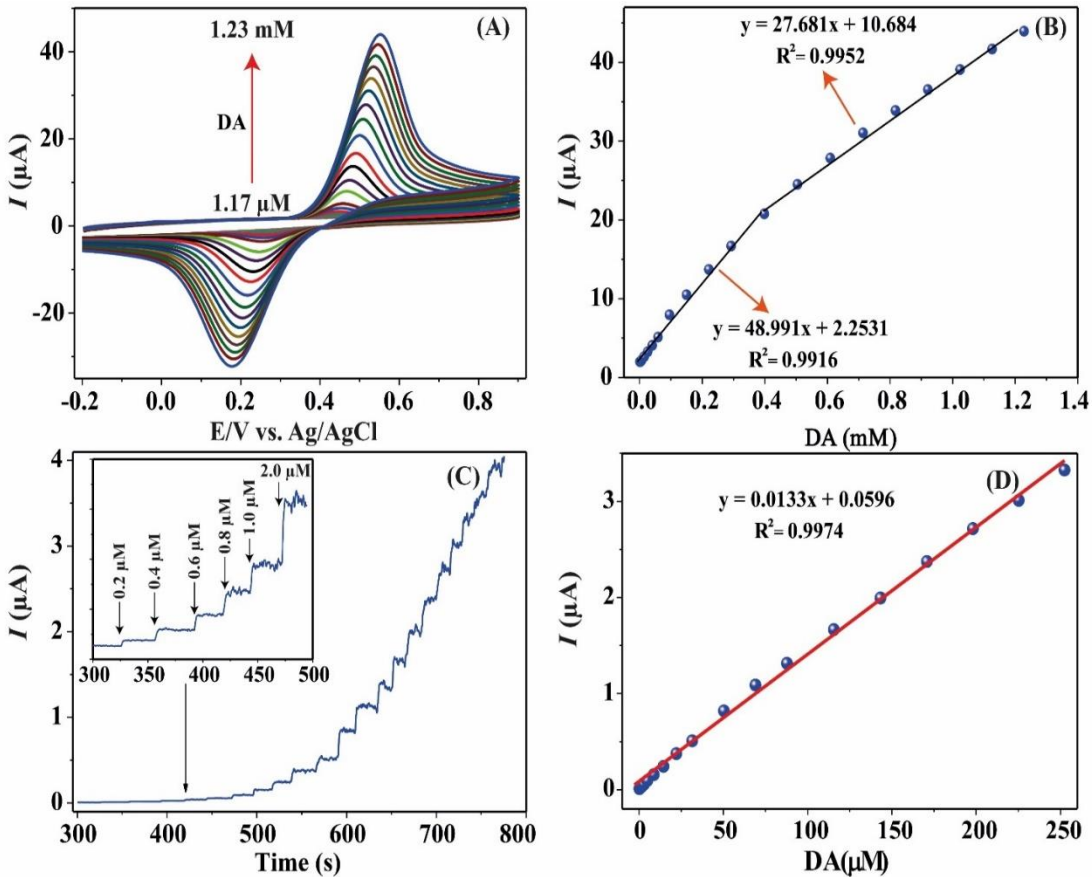


Figure 7. (A) CVs obtained by Nf/Ni@SiO_2-APTES/GCE in BR buffer (pH 3.5) containing various amount of DA at the scan speed of 50 mV/s. (B) I (µA) vs. DA concentration graph. (C) Amperometric response obtained using Nf/Ni@SiO_2-APTES/GCE on successive addition of DA to BR buffer (pH 3.5) at the performed potential of 0.44 V. (D) I (µA) vs. DA concentration graph.

sensor has excellent selectivity toward DA against these substances used.

To investigate the repeatability, 0.4, 1.0, 4.0 and 10.0 µM DA were measured using the same Nf/Ni@SiO_2-APTES/GCE working electrode. Figure 8B shows the current-DA concentration graph. The relative standard deviation (RSD) values were 5.46% for 0.4 µM DA, 3.22% for 1.0 µM DA, 2.7% for 4.0 µM DA, and for 2.6% for 10.0 µM DA. These findings indicated that the sensor has satisfactory repeatability.

To evaluate the reproducibility of the sensor, five Nf/Ni@SiO_2-APTES/GCE sensors were prepared under the same conditions. Then, electrochemical response of these electrodes to 0.4, 1.0, 4.0 and 10.0 µM DA were obtained. Figure 8C exhibits current-DA concentration graph. The RSD values were calculated as 3.9% for 0.4 µM DA, 1.87% for 1.0 µM DA, 0.6% for 4.0 µM DA, and 3.08% for 10.0 µM DA. According to these results, the sensor exhibited satisfactory reproducibility.

In our study, dopamine hydrochloride injection was used for real sample analysis. For this purpose, injection containing 200 mg/5mL dopamine hydrochloride were purchased from a local pharmacy and 0.1 and 1.0 mM DA solutions were prepared. Different concentrations of DA were put into the cell containing BR (pH 3.5). After that, the current-time graph was obtained. The mean value, standard deviation (SD) and RSD values was calculated for DA concentrations. Table 2 shows the RSD and recovery values with the DA concentrations added and found in the electrochemical cell, demonstrating that DA can be determined in dopamine injection using Nf/Ni@SiO_2-APTES/GCE sensor.

The storage stability of Nf/Ni@SiO_2-APTES/GCE was tested for 28 days. The sensor showed an electrocatalytic response of 95.33% after 28 days. It was concluded that the sensor was extremely stable within the studied time period.

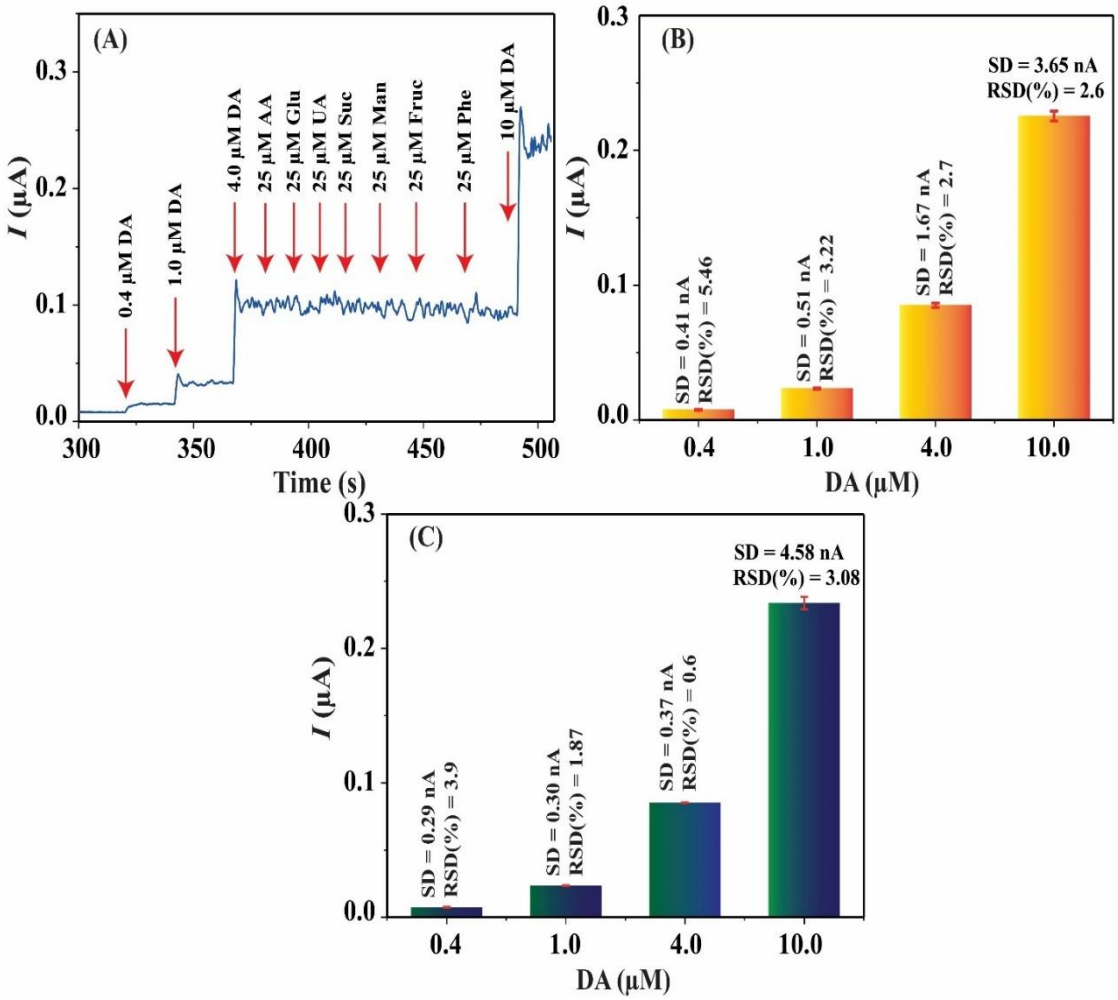


Figure 8. (A) Amperometric signals achieved at the Nf/Ni@SiO_2/GCE after adding 0.4, 1.0, 4.0 and 10.0 µM DA and 25 µM AA, UA, Phe, Glu, Fruc, Man, and Suc to BR buffer (pH 3.5). (B) The repeatability response of the sensor to DA with various amount. (C) The reproducibility of response of the sensor to DA with different concentrations.

Table 2. Electrochemical determination of dopamine in dopamine injection using Nf/Ni@SiO_2- APTES/GCE sensor

Samples	Added (µM)	Found (µM)	RSD (%)	Recovery (%)
1	0.4	0.4±0.015	3.75	100
2	1.0	1.04±0.05	4.81	104
3	4.0	4.0±0.23	5.75	100
4	10.0	10.4±0.78	7.50	104

4. Conclusion and Suggestions

In general, a novel DA electrochemical sensor was developed in this study. For this purpose, SiO_2 used as the support material was modified using APTES. Ni nanoparticles were deposited on the support. EIS, CV and amperometry were used for electrochemical studies. Ni (%5) peaks were not clearly formed in the XRD graph because Ni nanoparticles formed an amorphous structure (Ni-B) with B. The FTIR graph shows that the APTES compound was bound on the SiO_2 support. The Nf/Ni@SiO_2-APTES/GCE sensor displayed a linear detection range for DA in the range of $2x10^{-7}$ – $2.52x10^{-4}$ M. The sensitivity of the sensor was calculated to be 578.26 µA mM^{-1} cm^{-2} depending on the active surface area of the electrode using amperometric method. The LOD was computed to be 0.07 µM. The sensor exhibited extremely good

selectivity, repeatability, reproducibility and storage stability. This study showed that employing transition metal nanoparticles deposited on APTES functionalized SiO_2 is a potential method for fabricating an electrochemical sensor. Future studies may focus on new electrochemical sensors by depositing other transition metal or metal oxide nanoparticles on SiO_2-APTES.

Contributions of the authors

Yunus Emre Yıldırım: Investigation, Data curation. Muhammet Güler: Investigation, Writing, editing, Visualization, Methodology, Investigation, Data curation, Conceptualization.

Acknowledgment

This study was supported by Van Yüzüncü Yıl University, Scientific Research Projects Department (project number FYL-2021-9370).

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] S. Park, S. Park, R. A. Jeong, H. Boo, J. Park, H. C. Kim, T. D. Chung, "Nonenzymatic continuous glucose monitoring in human whole blood using electrified nanoporous Pt," *Biosens. Bioelectron.*, vol. 31, no. 1, pp. 284-291, January 2012.

[2] T. V. Maia, V. A. Conceição, "Dopaminergic disturbances in Tourette syndrome: an integrative account," *Biol. psychiatry*, vol. 84, no. 5, pp. 332-344, September 2018.

[3] O. D. Howes, R. McCutcheon, M. J. Owen, R. M. Murray, "The role of genes, stress, and dopamine in the development of schizophrenia," *Biol. psychiatry*, vol. 81, no. 1, pp. 9-20, January 2017.

[4] B. K. Madras, G. M. Miller, A. J. Fischman, "The dopamine transporter and attention-deficit/hyperactivity disorder," *Biol. psychiatry*, vol. 57, no. 11, pp. 1397-1409, October 2005.

[5] M. A. Peters, A. M. Walenkamp, I. P. Kema, C. Meijer, E. G. de Vries, S. F. Oosting, "Dopamine and serotonin regulate tumor behavior by affecting angiogenesis," *Drug Resist. Updat.*, vol. 17, no. 4-6, pp. 96-104, December 2014.

[6] J. A. Obeso, M. C. Rodríguez-Oroz, M. Rodríguez, J. Arbizu, J. M. Giménez-Amaya, "The basal ganglia and disorders of movement: pathophysiological mechanisms," *Physiol.*, vol. 17, pp. 51-55, April 2002.

[7] M. Groves, J. P. Vonsattel, P. Mazzoni, K. Marder, "Huntington's disease," *Science*, vol. 2003, pp. dn3, October 2003.

[8] M. Amiri, S. Dadfarnia, A. M. H. Shabani, S. Sadjadi, "Non-enzymatic sensing of dopamine by localized surface plasmon resonance using carbon dots-functionalized gold nanoparticles," *J. Pharm. Biomed. Anal.*, vol. 172, pp. 223-229, August 2019.

[9] X. Zhang, J. Zheng, "Hollow carbon sphere supported Ag nanoparticles for promoting electrocatalytic performance of dopamine sensing," *Sens. Actuators B: Chem.*, vol. 290, pp. 648-655, July 2019.

[10] T. W. Chen, S. Chinnapaiyan, S. M. Chen, M. A. Ali, M. S. Elshikh, A. H. Mahmoud, "A feasible sonochemical approach to synthesize $CuO@CeO_2$ nanomaterial and their enhanced non-enzymatic sensor performance towards neurotransmitter," *Ultrason. Sonochem.*, vol. 63, pp. 104903, May 2020.

[11] M. S. Wilson, "Electrochemical immunosensors for the simultaneous detection of two tumor markers," *Anal. Chem.*, vol. 77, no. 5, pp. 1496-1502, February 2005.

[12] P. D'Orazio, "Biosensors in clinical chemistry," *Clin. Chim. Acta*, Vol. 334, no. 1-2, pp. 41-69, August 2003.

[13] H. Huang, Y. Chen, Z. Chen, J. Chen, Y. Hu, J. J. Zhu, "Electrochemical sensor based on Ce-MOF/carbon nanotube composite for the simultaneous discrimination of hydroquinone and catechol," *J. Hazard. Mater.*, vol. *416*, pp. 125895, August 2021.

[14] H. V. Kiranakumar, R. Thejas, C. S. Naveen, M. I. Khan, G. D. Prasanna, S. Reddy, M. Oreijah, K. Guedri, O. T. Bafakeeh, M. Jamee, "A review on electrical and gas-sensing properties of reduced graphene oxide-metal oxide nanocomposites," *Biomass Convers. Biorefin.*, pp. 1-11, August 2022.

[15] M. Zhang, Y. Yang, W. Guo, "Electrochemical sensor for sensitive nitrite and sulfite detection in milk based on acid-treated Fe_3O_4@SiO_2 nanoparticles," *Food Chem.*, vol. 430, pp. 137004, January 2024.

[16] M. Kumar, B. K. Swamy, S. Reddy, W. Zhao, S. Chetana, V. G. Kumar, "ZnO/functionalized MWCNT and Ag/functionalized MWCNT modified carbon paste electrodes for the determination of dopamine, paracetamol and folic acid," *J. Electroanal. Chem.*, vol. 835, pp. 96-105, February 2019.

[17] L. Bayram, M. Guler, "An ultra-sensitive non-enzymatic hydrogen peroxide sensor based on SiO_2-APTES supported Au nanoparticles modified glassy carbon electrode," *Prog. Nat. Sci.: Mater. Int.*, vol. 29, no. 4, pp. 390-396, August 2019.

[18] J. Xu, J. Zhang, H. Peng, X. Xu, W. Liu, Z. Wang, N. Zhang, X. Wang, "Ag supported on meso-structured SiO_2 with different morphologies for CO oxidation: On the inherent factors influencing the activity of Ag catalysts," *Microporous Mesoporous Mater.*, vol. 242, pp. 90-98, April 2017.

[19] D. Gao, Z. Zhang, M. Wu, C. Xie, G. Guan, D. Wang, "A surface functional monomer-directing strategy for highly dense imprinting of TNT at surface of silica nanoparticles," *J. Am. Chem. Soc.*, vol. 129, no. 25, pp. 7859-7866, June 2007.

[20] M. Celebi, M. Yurderi, A. Bulut, M. Kaya, M. Zahmakiran, "Palladium nanoparticles supported on amine-functionalized SiO_2 for the catalytic hexavalent chromium reduction," *Appl. Catal. B: Environ.*, vol. 180, pp. 53-64, January 2016.

[21] I. B. Bwatanglang, S. T. Magili, I. Kaigamma, "Adsorption of phenol over bio-based silica/calcium carbonate (CS-SiO_2/$CaCO_3$) nanocomposite synthesized from waste eggshells and rice husks," *PeerJ Physical Chem.*, vol. 3, pp. e17, March 2021.

[22] C. Pereira, J. F. Silva, A. M. Pereira, J. P. Araujo, G. Blanco, J. M. Pintado, C. Freire, "[$VO(acac)_2$] hybrid catalyst: from complex immobilization onto silica nanoparticles to catalytic application in the epoxidation of geraniol," *Catal. Sci. Techno.*, vol. 1, no. 5, pp. 784-793, May 2011.

[23] Y. He, M. Qiao, H. Hu, Y. Pei, H. Li, J. Deng, K. Fan, "Preparation of amorphous Ni–B alloy: the effect of feeding order, precursor salt, pH and adding rate," *Mater. Lett.*, vol. 56, no. 6, pp. 952-957, November 2002.

[24] K. Chou, S. Chang, K. Huang, "Study on the characteristics of nanosized nickel particles using sodium borohydride to promote conversion," *Adv. Techno. Mater. Mater. Process. J.*, vol. 8, no. 2, pp. 172, January 2007.

[25] Y. Li, Y. Gu, B. Zheng, L. Luo, C. Li, X. Yan, Z. Tingting, L. Nannan, Z. Zhang, "A novel electrochemical biomimetic sensor based on poly (Cu-AMT) with reduced graphene oxide for ultrasensitive detection of dopamine," *Talanta*, vol. pp. 162, 80-89, January 2017.

[26] H. S. Jang, D. Kim, C. Lee, B. Yan, X. Qin, Y. Piao, "Nafion coated Au nanoparticle-graphene quantum dot nanocomposite modified working electrode for voltammetric determination of dopamine," *Inorg. Chem. Commun.*, vol. 105, pp. 174-181, July 2019.

[27] C. Rajkumar, B. Thirumalraj, S. M. Chen, H. A. Chen, "A simple preparation of graphite/gelatin composite for electrochemical detection of dopamine," *J. Colloid. Interface Sci.*, vol. 487, pp. 149-155, February 2017.

[28] H. Yang, D. Liu, X. Zhao, J. H. Yang, H. Chang, R. Xing, S. Liu, "AuPd bimetallic nanoparticle-supported carbon nanotubes for selective detection of dopamine in the presence of ascorbic acid," *Anal. methods*, vol. 9, no. 21, 3191-3199, May 2017.

[29] S. Hu, Q. Huang, Y. Lin, C. Wei, H. Zhang, W. Zhang, Z. Guo, X. Bao, J. Shi, A. Hao, "Reduced graphene oxide-carbon dots composite as an enhanced material for electrochemical determination of dopamine," *Electrochim. Acta*, vol. 130, pp. 805-809, June 2014.

[30] E. J. J. Laviron, "General expression of the linear potential sweep voltammogram in the case of diffusionless electrochemical systems," *J. Electroanal. Chem. Interfacial Electrochem.*, vol. 101, no. 1, pp. 19-28, July 1979.

[31] Q. J. Gong, H. X. Han, Y. D. Wang, C. Z. Yao, H. Y. Yang, J. L. Qiao, "An electrochemical sensor for dopamine detection based on the electrode of a poly-tryptophan-functionalized graphene composite," *New Carbon Mater.*, vol. *35, no.* 1, pp. 34-41, February 2020.

[32] R. Nehru, S. M. Chen, "Carbon supported olivine type phosphate framework: a promising electrocatalyst for sensitive detection of dopamine," *RSC adv.*, vol. 8, pp. 27775-27785, August 2018.

[33] A. García-Miranda Ferrari, C. W. Foster, P. J. Kelly, D. A. Brownson, C. E. Banks, "Determination of the electrochemical area of screen-printed electrochemical sensing platforms," Biosensors, vol. 8, no. 2, 53. (2018).

[34] Q. J. Gong, H. X. Han, Y. D. Wang, C. Z. Yao, H. Y. Yang, J. L. Qiao, "An electrochemical sensor for dopamine detection based on the electrode of a poly-tryptophan-functionalized graphene composite," *New Carbon Mater.*, vol. 35, pp. 34-41, February 2020.

[35] Y. Zheng, Z. Huang, C. Zhao, S. Weng, W. Zheng, X. Lin, "A gold electrode with a flower-like gold nanostructure for simultaneous determination of dopamine and ascorbic acid," *Microchim. Acta*, vol. 180, pp. 537-544, February 2013.

[36] N. S. Anuar, W. J. Basirun, M. Shalauddin, S. Akhter, "A dopamine electrochemical sensor based on a platinum–silver graphene nanocomposite modified electrode," *RSC adv.*, vol. 10, pp. 17336-17344, May 2020.

[37] Y. Y. Li, P. Kang, S. Q. Wang, Z. G. Liu, Y. X. Li, Z. Guo, "Ag nanoparticles anchored onto porous CuO nanobelts for the ultrasensitive electrochemical detection of dopamine in human serum," *Sens. Actuators B: Chem.*, vol. 327, pp. 128878, January 2021.

[38] C. Xue, Q. Han, Y. Wang, J. Wu, T. Wen, R. Wang, J. Hong, X. Zhou, H. Jiang, "Amperometric detection of dopamine in human serum by electrochemical sensor based on gold nanoparticles doped molecularly imprinted polymers," *Biosens. Bioelectron.*, vol. 49, pp. 199-203, November 2013.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1013-1022 YEAR: 2024
DOI: 10.17798/bitlisfen.1494090

Enhancing Currency, Commodity and Energy Price Forecasting Using the LSTM Model: A Case Study of EUR/NZD, GAS and SUGAR Prices

Bashir ALWESH[1], Fuat TÜRK[2], Mahmut KILIÇASLAN[3*]

[1] *Çankırı Karatekin University, Department of Computer Engineering, Çankırı, Türkiye*
[2] *Kırıkkale University, Kırıkkale, Department of Computer Engineering, Kırıkkale, Türkiye*
[3] *Ankara University, Department of Computer Technologies, Nallıhan Vocational School, Ankara, Türkiye*
(ORCID: 0009-0001-9517-4024) (ORCID: 0000-0001-8159-360X) (ORCID: 0000-0003-1117-7736)

Keywords: Time series analysis, EUR/NZDUSD exchange rate, deep learning.

Abstract

Forecasts from machine and deep learning models are vital for traders and investors in the global financial markets. Many different forecasting methods rely on technical patterns. In this study, the LSTM model based on candlesticks and financial variables was used to improve trading forecasts of different types. Japanese candlesticks are among the most widely used tools for evaluating financial markets. Therefore, these candlesticks, which show price patterns and differences between buying and selling, provide important data for predicting future price fluctuations. A 15-minute candlestick or 15-minute frame is used. The model showed excellent performance in predicting currency rates (EUR/NZDUSD), with an accuracy based on mean square error (MSE = 1.377e-07). The model also showed better accuracy in predicting sugar prices compared to other models, reaching (MSE = 1.419836). The same results were obtained with the GAS model, where the value was (MSE = 0.000173). This superior performance of the model indicates its ability to generate historical patterns and use them effectively in forecasting financial markets. These results provide promising opportunities for traders and investors to make more guided and intelligent investment decisions based on future trends based on these patterns. By using historical patterns and financial data, LSTM's deep learning model shows exceptional predictive performance. It outperforms traditional machine learning methods such as XGBoost. XGBoost achieved a score on the EUR/NZDUSD exchange rate (MSE = 9.537e-07). The error rate for the presented model is considered to be high. This confirms the success of the represented approach and its ability to enable traders and investors to make more informed and strategic decisions. This ultimately contributes to improving trading conditions and investment outcomes in global financial markets.

1. Introduction

Forecasting asset prices in financial markets is essential for making informed investment decisions, managing risks, and formulating policies. With the increasing availability of financial data and advancements in computational techniques, researchers and practitioners in economics and finance have turned to sophisticated machine learning models to enhance the accuracy of price forecasts. In recent years, artificial intelligence models such as the Long Short-Term Memory (LSTM) model have garnered significant attention due to their ability to capture complex temporal dependencies in sequential data. Originally developed for natural language processing tasks, LSTM has been successfully applied to various time series prediction problems, including financial markets forecasting.

This study aims to explore the effectiveness of the LSTM model in improving the forecasting accuracy of currency, commodity, and energy prices. The study focuses on three distinct markets: the EUR/NZDUSD currency pair, natural gas (GAS) futures, and sugar (SUGAR) futures. These markets represent different asset classes with unique

*Corresponding author: m.kilicaslan@ankara.edu.tr
Received: 01.06.2024, Accepted: 26.12.2024

characteristics, providing a diverse set of challenges and opportunities for forecasting.

2. Related Works

There are many studies in the literature stating that deep learning is widely used in financial market forecasting such as volatility forecasting, price forecasting and trend forecasting for various financial assets such as stock transactions, futures, and interests. In 2019, Yan Hu and his colleagues conducted a study entitled A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction about copper prices. The classical GARCH model is combined with a deep neural network based on LSTM and ANN. In turn, the researchers point out that the challenges and limitations they faced in the research are related to the non-linear nature and change over time of the factors affecting copper prices. This is important to understand the prediction capabilities and potential limitations of the model. This study concluded that the hybrid model achieves a significant improvement in the ability to predict copper prices [1]. In 2021, both M.S. Islam, Hossainb. By conducting a study on Forex market data and massive daily trading volume of over $5.1 trillion for major currency pairs (EUR/USD, GBP/USD, USD/CAD, USD/CHF) and specific history for analysis. It was titled Foreign exchange currency rate prediction using a GRU-LSTM hybrid network, where a new model combining GRU and LSTM is presented to predict currency rates was used. With 20 cells selected for the first GRU layer and 256 cells for the second LSTM layer. The model is applied over different time periods (10 minutes and 30 minutes) using data from the different time periods to show effectiveness. However, the study indicates that the market remains difficult and unstable, as well as the specific time period for the analysis and the specific currencies on which the model was tested. However, in this study it was shown that the performed model outperforms the individual models GRU and LSTM, and also outperforms the simple moving average (SMA) model. The new model achieves relatively good performance using different metrics such as MSE, RMSE, MAE, and R^2. It is the least risky based on its R^2 score, making it a good choice for investors who want to reduce risk in the Forex market [2]. In 2021, Mohammad J. Hamayel and Amani Yousef Owda conducted a study entitled A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms on three different cryptocurrencies: Bitcoin (BTC), Litecoin (LTC), And Ethereum (ETH). Three models, RNN, GRU, and Bidirectional LSTM (bi-LSTM), were used to predict cryptocurrency prices. Emphasis is placed on forecast accuracy using the Mean Percentage Error (MAPE) criterion. However, researchers point out that current models do not take into account other factors that may affect cryptocurrency prices, such as social media and trading volume. However, the GRU model showed the advantage in predicting cryptocurrency prices, as it showed better accuracy based on MAPE. In the future, additional research is suggested to explore other factors that may influence cryptocurrency prices such as social media, user tweets, and trading volume [3]. In 2022, the Hachmi Ben Ameur Group conducted a study on daily data for commodity markets extending from January 2002 to December 2020. The Bloomberg Commodity Index and five other sub-indices are used. The use of artificial intelligence analysis and deep learning techniques has been used to predict commodity prices. The focus was on using the LSTM algorithm. The focus was solely on using the Bloomberg Commodity Index (to which the Global Aggregate Index refers) and its five sub-indices: the Agricultural Wealth Index, the Precious Metals Index, the Livestock Index, the Industrial Minerals Index, and the Energy Index. The research successfully demonstrated the effectiveness of the Long Short Memory Method (LSTM) as a tool for forecasting commodity prices. The superiority of the Bloomberg Livestock Sub-Index and the Bloomberg Industrial Metals Sub-Index is evident in evaluating other commodity indices. These results are important for investors in terms of risk management as well as for policy makers in adjusting public policy, especially during the Russian-Ukrainian war [4]. In 2023, FEI LI and colleagues conducted a study entitled A Medium to Long-Term Multi-Influencing Factor Copper Price Prediction Method Based on CNN-LSTM on copper prices with 11 factors affecting copper price fluctuations selected as explanatory variables. CNN was used to extract spatial features of the data and LSTM to extract temporal features, and these features are fed into the CNN-LSTM network to predict monthly copper prices. The explanatory variables were carefully selected based on an analysis of the characteristics of the variables and quantitative relationships with influencing factors. The researchers point out the challenges of analyzing primary data using traditional methods, which may require adjustment to the selected influencing factors and final results. However, our approach outperforms existing methods by using the effective ability to extract space features of CNNs and extract temporal features of LSTMs. In improving the predictive power of copper prices [5].

In the same year 2023, Carlo Mari and Emiliano Mari conducted a study on a variety of time series data for US energy prices, including a daily time series for electricity prices, a time series for natural gas prices, and a time series for crude oil prices. In that study, they used the method of describing the primary system through the process of back propagation, and the second system is driven by predictions of a deep neural network trained on time series of market returns. A statistical technique based on the moments simulation method was also proposed to estimate the model on market data. Although there may be challenges in applying this model to real market data because market changes are not accurately predicted. However, the researchers find that the proposed model produces results that agree well with the empirical data, as it appears to reproduce the first four central means of the empirical distributions of log returns well, and it models the observed price series well as well [6]. Another noteworthy study was presented by Fischer and Krauss in 2018. In their study, they implemented the application of LSTM networks to financial market predictions [7]. Chong, Han, and Park merged deep learning networks with three unsupervised feature extraction methods. As a result, they improved the financial forecasting performance, but it is not yet sufficient to achieve high accuracy. However, it has application advantages over traditional methods that are not naturally generated in financial market data [8]. Li and Tam developed a hybrid model in 2017. The authors combined LSTM with real-time wavelet denoising functions to predict stock indices. Here, a sliding window mechanism was adopted for wavelet denoising. Experimental results showed that their proposed model outperformed the LSTM model without wavelet denoising module [9].

Among deep learning architectures, long short-term memory (LSTM) networks, which represent a specific type of recurrent neural network (RNN), are particularly well-suited for modeling temporal patterns in various time series tasks [10-12]. In this study, LSTM model based on candlesticks and financial variables was used to improve various types of price predictions. Using this data showing price patterns and differences between buy and sell, an existing model was applied to predict future price fluctuations. The model was used to predict exchange rates, sugar prices and GAS values. The deep learning technique was compared with some traditional models. The rest of the study consists of material and methods, methodology, experimental findings and conclusions-recommendations sections.

2.1. Literature Gap

Despite advances in predictive models for analysing market data, there is a gap in understanding the limitations associated with the non-linear nature and temporal volatility of factors affecting commodity prices, particularly in highly volatile and dynamic markets such as cryptocurrencies. Existing research focuses primarily on traditional methods and does not take into account emerging factors such as social media and trading volumes. In addition, relying on deep neural networks for prediction poses challenges due to potential statistical constraints and reliance on training data that may not accurately reflect real market dynamics. This gap in the literature highlights the need for further research and development of models that address these challenges and provide more robust predictions for market analysis.

3. Material and Methods

3.1. Dataset

Time series type data were used in the financial variables category for the date and time of each trading session, in the commodities category for sugar trading prices, in the energy category for gas trading prices and in the currencies category for New Zealand dollar trading prices. They are displayed in the form of Japanese candlesticks. Each time period is defined by the opening price, the high price, the low price and the closing price. The time period is 15 minutes per frame.

3.2. Performance Evaluation Criteria

Accuracy and performance measures is crucial step for experimental results. The following performance metrics were used to evaluate the performance of the LSTM model and to measure error rates [13]. The evaluation criteria used in this study are frequently preferred and these are Mean Squared Error (MSE) [14], Mean Absolute Error (MAE) [15], Root Mean Squared Error (RMSE) [16], Mean Absolute Percentage Error (MAPE) [17], Coefficient of Determination (R-squared).

3.3 Machine Learning and Deep Learning Algorithms

In this section, the approaches compared to the presented model are explained. The relevant techniques are classified as machine learning or deep learning algorithms.

Machine Learning Algorithms encompass a diverse set of techniques and tools employed for training models to comprehend data and leverage this comprehension for predictive or classification purposes. What distinguishes machine learning is its inherent ability to learn from data, continuously enhancing model performance over time. In the context of this study, various machine learning algorithms were employed as models to predict exchange rates using the dataset under examination. These algorithms are then compared with the outcomes derived from the offered algorithm. Noteworthy among them are Random Forest (RF), Adaptive Boosting (AdaBoost) and eXtreme Gradient Boosting (XGBoost), each contributing to the comparative analysis [18]. RF is a machine learning model grounded in the concept of random trees. It is a powerful tool extensively employed for classification and forecasting purposes. This model operates by utilizing multiple independent decision trees, a characteristic that enhances accuracy and mitigates the overfitting problem, contributing to its effectiveness in diverse applications [19]. XGBoost is a robust machine learning framework designed for predictive and classification tasks. Renowned for its exceptional performance, XGBoost excels in handling large datasets. This framework operates on the principle of ensemble learning, leveraging the strengths of multiple models to construct a powerful predictive model [20]. AdaBoost is a traditional machine learning model that aims to enhance performance by creating a set of weak models. It can also be used in deep machine learning [21].

Deep Learning algorithms constitute a specialized category within the broader field of machine learning, and they belong to the family of deep neural networks. Specifically designed to tackle complex data and structures with intricate details, these algorithms excel in extracting valuable features from such data. Their remarkable capability lies in learning from diverse datasets, recognizing intricate patterns, and constructing accurate predictive models. In the context of this study, we will compare the LSTM algorithm with the following noteworthy deep learning algorithms [22]. Gated Recurrent Unit (GRU) stands as a prevalent deep neural network model tailored for processing sequential data, including text and audio. Recognized for its efficiency in handling temporal patterns, GRU finds applications in machine translation and voice recognition [23]. Gradient Boosting Decision Trees (GBDT) is considered a model for traditional machine learning, it is also used in deep machine learning. It is based on decision trees and is widely used in classification and prediction [24]. Long Short-Term Memory (LSTM) is a deep neural network that is specifically used to process data sequences such as time sequences. It has the ability to handle both long and short delays in data, making it suitable for forecasting temporal data such as currency exchange rates.

4. Methodology

The study aims to predict time series data for each trade of commodities, energy and currencies based on the time series data represented. The overall process includes several key stages, each of which contributes to the development of an accurate predictive model, as shown in Figure 1

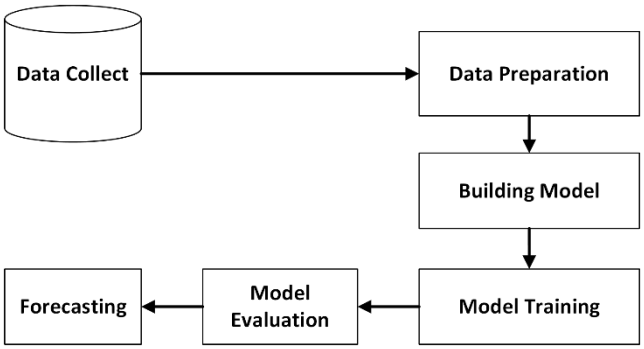


Figure 1. Flowchart of the offered system

4.1 Data collection and Data preparation

Data collection and preparation is a very important step in the study methodology. Raw data must be processed to ensure it is ready for training and evaluation. This involves the following steps:

4.1.1 Reading data from the source

In this step, raw data is obtained from the forex market for real-time and live trading. It includes a time series dataset for three categories of data: date and time of each trading session, sugar trading prices from the commodity category, gas trading prices from the energy category, and New Zealand dollar trading prices from the currency category. Each category consists of the highest price, lowest price, opening price, closing price, and trading volume. Which is retrieved from Dukascopy website, one of the most prominent forex and online trading sites. (https://www.dukascopy.com/swiss/english/marketwatch/historical/). The site provides many useful services and tools for traders, including historical price data and is a popular Swiss site.

4.1.2 Time and date formatting

Since time and date can be critical factors in temporal data analysis, the columns for time and date must be properly formatted to enable the model to accurately understand temporal patterns. Therefore, some data mining techniques such as data cleaning, data merging and data transformation were applied. The time zone facility for each session for the GMT date and time was removed, as with it, the field for each session was considered as a string. After removing it as a whole, we were able to apply operations to this field comfortably and treat it as a date and time field. Then, the direction or position of the date and time for each session was modified to match the style used.

4.1.3 Data organization

In this step, the initial financial data for each of the three types of data is prepared to be suitable for training and evaluation of the LSTM model. A package of 20 pre-readings is chosen, which consists of 20 candles. Each candle is a 15-minute price period per candle. The data is organized in a format consistent with time for the time series.

4.2 Building Model

The structure of the LSTM model is built through a deep study of the key parameters. The number of LSTM units was set to 20 in each layer, allowing the model to capture complex temporal patterns within the data. Activation functions, such as the sigmoid function, are automatically configured within each LSTM module to organize the flow of information, and by arranging the layers in a sequential manner, we created an information flow through the network.

4.3 Training Model

After establishing the structure of the LSTM model, we move to the critical and sensitive phase of training the model. In this stage, we prepare the model to make accurate predictions by exposing it to pre-processed training data. This process involves several key steps, each of which contributes to the model's ability to learn from historical patterns and make informed predictions.

Important training parameters are configured before starting the training process. The batch size, set to 120, controls the amount of training data that is processed before the model's internal weights are changed. We also specify the number of training periods, which is 20 in this case. This refers to the number of times the full training data set is applied to the model. Through this iterative approach, the model can enhance internal representations and reveal hidden patterns. Choosing a loss function is a critical step in the training process. The mean square error (MSE) loss function was used because it calculates the difference between the predicted and actual values of the target variable. The model adjusts its parameters to enhance the predicted accuracy by minimizing this loss.

4.4 Model evaluation

After the training phase, the model moves to the main evaluation phase, where its performance is carefully examined. To measure a model's ability to predict outcomes based on unseen data. A wide range of rating scales are used. These metrics include mean square error (MSE), mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R-squared). Based on the values of the resulting metrics, if the results are acceptable for those metrics, one can move to the prediction stage, and if unacceptable results are achieved, one can move to the data preparation stage again and carry out the rest of the stages until acceptable results are obtained.

4.5 Forecasting

The final step is to predict future data points using the trained LSTM model. In order to achieve this, new data must be prepared in a manner similar to the training data, predictions must be produced using the trained model, and it may be necessary to undo any changes to the data in order to obtain predictions in their original size. The accuracy of the predictions is then evaluated after further analysis using pre-defined performance criteria.

5. Experimental Results and Discussion

In this section, the results are discussed in two parts. While the first part focuses on the results of machine learning models on the three previously mentioned data categories, the second part discusses the results of deep learning models on energy, commodity and foreign exchange trading prices.

5.1. Results of machine learning models

AdaBoost, XGBoost and RF machine learning models were used with time series data on commodities, energy and currencies. In this study, the accuracy of machine learning models was tested using error and performance measurements. Table1 indicates performance and error metrics for the machine learning algorithms.

Table 1. Performance and error metrics for the AdaBoost, XGBoost, and RF

MODELS	MSE	MAE	RMSE	MAPE	R-SQUARED	Type of Data
AdaBoost	0.0034	0.05658	0.0583	0.37185	0.995697	EUR/NZDUSD
	4.8513	2.1766	2.2025	78.9842	-38.914	GAS
	23848.7	147.688	154.43	21.3163	-10.905	SUGAR
XGBoost	9.537e-07	0.00069	0.00097	0.11674	0.99536	EUR/NZDUSD
	0.000405	0.01059	0.0201	0.395931	0.99642	GAS
	48.4896	4.34413	6.9634	0.61338	0.97619	SUGAR
RF	0.00456	0.05392	0.0675	8.95	-21.2112	EUR/NZDUSD
	6.43851	1.77175	2.5374	66.12	-55.73	GAS
	25661.06	138.550	160.19	19.95	-11.59863	SUGAR

Details in Table 1, the best model is XGBoost as it has the lowest percentage error. The MSE for EUR/NZDUSD was 9.537e-07. Furthermore, for GAS, the MSE is 0.000405, which is considered to be one of the best results obtained by the XGBoost model for GAS. The similarly resluslts for SUGAR, where MSE=48.4896 was obtained. This is also true for other error measurements for the same model. Figure 2 shows the plot patterns for the actual and predicted predictive model performance of AdaBoost on SUGAR prices. Figure 3 and Figure 4 show the actual and predicted chart patterns for GAS and NZDUSD prices respectively for the same models. Based on the charts showing the chart patterns of SUGAR, GAS, and NZDUSD prices for the performance of machine learning prediction models, the best model that shows almost the same pattern of actual value is the XGBoost model across all of them. Thus, it can be said that the XGBoost machine learning model can compete with the LSTM deep learning model.

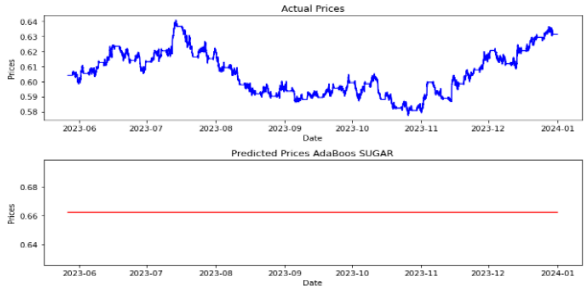


Figure 2. Chart of actual values and predicted values of SUGAR prices with AdaBoost

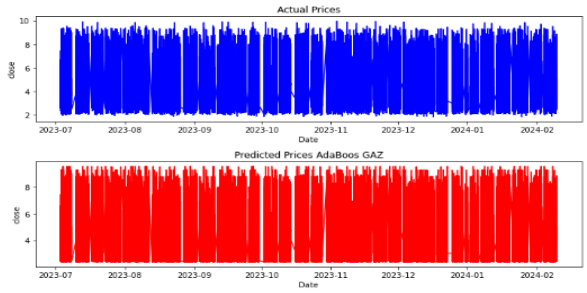


Figure 3. Chart of actual values and predicted values of GAS prices with AdaBoost

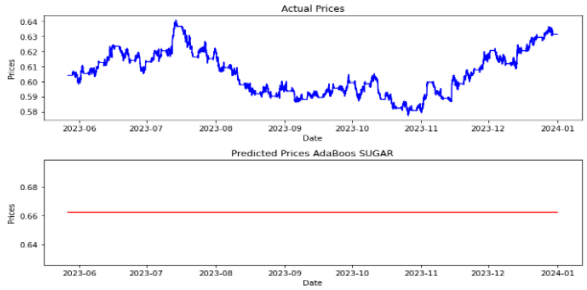


Figure 4. Chart of actual values and predicted values of NZDUSD prices with AdaBoost

5.2 Results of deep learning models

Deep learning models LSTM, GRU and GBDT were used with time series data on commodities, energy and currencies. In this study, the accuracy of deep learning models was tested using error and performance measurements. Table 2 represents the error and performance metrics of deep learning models.

Table 2. Performance and error metrics for the LSTM, GRU and GBDT

MODELS	MSE	MAE	RMSE	MAPE	R-SQUARED	TYPE OF DATA
LSTM	1.377e-07	0.00024	0.0003	0.0411	0.999330	EUR/NZDUSD
	0.000173	0.00754	0.0131	0.2706	0.998470	GAS
	1.419836	0.61444	1.1915	0.0902	0.999302	SUGAR
GRU	1.3772e-07	0.00025	0.0003	0.0414	0.999330	EUR/NZDUSD
	0.00019	0.00961	0.0138	0.3505	0.998308	GAS
	1.537671	0.75178	1.24002	0.1106	0.999245	SUGAR
GBDT	0.002162	0.04175	0.0465	6.4133	-4.093e-05	EUR/NZDUSD
	3.933846	1.66492	1.9833	43.108	-8.96e-05	GAS
	7987.260	73.2919	89.371	13.0054	-1.824e-05	SUGAR

Based on the error and performance metrics resulting from the three previous deep learning model in Table 2. We find that the best model that achieved the lowest error rate in the prediction process is the LSTM model compared to the rest of the other deep learning models. The MAE ratio was = 0.00024 for EUR/NZDUSD. Also, when using the same measure but on the GAS, we found that MAE = 0.00754, which is also considered one of the best results achieved by the LSTM model on the GAS. The same applies to SUGAR, which achieved MAE = 0.61444. This applies to the rest of the error measures for the same model. Noting that the GRU model achieved error measures that were very close to the model. Figure 5 shows the prediction behavior of LSTM, GRU and GBDT models on SUGAR prices.

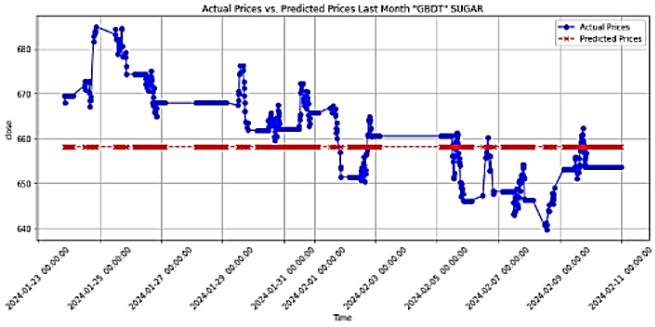


Figure 5. Chart of actual values and predicted values of SUGAR prices with LSTM

This is based on chart patterns for the performance of actual and forecast SUGAR, GAS and NZDUSD price prediction models for deep learning models. The LSTM and GRU model showed remarkable success in achieving perfect matching between actual and expected value in all relevant data categories. The GRU model showed high convergence with the model.

Model is considered to have succeeded in achieving the best results from other deep learning models. Therefore, to prove the strength and accuracy of the presented model, we can observe or compare some other results with the machine learning model that achieved the best results among other models, which is XGBoost. Therefore, error rates and performance rates, as well as charts and price distribution curves, can be compared to find out which is the best deep learning model or machine learning model with these types of data. Table 3 shows the error rates between the two models.

Table 3. Performance and error metrics between both the XGBoost and LSTM

MODELS	MSE	MAE	RMSE	MAPE	R-Squared	Type of Data
LSTM	1.377e-07	0.00024	0.0003	0.0411	0.999330	EUR/NZDUSD
	0.000173	0.00754	0.0131	0.2706	0.998470	GAS
	1.419836	0.61444	1.1915	0.0902	0.999302	SUGAR
XGBoost	9.537e-07	0.00069	0.00097	0.11674	0.99536	EUR/NZDUSD
	0.000405	0.01059	0.0201	0.395931	0.99642	GAS
	48.4896	4.34413	6.9634	0.61338	0.97619	SUGAR

The large differences in error rates can be seen in Table 3 between the LSTM deep learning model and the XGBoost machine learning model on all data types: GAS, EUR/NZDUSD, SUGAR. The LSTM deep learning model has achieved lower results in error rates than the machine learning model.

When the performance metrics are examined in general, LSTM performed better than XGBoost. Especially RMSE provides information about the fit of a regression line to the data points. LSTM created the best fitting line of the predictions and data points. In other words, it predicted better than XGBoost. MSE works according to the same principles as RMSE, the squared value of RMSE. A larger R-squared value means a better regression model since it explains most of the variance in the response variable values. For this reason, LSTM performed better than XGBoost as proven by other metrics.

When analyzing the performance of the models on different datasets, it is clear that LSTM outperforms XGBoost in most metrics, especially in the case of EUR/NZDUSD data. For example, LSTM achieves lower MSE and RMSE values, meaning that LSTM had a better fit to the regression line than XGBoost. This is particularly reflected in the higher R-squared value of the LSTM model, indicating that LSTM explains most of the variance in the dependent variable values.

For GAS and SUGAR data, LSTM remains superior in most metrics such as RMSE and R-Squared, indicating its ability to provide more accurate and reliable forecasts than XGBoost. However, it can be noted that XGBoost is not necessarily performing poorly, but it is less accurate compared to LSTM, especially in forecasts related to complex datasets such as EUR/NZDUSD.Thus, the table shows that LSTM outperforms XGBoost in providing more accurate predictions on different data types, making it the more effective model in these cases.

6. Conclusion and Suggestions

This study has achieved remarkable results in terms of trade forecasting for commodities, energy and currencies using the deep learning model. The LSTM based model showed excellent performance in predicting exchange rates (EUR/NZDUSD) and a high accuracy of around 1.377e-07 based on MSE value. The same measurement was obtained on the price of sugar, where it showed the best accuracy among other models with a value of MSE = 1.419836. This was also achieved with GAS, so the MSE value = 0.000173. In addition, a comparison was made between machine learning and deep learning models in the study. The XGBoost machine learning model produced better results than similar machine learning models in terms of having fewer errors on the three data types. It also achieves patterns of charts for actual and predicted values better than its counterparts. However, when compared with the deep learning model LSTM outperforms all metrics and models as well as the visible model and price distribution curve. This indicates the model's ability to absorb historical patterns and use them effectively to predict the future. These patterns and results provide promising opportunities for traders and investors in global trading markets to make more directed and smarter investment decisions based on the trend taken by these patterns. The success of this study reflects the ability to improve trading forecasts of various types using deep learning techniques and thus can contribute to improving the conditions of traders and investors.

Contributions of the authors

BA conceives twrites the draft of the paper, implements the algorithm, and performs analysis. FT and MK write the revision of the article, analyze the algorithm results in detail, and obtain the results for comparison. All authors read and approved the final manuscript.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] Y. Hu, J. Ni, ve L. Wen, "A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction," *Physica A: Statistical Mechanics and its Applications*, cilt. 557, s. 124907, 2020.

[2] M. S. Islam ve E. Hossain, "Foreign exchange currency rate prediction using a GRU-LSTM hybrid network," *Soft Computing Letters*, cilt. 3, s. 100009, 2021.

[3] M. J. Hamayel ve A. Y. Owda, "A novel cryptocurrency price prediction model using GRU, LSTM and bi-LSTM machine learning algorithms," *AI*, cilt. 2, sayı. 4, ss. 477-496, 2021.

[4] H. Ben Ameur ve diğerleri, "Forecasting commodity prices: empirical evidence using deep learning tools," *Annals of Operations Research*, 2023, ss. 1-19.

[5] F. Li ve diğerleri, "A medium to long-term multi-influencing factor copper price prediction method based on CNN-LSTM," *IEEE Access*, cilt. 11, ss. 69458-69473, 2023.

[6] C. Mari ve E. Mari, "Deep learning based regime-switching models of energy commodity prices," *Energy Systems*, cilt. 14, sayı. 4, ss. 913-934, 2023.

[7] T. Fischer ve C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, cilt. 270, sayı. 2, ss. 654-669, 2018.

[8] E. Chong, C. Han, ve F. C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies," *Expert Systems with Applications*, cilt. 83, ss. 187-205, 2017.

[9] Z. Li ve V. Tam, "Combining the real-time wavelet denoising and long-short-term-memory neural network for predicting stock indexes," 2017 *IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, 2017.

[10] B. Cortez ve diğerleri, "An architecture for emergency event prediction using LSTM recurrent neural networks," *Expert Systems with Applications*, cilt. 97, ss. 315-324, 2018.

[11] J. Jurgovsky ve diğerleri, "Sequence classification for credit-card fraud detection," *Expert Systems with Applications*, cilt. 100, ss. 234-245, 2018.

[12] N. C. Petersen, F. Rodrigues, ve F. C. Pereira, "Multi-output bus travel time prediction with convolutional LSTM neural network," *Expert Systems with Applications*, cilt. 120, ss. 426-435, 2019.

[13] I. K. Nti, A. F. Adekoya, ve B. A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," *Artificial Intelligence Review*, cilt. 53, sayı. 4, ss. 3007-3057, 2020.

[14] G. James ve diğerleri, *An Introduction to Statistical Learning*, cilt. 112, Springer, 2013.

[15] M. Kuhn ve K. Johnson, *Applied Predictive Modeling*, cilt. 26, Springer, 2013.

[16] W. Wang ve Y. Lu, "Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model," *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2018.

[17] R. J. Hyndman ve G. Athanasopoulos, *Forecasting: Principles and Practice*, OTexts, 2018.

[18] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly Media, Inc., 2022.

[19] J. M. Klusowski, "Sharp analysis of a simple model for random forests," arXiv preprint arXiv:1805.02587, 2018.

[20] T. Chen ve C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[21] Y. Freund, R. Schapire, ve N. Abe, "A short introduction to boosting," *Journal-Japanese Society For Artificial Intelligence*, cilt. 14, ss. 1612, 1999.

[22] J. Patterson ve A. Gibson, *Deep Learning: A Practitioner's Approach*, O'Reilly Media, Inc., 2017.

[23] K. Cho ve diğerleri, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014.

[24] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, ss. 1189-1232, 2001.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1023-1030 YEAR: 2024
DOI: 10.17798/bitlisfen.1495657

Numerical Solutions for Mixed Fractional Order Two-Dimensional Telegraph Equations

Fatih ÖZBAĞ[1*], Mahmut MODANLI[1], Sadeq Taha ABDULAZEEZ[2]

[1] *Harran University, Faculty of Arts and Sciences, Department of Mathematics, Şanlıurfa, Türkiye*
[2] *University of Duhok, College of Basic Education, Department of Mathematics, Duhok, Iraq*
(ORCID: 0000-0002-5456-4261) (ORCID: 0000-0002-7743-3512) (ORCID: 0000-0003-4515-1585)

Keywords: Two-Dimensional Telegraph Equation, Caputo and ABC Fractional Derivatives, Finite Difference Technique, Numerical Solution.

Abstract

This research presents an innovative numerical approach to solving two-dimensional telegraph equations of mixed fractional order by integrating the fractional derivatives of Caputo and Atangana-Baleanu Caputo (ABC) into a single model. Using MATLAB as its implementation, the research creates a customized first-order difference scheme and analyses stability. The ability to manage mixed fractional derivatives in 2D telegraph equations, a situation that has not been tackled in literature before, is the method's innovative aspect. This development shows that these complicated equations may be efficiently and reliably modelled, opening up new avenues for the study of complex physical phenomena. The work makes a substantial contribution to the numerical analysis of fractional differential equations with mixed derivative types and opens up possible applications in areas like wave propagation and anomalous diffusion processes.

1. Introduction

Initial value problems of fractional order can be used to model a variety of scientific disciplines, helping us better understand how to characterize natural phenomena in fields like engineering, physics, economics, biology, and seismology. According to its applications in numerous engineering and scientific domains, such as [1]-[4], fractional calculus theory has grown in popularity and importance during the last few decades. Additionally, it has demonstrated success in simulating actual issues that arise in the field of chemistry [5], physics [6], biology [7], and other fields.

Over the past few years, researchers have extensively studied different aspects of fractional models applied to the telegraph equation, for example, see [8]-[14]. The authors in [8] provided a hybrid approach built on the finite difference scheme and Sinc- Galerkin approach allows for a partial solution of the two-dimensional hyperbolic telegraph problem. The B-spline collocation method was introduced by the authors in [9] as a method for solving the 1D hyperbolic telegraph equation, additionally, they demonstrated how the suggested method converges. The authors in [10] used a Galerkin-like approach to numerically present the two-dimensional hyperbolic telegraph equations. To solve a two-dimensional hyperbolic telegraph equation with both Dirichlet and Neumann boundary conditions, the authors in [11] developed a numerical solution based on the polynomial differential quadrature method. In [12], Oruç used the Hermite wavelet-based method for obtaining 2D hyperbolic telegraph equation solutions in numerical form. The approximate numerical solution and stability estimates for the two-dimensional telegraph equation were obtained employing finite difference schemes in [13]. In [14], the authors proposed a semi-discrete method utilizing shifting shape functions of least squares to provide numerical results of complicated variable-order time fractional 2-dimensional telegraph problems. Through the use of the Legendre wavelet collocation technique, the authors in [15], a collection of time-fractional telegraph equations that take the Caputo fractional derivative. The authors of

*Corresponding author: fozbag@harran.edu.tr
Received: 04.06.2024, Accepted: 16.10.2024

[16] employed a hybrid approach that incorporated the integration of the Legendre polynomials and the block-pulse functions.

The finite difference method has been successfully used to solve a variety of issues, including fractional order telegraph equations [17], fractional chaotic systems in the sense of Caputo [18], and pseudo-hyperbolic telegraph equations based on the fractional operators [19]-[20], advection–dispersion equations [21], generalized fractional derivative terms in the fractional telegraph equation [22] and the higher order mixed fractional differential equations [23]. One dimensional linear and nonlinear hyperbolic telegraph equation is studied in [24-30]. Moreover, our work extends the application of finite difference methods to the particular case of mixed Caputo and ABC fractional derivatives in the context of 2D telegraph equations, whereas previous studies like [17]–[23] have used these methods to various fractional order equations. This development makes it possible to represent complex physical systems more accurately in fields like engineering, physics, and biology, where the interaction between different types of fractional derivatives can be important for modeling real-world phenomena. In the present study, our main goals are twofold: firstly, to introduce an innovative formulation of the fractional order two-dimensional telegraph partial differential equation, and secondly, to propose a finite difference technique for obtaining numerical approximations of these equations. Notably, since no previous research has directly addressed the fractional order two-dimensional telegraph equation using the fractional derivatives of Caputo and ABC, our approach is unique in this regard.

For this, we take into account the fractional order two-dimensional telegraph equation depending on the fractional derivatives, as follows

$$\begin{aligned} &{}_0^C D_t^\alpha v(t,x,y) + {}^{ABC}_0 D_t^\beta v(t,x,y) + v(t,x,y) = \\ &v_{xx}(t,x,y) + v_{yy}(t,x,y) + f(t,x,y) \qquad (1)\\ &0 < x,y < L,\ 0 < t, 1 < \alpha \le 2, 0 < \beta \le 1, \end{aligned}$$

with initial and boundary conditions

$$v(0,x,y) = v_t(0,x,y) = \varphi(x),\ 0 \le x,y \le L \qquad (2)$$

$$v(t,0,y) = v_t(t,L,y) = \sigma(x),\quad 0 \le t \le C, 0 \le y \le L \qquad (3)$$

$$v(t,x,0) = v_t(t,x,L) = \mu(x),\quad 0 \le t \le C, 0 \le x \le L \qquad (4)$$

where, $\varphi(x)$, $\mu(x)$, $\sigma(x)$, and $f(t,x,y)$ are known functions, but $v(t,x,y)$ is an unidentified function that should be examined. The initial and boundary conditions of the telegraph partial differential equation play a crucial role in determining the behavior of the solution, as they define how signals propagate in time and space along the transmission line. These are essential in ensuring the solution remains physically meaningful over the entire spatial domain.

${}_0^C D_t^\alpha v(t,x,y)$ is the Caputo fractional derivative that is defined as

$${}_0^C D_t^\alpha v(t,x,y) = \frac{1}{\Gamma(n-\alpha)} \int_0^t \frac{v^{(n)}(s,x,y)}{(t-s)^{\alpha-n+1}} ds,$$

$n-1 < \alpha \le n$ and $n = 1,2,\ldots \in N$. Also, ${}^{ABC}_0 D_t^\beta v(t,x,y)$ is the Atangana–Baleanu Caputo fractional derivative and defined as

$${}^{ABC}_0 D_t^\beta v(t,x,y) = \frac{B(\beta)}{1-\beta} \int_0^t v'(s,x,y) E_\beta \left[\frac{-\beta}{1-\beta}(t-s)^\beta\right] ds.$$

Here $B(\beta) = 1 - \beta + \frac{\beta}{\Gamma(\beta)}$ and $E_\beta \left[\frac{-\beta}{1-\beta}(t-s)^\beta\right] = \sum_{k=0}^{\infty} \frac{\left[\frac{-\beta}{1-\beta}(t-s)^\beta\right]^k}{\Gamma(\beta k+1)}$ is the Mittag-Leffler function.

The improved Caputo and ABC fractional derivatives facilitate the application of derivatives to non-integer orders, effectively representing memory effects in physical systems. The Caputo derivative is extensively utilized in domains such as viscoelasticity, anomalous diffusion, and control theory because it offers beginning conditions analogous to those of integer-order equations. The ABC fractional derivative improves this by incorporating a non-local and non-singular kernel, rendering it appropriate for intricate systems with long-range interactions. It is utilized in epidemiology, chaos theory, and financial modeling. By integrating these two derivatives in our examination of the two-dimensional telegraph equation, we develop a more adaptable model that encompasses many physical phenomena, tackling situations where conventional fractional derivatives are unsatisfactory.

To solve problem (1), the finite difference technique is implemented. First-order difference schemes were built for the suggested model. The explicit finite difference method is used to analyze the error estimates for the two-dimensional telegraph equation dependent on the Caputo and ABC fractional derivatives. The Von-Neumann analysis approach is then provided to generate stability estimations for the stated problem.

This methodology provides a reliable and efficient mechanism for modeling these complex equations, as highlighted in the papers focused on

stability analysis. Given that the existing literature has not thoroughly explored this specific issue, the finite difference method's capacity to accommodate fractional derivatives of Caputo and Atangana-Baleanu Caputo (ABC) makes it an appropriate selection for this research.

The following is how the current paper is structured: Part 2 presents a finite difference technique for the proposed model with the addition of a stability analysis. Part 3 discusses the numerical calculations of the introduced model. The article's conclusions are provided in Part 4.

2. Material and Method

2.1. Finite Difference Technique and Stability Analysis

We analyze the stability of the proposed model and devise a numerical method known as a finite difference scheme. The scheme is designed to solve a mixed fractional order two-dimensional telegraph partial differential equation. It incorporates the use of Caputo and ABC fractional derivative.

To compute the numeric solution for the problem (1), the initial stage involves creating a difference scheme with first-order accuracy. This is achieved by adopting a specific grid spacing as $G_{\tau,h} = [0,C]_\tau \times [0,L]_h$ with $t_k = k\tau, x_n = nh, y_m = mh$ and $k = 1,2,\dots,N, n,m = 1,2,\dots,M$. Also for x and y axes, we take $h = \frac{L}{M}$ and for t axes, $\tau = \frac{C}{N}$. The finite difference techniques for Caputo and ABC fractional derivatives are then demonstrated. The following is how the Caputo fractional derivative of order $1 < \alpha \leq 2$ is constructed using a first-order difference technique.

$$
\begin{aligned}
&{}_0^C D_t^\alpha v(t_k, x_n, y_m) \\
&= \frac{\tau^{-\alpha}}{\Gamma(3-\alpha)} \sum_{j=0}^{k-1} ((j+1)^{2-\alpha} - j^{2-\alpha}) \big(v_{n,m}^{k-j+1} \\
&\qquad - 2v_{n,m}^{k-j} + v_{n,m}^{k-j-1}\big).
\end{aligned}
$$

For simplicity, we call $b_j^\alpha = ((j+1)^{2-\alpha} - j^{2-\alpha})$.

Next, the ABC fractional derivative of order $0 < \beta \leq 1$ is thus supplied with a first-order difference technique as

$$
\begin{aligned}
&{}^{ABC}_0 D_t^\beta v(t_k, x_n, y_m) \\
&= \frac{1}{\Gamma(\beta)} \left(\sum_{j=0}^{k} \frac{v_{n,m}^{k+1} - v_{n,m}^{k}}{\tau} ((t_j - t_{k+1})^{1-\beta} \right. \\
&\qquad \left. - (t_j - t_k)^{1-\beta}) \right).
\end{aligned}
$$

For simplicity, we call

$$d_j^k = ((t_j - t_{k+1})^{1-\beta} - (t_j - t_k)^{1-\beta}).$$

Then, we need to give the following difference formula of the model (1), as

$$
\begin{cases}
\frac{\tau^{-\alpha}}{\Gamma(3-\alpha)} \sum_{j=0}^{k-1} b_j^\alpha \left(v_{n,m}^{k-j+1} - 2v_{n,m}^{k-j} + v_{n,m}^{k-j-1}\right) \\
+ \frac{1}{\Gamma(\beta)} \left(\sum_{j=0}^{k} \frac{v_{n,m}^{k+1} - v_{n,m}^{k}}{\tau} d_j^k \right) + v_{n,m}^k \\
\qquad = \frac{v_{n+1,m}^k - 2v_{n,m}^k + v_{n-1,m}^k}{h^2} + \\
\qquad \frac{v_{n,m+1}^k - 2v_{n,m}^k + v_{n,m-1}^k}{h^2} + f_n^k.
\end{cases} \tag{5}
$$

After rewriting the formula (5), we obtain

$$
\begin{cases}
\frac{\tau^{-\alpha}}{\Gamma(3-\alpha)} \sum_{j=1}^{k-1} b_j^\alpha \left(v_{n,m}^{k-j+1} - 2v_{n,m}^{k-j} + v_{n,m}^{k-j-1}\right) \\
-\frac{1}{h^2} v_{n+1,m}^k - \frac{1}{h^2} v_{n-1,m}^k + \left(\frac{\tau^{-\alpha}}{\Gamma(3-\alpha)} + \frac{d_j^k}{\tau\Gamma(\beta)} \right) v_{n,m}^{k+1} \\
\qquad + \frac{\tau^{-\alpha}}{\Gamma(3-\alpha)} v_{n,m}^{k-1} + \\
\qquad \left(\frac{-2\tau^{-\alpha}}{\Gamma(3-\alpha)} - \frac{d_j^k}{\tau\Gamma(\beta)} + 1 + \frac{4}{h^2} \right) v_{n,m}^k \\
\qquad -\frac{1}{h^2} v_{n,m+1}^k - \frac{1}{h^2} v_{n,m-1}^k = f_n^k.
\end{cases} \tag{6}
$$

For the stability of scheme (6) we give the following theorem.

Theorem 2.1. If the following condition is satisfied, the formula (6) have the stability estimates:

$$\tau^{-\alpha} < \Gamma(3-\alpha)(1 + \tfrac{8}{h^2}).$$

Proof. Using the Von-Neumann analysis method for the formula (6) and with the given condition, the proof of Theorem 2.1. can be shown easily.

2.2. Numerical Results

The present section offers the substantially numerical solutions for the mixed fractional order two-dimensional telegraph problem that depend on the Caputo and ABC fractional derivatives. In the section that follows, we solve a test problem using the

numerical method to demonstrate the usefulness of the strategy. We assess the method's performance by computing the maximum norm errors. Our focus is a computational evaluation of the mixed fractional order two-dimensional telegraph equation's initial boundary value problem.

Example 1. We examine a set of mixed fractional order partial differential equations in a two-dimensional telegraph system.

$$\begin{cases} {}_0^C D_t^{\alpha} v(t,x,y) + {}^{ABC}{}_0 D_t^{\beta} v(t,x,y) + v(t,x,y) \\ = v_{xx}(t,x,y) + v_{yy}(t,x,y) + f(t,x,y) \\ f(t,x,y) = \frac{1}{B(\beta)}\left(\frac{6(1-\beta)}{\Gamma(4-\alpha)} t^{3-\alpha} + \frac{6\beta t^{\beta+3-\alpha}}{\Gamma(\beta+4-\alpha)}\right) sin(x)\sin(y) \\ +t^3 sin(x) sin(y) \\ + \frac{3}{B(\beta)}\left((1-\beta)t^3 + 6\beta \frac{t^{\beta+3}}{\Gamma(\beta+4)}\right) sin(x)\sin(y) \\ 1 < \alpha \le 2, 0 < \beta \le 1, \\ v(0,x,y) = v_t(0,x,y) = 0, \ \ 0 \le x,y \le \pi, \\ v(t,0,y) = v_t(t,L,y) = 0, \ \ 0 \le t \le 1, 0 \le y \le \pi, \\ v(t,x,0) = v_t(t,x,L) = 0, \ \ 0 \le t \le 1, 0 \le x \le \pi. \end{cases} \quad (7)$$

One can easily check that the exact solution of (7) is

$$v(t,x,y) = \frac{1}{B(\beta)}\left((1-\beta)t^3 + 6\beta \frac{t^{\beta+3}}{\Gamma(\beta+4)}\right) sin(x) sin(y).$$

We use the formula (6), apply the similar numerical procedure in [13] and utilize a modified Gauss elimination method to solve the problem (7) for n and m. The MATLAB software is employed to acquire solutions for equation (7). We compute numerical solutions for various grid points of N and M, and the error is calculated using the subsequent formula;

$$\varepsilon = \max_{1 \le k \le N-1, 1 \le n \le M-1} |v(t_k, x_n, y_m) - v(t,x,y)|$$

where $v(t_k, x_n, y_m)$ is approximate and $v(t,x,y)$ is exact solution. In problem (7), we take $\alpha = 1.9$ and $\beta = 0.9$ then calculate the error values in Table 1.

Table 1. Error table for problem (7)

N, M	Error values
N=M=5	0.1873
N=M=10	0.0869
N=M=15	0.0507
N=M=20	0.0358
N=M=25	0.0282
N=M=30	0.0246
N=M=35	0.0228
N=M=40	0.0221
N=M=50	0.0221

The error margin can be seen in the error analysis table in the numerical calculations for the created difference scheme that existed less than one. The decreasing trend in maximum norm errors as the number of grid points increases highlights the sensitivity of the scheme. Table 1 confirms the accuracy of the generated difference scheme. Furthermore, numerical simulations are provided for N and M values to demonstrate the similarity between the exact and approximate solutions. Figure 1 and Figure 2 represent the exact solution and numerical solution of (7) respectively. From these figures, one can conclude that the solutions are almost identical.

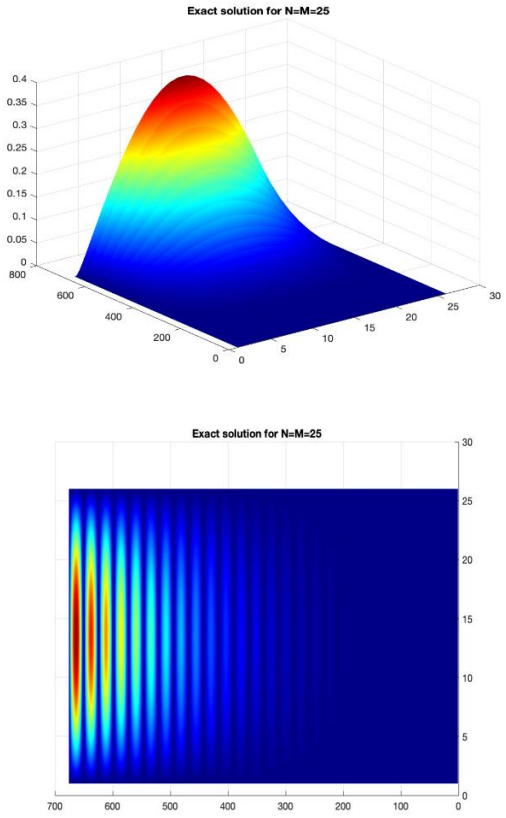


Figure 1. Gives the exact solution of (7) from different angles when $N = M = 25$, $\beta = 0.9$, and $\alpha = 1.9$.

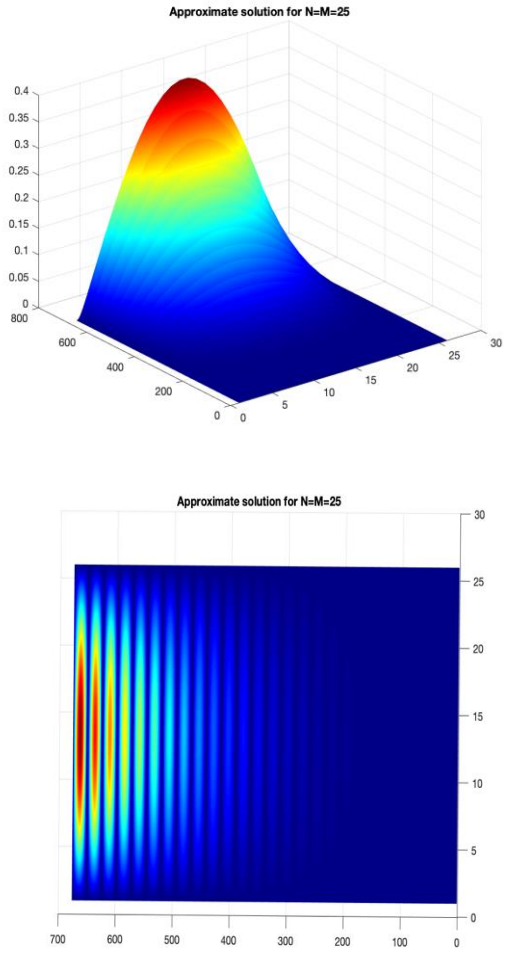


Figure 2. Gives the approximate solutions of (7) from different angles when $N = M = 25$, $\beta = 0.9$, and $\alpha = 1.9$.

Example 2. Consider the following mixed fractional order partial differential equations in a two-dimensional telegraph system.

$$\begin{cases} {}_0^C D_t^{\alpha} v(t,x,y) + {}_0^{ABC} D_t^{\beta} v(t,x,y) + v(t,x,y) \\ = v_{xx}(t,x,y) + v_{yy}(t,x,y) + f(t,x,y) \\ f(t,x,y) = \frac{1}{B(\beta)}\left(\frac{6(1-\beta)}{\Gamma(4-\alpha)} t^{3-\alpha} + \frac{6\beta t^{\beta+3-\alpha}}{\Gamma(\beta+4-\alpha)}\right)(x-x^2)(y-y^2) \\ +t^3(x-x^2)(y-y^2) + \frac{1}{B(\beta)}\left((1-\beta)t^3 + 6\beta\frac{t^{\beta+3}}{\Gamma(\beta+4)}\right) \\ \left((x-x^2)(y-y^2) + 2(x-x^2) + 2(y-y^2)\right) \qquad (8) \\ 1 < \alpha \le 2, 0 < \beta \le 1, \\ v(0,x,y) = v_t(0,x,y) = 0, \ \ 0 \le x,y \le 1, \\ v(t,0,y) = v_t(t,L,y) = 0, \ 0 \le t \le 1, 0 \le y \le 1, \\ v(t,x,0) = v_t(t,x,L) = 0, \ 0 \le t \le 1, 0 \le x \le 1. \end{cases}$$

The exact solution of (8) is

$$v(t,x,y) = \frac{1}{B(\beta)}\left((1-\beta)t^3 + 6\beta\frac{t^{\beta+3}}{\Gamma(\beta+4)}\right)(x-x^2)(y-y^2).$$

By applying the same procedure as in the first example, we compute numerical solutions and the error is calculated using the subsequent formula;

$$\varepsilon = \max_{1\le k\le N-1, 1\le n\le M-1} |v(t_k, x_n, y_m) - v(t,x,y)|$$

where $v(t_k, x_n, y_m)$ is approximate and $v(t,x,y)$ is exact solution. We take $\alpha = 1.9$ and $\beta = 0.9$ then calculate the error values in Table 2.

Table 2. Error table for problem (8)

N, M	Error values
N=10, M=5	0.0073
N=20, M=10	0.0033
N=30, M=15	0.0020
N=40, M=20	0.0015
N=50, M=25	0.0011
N=60, M=30	0.0010
N=70, M=35	0.0009

In Table 2, again the error margin can be seen that they are less than 1. Moreover, error values keep decreasing as grid values keep increasing. Table 2 also confirms the accuracy of the generated difference scheme.

We conclude that the empirical data gathered in this study provides strong confirmation for the suggested methodology based on what is evident from example 1 and 2. The accuracy and dependability of the recommended technique are highlighted by the strong agreement between the numerical results and the predicted theoretical outcomes. This innovation not only solves the issue at hand but also paves the way for prospective applications in related fields. Its accuracy and applicability represent a substantial improvement in the discipline and provide both researchers and practitioners with an invaluable tool. This novel strategy has the potential to completely alter how we tackle comparable problems in the future.

3. Conclusion and Suggestions

This research paper focuses on evaluating mixed fractional order two-dimensional telegraph equations using a combination of Caputo and ABC fractional derivatives. The authors employed a finite difference approach to analyze the problem presented in the study. Stability estimates were obtained via Von-Neumann analysis method. Error analysis data was computed by using the finite difference technique. MATLAB applications were used for calculating the error analysis results.

The results of our analysis illustrate the suggested method's exceptional accuracy and underscore its adaptability in addressing a range of intricate issues. This efficacy highlights the method's prospective uses in domains such as wave propagation, anomalous diffusion processes, and other intricate physical phenomena that incorporate mixed fractional derivatives. This strategy is especially effective in situations when conventional procedures are inadequate for addressing the complexities of fractional calculus.

This approach might be expanded to higher-dimensional systems, nonlinear situations, and various forms of fractional derivatives in future study to enhance its application. Furthermore, investigating actual applications in telecommunications, materials science, and signal processing may enhance understanding of its importance and promote additional improvements to the approach.

Contributions of the authors

F. Ozbag: software development, methodology, visualization, revision. M. Modanlı: designing the study, methodology, stability analysis. S.T. Adulazeez: literature review, writing, editing and resources.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics

References

[1] A. Atangana and J. F. Gómez-Aguilar, "Decolonisation of fractional calculus rules: Breaking commutativity and associativity to capture more natural phenomena," *Eur. Phys. J. Plus*, vol. 133, no. 4, 2018.

[2] A. Atangana, "Blind in a commutative world: simple illustrations with functions and chaotic attractors, Chaos," *Chaos, Solitons & Fractals*, vol. 114, pp. 347–363, 2018.

[3] A. Atangana, "RETRACTED ARTICLE: Derivative with two fractional orders: A new avenue of investigation toward revolution in fractional calculus," *Eur. Phys. J. Plus*, vol. 131, no. 10, 2016.

[4] M. Modanli, K. Karadag, and S. T. Abdulazeez, "Solutions of the mobile–immobile advection–dispersion model based on the fractional operators using the Crank–Nicholson difference scheme," *Chaos Solitons Fractals*, vol. 167, no. 113114, p. 113114, 2023.

[5] S. B. Yuste, L. Acedo, and K. Lindenberg, "Reaction front in an A+ B→ C reaction-subdiffusion process," *Physical Review* E, vol. 69, no. 3, 2004.

[6] A. Atangana, "Non validity of index law in fractional calculus: A fractional differential operator with Markovian and non-Markovian properties," Physica A, vol. 505, pp. 688–706, 2018.

[7] S. B. Yuste and K. Lindenberg, "Subdiffusion-limited A+A reactions," *Phys. Rev. Lett*., vol. 87, no. 11, p. 118301, 2001.

[8] E. Hesameddini and E. Asadolahifard, "A new spectral Galerkin method for solving the two dimensional hyperbolic telegraph equation," *Comput. Math. Appl*., vol. 72, no. 7, pp. 1926–1942, 2016.

[9] M. Zarebnia and R. Parvaz, "A new approach for solution of telegraph equation," *International Journal of Nonlinear Analysis and Applications*, vol. 12, no. 1, pp. 385–396, 2021.

[10] Ş. Yüzbaşı and M. Karaçayır, "A Galerkin-like scheme to solve two-dimensional telegraph equation using collocation points in initial and boundary conditions," *Comput. Math. Appl*., vol. 74, no. 12, pp. 3242–3249, 2017.

[11] R. Jiwari, S. Pandit, and R. C. Mittal, "A differential quadrature algorithm to solve the two dimensional linear hyperbolic telegraph equation with Dirichlet and Neumann boundary conditions," *Appl. Math. Comput.*, vol. 218, no. 13, pp. 7279–7294, 2012.

[12] Ö. Oruç, "A numerical procedure based on Hermite wavelets for two-dimensional hyperbolic telegraph equation," *Eng. Comput.*, vol. 34, no. 4, pp. 741–755, 2018.

[13] M. Modanli and F. Ozbag, "Stability of finite difference schemes for two-space dimensional telegraph equation," *Pramana*, vol. 96, no. 4, 2022.

[14] M. Hosseininia and M. H. Heydari, "Meshfree moving least squares method for nonlinear variable-order time fractional 2D telegraph equation involving Mittag–Leffler non-singular kernel," *Chaos Solitons Fractals*, vol. 127, pp. 389–399, 2019.

[15] X. Xu and D. Xu, "Legendre wavelets direct method for the numerical solution of time-fractional order telegraph equations," *Mediterr. J. Math.*, vol. 15, no. 1, 2018.

[16] N. Mollahasani, M. M. (mohseni) Moghadam, and K. Afrooz, "A new treatment based on hybrid functions to the solution of telegraph equations of fractional order," *Appl. Math. Model.*, vol. 40, no. 4, pp. 2804–2814, 2016.

[17] F. Ozbag and M. Modanli, "On the stability estimates and numerical solution of fractional order telegraph integro-differential equation," *Phys. Scr.*, vol. 96, no. 9, p. 094008, 2021.

[18] D. Baleanu, S. Zibaei, M. Namjoo, and A. Jajarmi, "A nonstandard finite difference scheme for the modeling and nonidentical synchronization of a novel fractional chaotic system," *Adv. Differ. Equ.*, vol. 2021, no. 1, 2021.

[19] F. Ozbag and M. Modanli, "Numerical solutions of fractional order pseudo hyperbolic differential equations by finite difference method," *Afyon Kocatepe Univ. J. Sci. Eng.*, vol. 22, no. 5, pp. 998–1004, 2022.

[20] M. Modanli, F. Ozbag, and A. Akgülma, "Finite difference method for the fractional order pseudo telegraph integro-differential equation," *J. Appl. Math. Comput. Mech.*, vol. 21, no. 1, pp. 41–54, 2022.

[21] T. Liu and M. Hou, "A fast implicit finite difference method for fractional advection-dispersion equations with fractional derivative boundary conditions," *Adv. Math. Phys.*, vol. 2017, pp. 1–8, 2017.

[22] K. Kumar, R. K. Pandey, and S. Yadav, "Finite difference scheme for a fractional telegraph equation with generalized fractional derivative terms," *Physica A*, vol. 535, no. 122271, p. 122271, 2019.

[23] S. O. Abdulla, S. T. Abdulazeez, and M. Modanli, "Comparison of third-order fractional partial differential equation based on the fractional operators using the explicit finite difference method," *Alex. Eng. J.*, vol. 70, pp. 37–44, 2023.

[24] M. Modanli and F. Ozbag, "Stability of finite difference schemes to pseudo-hyperbolic telegraph equation," *Journal of Mathematical Sciences and Modelling*, vol. 5, no. 3, pp. 92–98, 2022.

[25] W. M. Abd-Elhameed, E. H. Doha, Y. H. Youssri, and M. A. Bassuony, "New Tchebyshev-Galerkin operational matrix method for solving linear and nonlinear hyperbolic telegraph type equations," *Numer. Methods Partial Differ. Equ.*, vol. 32, no. 6, pp. 1553–1571, 2016.

[26] E. H. Doha, W. M. Abd-Elhameed, and Y. H. Youssri, "Fully Legendre spectral Galerkin algorithm for solving linear one-dimensional telegraph type equation," *Int. J. Comput. Methods*, vol. 16, no. 08, p. 1850118, 2019.

[27] Y. H. Youssri, W. M. Abd-Elhameed, and A. G. Atta, “Spectral Galerkin treatment of linear one-dimensional telegraph type problem via the generalized Lucas polynomials,” *Arab. J. Math.*, vol. 11, no. 3, pp. 601–615, 2022.

[28] A. G. Atta, W. M. Abd-Elhameed, G. M. Moatimid, and Y. H. Youssri, “Advanced shifted sixth-kind Chebyshev tau approach for solving linear one-dimensional hyperbolic telegraph type problem,” *Math. Sci.*, vol. 17, no. 4, pp. 415–429, 2023.

[29] H. T. Taghian, W. M. Abd-Elhameed, G. M. Moatimid, and Y. H. Youssri, “Shifted Gegenbauer–Galerkin algorithm for hyperbolic telegraph type equation,” *Int. J. Mod. Phys. C.*, vol. 32, no. 09, p. 2150118, 2021.

[30] R. M., Hafez, and Y. H. Youssri. "Shifted Jacobi collocation scheme for multidimensional time-fractional order telegraph equation." *Iranian Journal of Numerical Analysis and Optimization,* 10(1), 195-223, 2020.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1031-1046 YEAR: 2024
DOI: 10.17798/bitlisfen.1497810

Evaluation of Recycled Steel Fiber Effect on Concrete Performance Using Artificial Intelligence Technique

Şevin EKMEN[1*]

[1]*Harran University, Faculty of Engineering, Department of Civil Engineering, Şanlıurfa, Türkiye (ORCID:0000-0002-2577-696X)*

Keywords: Waste tire, Recycled steel fiber, Artificial neural network, Engineering properties

Abstract

Reusing waste materials is critical for sustainability and preventing adverse impacts on human life and the environment. Waste vehicle tires have become a big problem due to high consumption. It is possible to separate waste tires into different materials through technological means. Recycled steel fiber is a material obtained from these tires, and various studies have been conducted on its use in concrete. In addition to the geometric properties, such as the length and diameter, the percentage of steel fiber also affects the strength of concrete. In this study, the effect of recycled steel fiber on concrete's compressive and flexural strength values was estimated using artificial intelligence functions with high statistical significance. The relationship between the strength results and the recycled steel fiber properties was determined using literature data. The model's accuracy was demonstrated by comparing the obtained compressive and flexural strengths with the laboratory results. Thanks to the model with a high correlation coefficient created as a result of the study, the effect of recycled steel fiber on concrete performance as an alternative to laborious laboratory tests can be predicted with artificial intelligence-supported functions. With the proposed neural network method, R^2 values of 0.83 for compressive strength measurements and 0.96 for flexural strength measurements were obtained. Based on the findings, it is concluded that the recycled steel fiber-reinforced concrete parameters can be well represented by artificial neural networks, and the presented model can be used as a good alternative to laboratory studies for further research.

1. Introduction

Worldwide, there is worry over the detrimental effects of overconsumption of various materials. Thus, waste management plays a significant role in many nations' management systems. For this reason, creating various usage areas to reuse waste materials is the subject of much research. Tires of vehicles, which provide great convenience in people's daily lives, are also included in this list of waste materials as they complete their lives [1]. Approximately 1.5 thousand tires are produced yearly, and most are left as garbage after use [2]-[5]. Researchers aim to reduce resource consumption and carbon footprint by reusing waste materials. In addition, environmental pollution is also prevented by clearing various large areas of waste [6-7]. In addition, waste vehicle tires must be kept under control due to their flammability. When the chemical composition of vehicle tires is examined, it is seen that 14% by weight is natural rubber, 27% is a synthetic rubber, 28% is carbon black, 14-15% is steel, and 16-17% is filler, etc. [8-9|.

Concrete is a brittle material that can crack or fracture due to loads acting on it above its maximum bearing capacity. The disadvantages caused by this low ductility and limited tensile strength properties of concrete are tried to be reduced by using various fibers [11]-[16]. Many comprehensive laboratory studies are carried out to determine the effect of fibers with different types, proportions, and properties on the engineering properties of concrete [17]-[22]. The most common fiber types examined in the studies are propylene, steel, glass, and carbon fiber. In the studies where the use of fibers in concrete mix design was

*Corresponding author: ekmensevin@harran.edu.tr
Received: 07.06.2024, Accepted: 28.10.2024

carried out, it is seen that the use of steel fiber shows superior performance compared to other fiber types in improving some properties of concrete. Thanks to the use of steel fiber, the energy absorption capacity of concrete is increased, and the deficiencies caused by this fragile structure can be reduced [23]. In addition to the advantages provided by the use of steel fiber, it is also a fact that it creates an extra cost. Since the improvement in the performance of concrete caused by steel fiber cannot be ignored, the use of recycled steel fiber has gained importance in terms of cost reduction and sustainability [24]. Steel fibers obtained from waste vehicle tires by various methods in recycling facilities have different usage areas. Studies on the use of recycled steel fiber in concrete and the evaluation of its effect on the performance of concrete continue to be the focus of interest. Doğruyol et al. [25] examined how the use of waste steel fiber affected the behavior of concrete at high temperatures. Waste tyre steel was substituted for fine aggregate in one experimental group's concrete at a volume of 0.4%, while it was added at 0.8% in the other. Specifically, 400, 600, and 800 °C were the study's target temperatures. The overall findings indicated that fiber-added concrete significantly improved the performance of the concrete at high temperatures. Abed et al. [26] studied the effects of adding crumb rubber and recycled steel fibers to roller compacted concrete. They found that adding recycled steel fibers increased RCC toughness by 909% compared to reference specimens. The study also showed that roller compacted concrete mixtures with these additions are more durable against harmful liquids and more ductile. Ayhan et al. [27] conducted a study on the effect of waste vehicle tire steel fiber on concrete behavior. They used steel fibers at 0.4% and 0.8% by volume in place of fine aggregate in concrete. The results of the compression, flexural, and ultrasonic pulse velocity experiments were analyzed and discussed. The experimental studies revealed that the addition of 0.4% and 0.8% waste tire steel fibers decreased the compressive strength of concrete by 23% and 15%, respectively. On the other hand, the 0.4% and 0.8% waste tire steel fiber additives increased the flexural strength by 5% and 16%, respectively.

Two methods are used to extract recycled steel fiber from waste vehicle tires. These are mechanical recycling and thermo-chemical decomposition [28]. Studies have found that using recycled steel fibers has similar effects on improving the engineering properties of concrete compared to industrial steel fibers. It is believed that recycled fibers can be used as an alternative to industrial fibers in concrete production if the recycling method and source are appropriate. Liew and Akbar [29] conducted studies to examine how recycled fiber affects the engineering properties of concrete. The improving impact of recycled fiber on compressive strength, flexural toughness, shrinkage behavior, and impact resistance was revealed. Leona et al. [30] compared recycled and industrial fibers used in concrete. Using recycled fiber resulted in satisfactory shear deformation and crack behavior. Golpasand et al. [31] determined that the damage seen in concrete under repeated and biaxial loads was reduced due to the addition of recycled fiber. Frančić Smrkić et al. [32] showed that combining recycled and industrial fiber in concrete can achieve the desired mechanical properties and fatigue behavior.

Artificial Neural Networks (ANNs) have a working system that mimics the functioning of the human brain [33]. ANNs create a model based on the biological processes of the brain, such as learning, thinking, remembering, reasoning, and problem-solving [34]. In addition to solving complex relationships, ANNs can obtain accurate or near-accurate results by creating the appropriate input-output model in case of missing data. ANNs can also produce generalized responses for low-quality, noisy data, and new scenarios [35]. There have been many studies showings how effective ANN models are in problem-solving and how consistent results are obtained [36]-[42]. Awolusiki et al. [43] used an artificial neural network technique to evaluate the performance of steel and coconut fibers used in discontinuous reinforcement in concrete. It is consequently advised to use ANN for the prediction and assessment of fiber-reinforced concrete that contains steel and coconut fibers, since this strongly shows that ANN can comprehend the current link between the input variables and the output. Golafshani and Behnood [44] conducted a study using an artificial neural network assisted by multi-objective multi-verse optimizer algorithm to predict the mechanical properties of concrete containing waste foundry sand. The study achieved several optimal ANN models for compressive strength, splitting tensile strength, modulus of elasticity, and flexural strength, showing the potential for accurately estimating these properties. Jay et al. [45] investigated the suitability of response surface methodology and artificial neural network in predicting the mechanical strength of concrete with fine glass aggregate and condensed milk can fibers. Both techniques closely predicted the experimental values for compressive and splitting tensile strength. Statistical parameters indicated the effectiveness of both modeling approaches for concrete strength prediction. Almasaeid et al. [46] developed an artificial neural

network model to assess concrete strength after exposure to high temperatures without further destructive testing. They used destructive and non-destructive testing methods to investigate the effect of high temperatures (200–800 °C) on concrete compressive strength. The artificial neural network analysis showed that concrete compressive strength and the level of exposure temperature can be predicted accurately using non-destructive test results, with a coefficient of determination of 0.944.

The properties of the steel fiber used in concrete are highly influential on the change in the performance of concrete. It is crucial to accurately determine the concrete's length, diameter, length-to-diameter ratio, and fiber content to ensure compliance with desired specifications. Many laboratory studies are needed to determine the full effect of these properties. Experimental studies bring disadvantages such as time and energy loss, cost increase, and resource consumption. Considering these situations, artificial intelligence studies are of great importance. Thanks to the studies on this subject, these negative situations are minimized, and the studies to be carried out are accelerated. Recently, many artificial intelligence studies, including design criteria for various types of concrete, have been carried out [47]-[50]. As a result of the research, it was seen that there is almost no artificially assisted study that considers the use of recycled steel fiber obtained from tires in concrete. However, it is imperative to know the change in concrete performance depending on the properties of the recycled steel fiber. Thanks to the artificial intelligence model based on the function exhibiting the relationship between input and output parameters, highly correlated results for different data sets can be obtained, offering an alternative to arduous laboratory studies.

In this study, the compressive strength and flexural strength values of concrete corresponding to the length, diameter, and ratio of recycled fibers were predicted by artificial neural networks. Artificial neural networks were employed to simulate networks by utilizing experimental data from literature. Prediction functions were created to determine the correlations between the properties of recycled fiber and the mechanical properties of concrete. These correlations were then evaluated for their accuracy.

2. Material and Method

End-of-life waste vehicle tires go through sorting stages in various recycling facilities. Thanks to shredding machines, the tires, which are divided into rubber, steel fiber, and other wastes, can be used as raw materials in different fields of activity. Rubber tires are first cut into pieces of 5 cm in size for ease of transportation and storage. In the following stages, they are classified into different sizes: coarsely shredded, crumbled, granulated, and powdered [51]. After the tire pieces are brought to various sizes and passed through sieves, they are used in sections such as walking trails, playground floors, and sports fields.

In this study, the use of steel fibers obtained from waste vehicle tires by various recycling methods in concrete was examined. The geometric properties of the recycled steel fibers vary depending on the type of waste tire used and the steel fiber's separation process. Upon review of various studies, it is evident that the use of recycled steel fibers varies in percentage, length, and diameter. The studies also revealed the effect of these parameters on concrete's mechanical properties. In this study, artificial neural network functions with excellent correlation coefficients were used to predict concrete's mechanical properties corresponding to recycled steel fiber's properties. Many studies in the literature were reviewed, and data sets from various studies [52]-[62] were used to be simulated in artificial neural networks. The compressive and flexural strength values corresponding to the length, diameter, and percentage of recycled steel fiber were obtained through artificial neural networks.

In experimental studies, fibers extracted from waste vehicle tires are used in concrete mixtures to examine their impact on concrete performance through various laboratory tests. The aim of using this waste as recycled steel fiber is to determine the effect of its use on concrete performance. After processing the data from the laboratory tests into a computer, prediction values were calculated using an appropriate model created with ANN in this study. Figure 1 provides a representative diagram of the entire process. Waste steel fibers can be extracted from tires using various methods, resulting in diverse geometric properties. To evaluate their effectiveness, concrete samples are produced in a laboratory using varying proportions of waste steel fibers with different length and diameter ratios. Subsequently, the samples are subjected to relevant tests. The ANN's desired data stock is achieved by meeting compressive and bending strength values, the two most crucial concrete-use criteria. It is essential to consider how the input parameters affect the output data. The criterion of whether the laboratory results obtained are only within certain limits may be insufficient. Processing the laboratory data to create an appropriate model in artificial neural networks is necessary. This model will allow for a detailed and comprehensive statistical evaluation of strength values. After the program's essential categorization

process, these processed data become suitable for model creation. The precision of the model is crucial for obtaining results that closely align with actual values. For this reason, the correlation coefficient of the resulting model is aimed to be close to 1. Thus, depending on the model created by artificial neural networks, prediction values reflecting the performance criteria of concrete are reached.

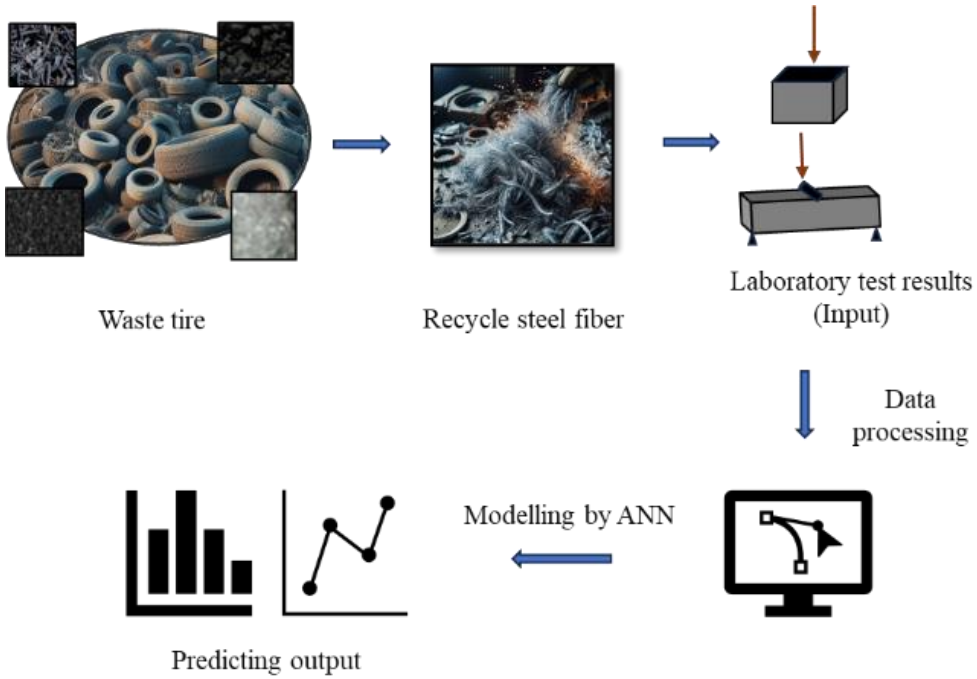


Figure 1. Flowchart for extraction of recycled steel fiber and modeling by ANN.

Some studies have been conducted on the usability of steel fibers obtained from waste tires in concrete production. Within the scope of these studies, along with the changing geometric properties of the recycled steel fiber, there has been a change in the rate of recycled steel fiber incorporation into concrete. While the amount of use was determined by mass in some studies, most studies calculated the incorporation rates by volume. The selection criterion for this study was the literature research conducted using recycled steel fiber by volume. Studies in the literature have shown that varying the length, diameter, and percentage of recycled steel fibers affects the performance of concrete. The 2D graph containing all the experimental data used in generating the function with artificial neural networks is given in Figure 2. Three different representations of the compressive and flexural strength values obtained are presented depending on each input parameter. Figure 2 (a) shows the compressive and flexural strength values corresponding to different recycled steel fiber lengths. In the literature studies evaluated within the scope of this search, the lengths of recycled steel fibers used in concrete ranged from 12 mm to 60 mm.

Figure 2 (b) displays the varying diameter ratios of the recycled steel fibers used in these studies and the compressive and flexural strength values obtained accordingly. The diameters of the steel fibers incorporated into the concrete varied between 0.22 mm and 1 mm. When the mix designs of the studies were examined, it was seen that different percentages of recycled steel fibers were used by volume. These steel fiber addition percentages and the corresponding compressive and flexural strength values are presented in Figure 2 (c). The recycled steel fiber percentages varied between 0.11% and 3.16%.

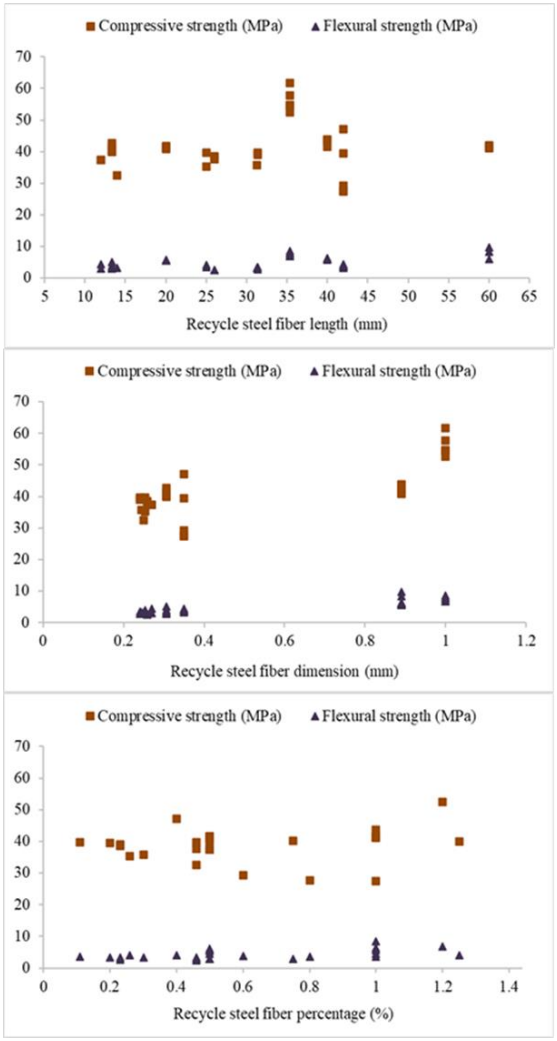


Figure 2. Compressive and flexural strength values in the literature studies depending on a) the length of recycled steel fiber (mm), b) the diameter of steel fiber (mm), and c) the percentage of steel fiber used (%).

The literature data graph to be used in the artificial neural network model is shown in Figure 3, represented in four dimensions. While recycled steel fiber length, diameter, and utilization rate constitute the input parameters, the output parameter obtained from the laboratory tests was the compressive strength. Compressive strength values vary depending on the geometric properties of the recycled steel fiber, and the amount of addition is shown in different colors. The compressive strength values increase as the color changes from blue to yellow. Thanks to this 4-dimensional graph, the effect of all the properties of the recycled steel fiber on the compressive strength values can be seen when examined together.

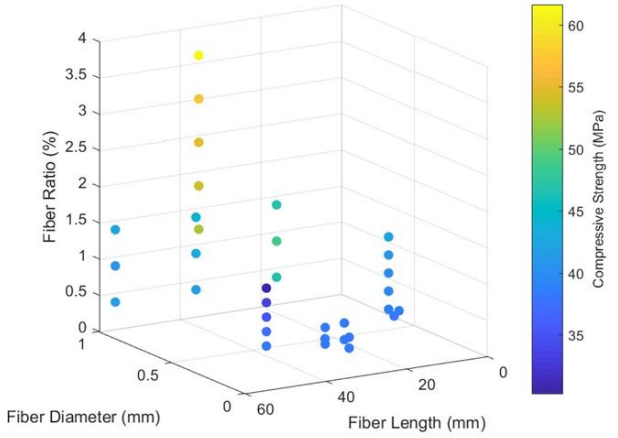


Figure 3. Compressive strength values depending on the recycled steel fiber properties in the literature studies.

The 4-dimensional graph showing the relationship between the recycled steel fiber properties and the flexural strength values of concrete is given in Figure 4. Functions with artificial neural networks were generated using the input parameters, including the properties of the recycled steel fiber, and the output parameter, including flexural strength. The flexural strength values vary depending on the recycled steel fiber's length, diameter, and usage rate, shown in different colors. The flexural strength values increase as the color changes from blue to yellow. Thanks to this 4-dimensional graph, the effect of all recycled steel fiber properties on the flexural strength values can be seen when examined together.

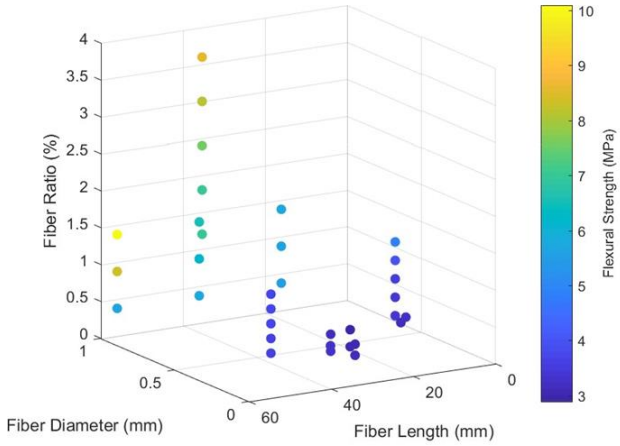


Figure 4. Flexural strength values depending on the recycled steel fiber properties in the literature studies.

3. Results and Discussion

Neural networks are a method of artificial intelligence inspired by the human brain. They can be used for many functions, such as solving complex problems, building models, grouping, and making predictions. Neural networks use a structure in which nerve cells called neurons are connected. These connections are expressed as weights and boundaries. The data is passed through the network, and each neuron produces inputs, processes, and results. Artificial neural networks progress through a learning process. Initially, weights and bounds are chosen randomly. Then, the significance increases as they are trained on the data. It is stated in various studies in the literature that working with a sufficient number of data points is one of the most critical factors for training a network [63]-[64].

In this study, prediction functions were created using artificial neural networks between recycled steel fiber properties, ratios, and compressive and flexural strength values. Network training was performed using the Matlab program [65]. Table 1 presents this study's options when using artificial neural networks.

Figure 5 shows the network diagram representing the model created with artificial neural networks. This network model consists of 3 layers: input, hidden, and output. The number of neurons in the hidden layer depends on the number of values and data used to create the model. In this study, as seen in Figure 5, this number is 10. The number of neurons in the hidden layer varies depending on the model's suitability created with artificial neural networks.

Table 1. Options used in ANN

Method	Levenberg-Marquardt
Neuron Number	10
Number of Hidden Layers	1
Data Distribution	Random
Training, Testing, and Validation Data Distribution	70%, 15%, 15%
Performance Checker Function	Mean Square Error (MSE)

In other words, it is designed to allow the creation of the output closest to the target values for the networks to be created. The input parameters that formed the input layer in the artificial neural network model were steel fiber length (mm), steel fiber diameter (mm), and steel fiber ratio (%). These input parameters are connected to the hidden layer with network connections established at different weights. After the learning and training stages in the artificial neural network model, the data of the output layer are predicted. In this study, concrete's compressive strength and flexural strength values constitute the output parameters.

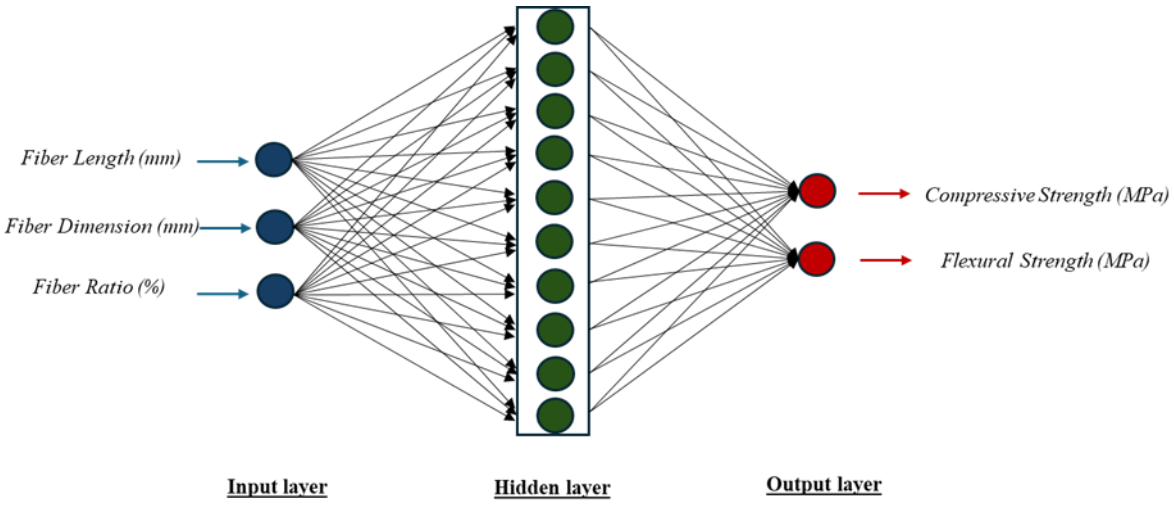


Figure 5. ANN working mechanism showing the output parameters obtained depending on the input parameters included in the study.

As shown in Table 1, the Levenberg-Marquardt method was preferred to generate the prediction function with artificial neural networks by this data set. A single hidden layer and ten neurons were used. Training, testing, and validation data distribution is random at 70%, 15%, and 15%. Mean Square Error (MSE) was considered to check and improve the performance. Fiber length, diameter, and ratio were independent variables, while compressive and flexural strength were chosen as separate

dependent variables. For this reason, two different analyses were performed, and two other functions were obtained for compressive and flexural strength. Figure 6 shows the correlation coefficient values of the function for compressive strength. The closeness of this correlation coefficient (R) to 1 gives information about the model's accuracy. Figure 1 shows how the data used in creating a function relates to prediction values for training, testing, validation, and all data. The graph also displays how closely the function aligns with the ideal curve. The obtained correlation coefficients were within the range of 0.76 to 0.95. The highest correlation coefficient was obtained in the test phase, while the lowest was obtained in the validation phase. However, the correlation coefficient obtained in validation is high enough to show that the model is usable. The high correlation coefficients indicate that the function is a good indicator of the data model and has a high prediction success.

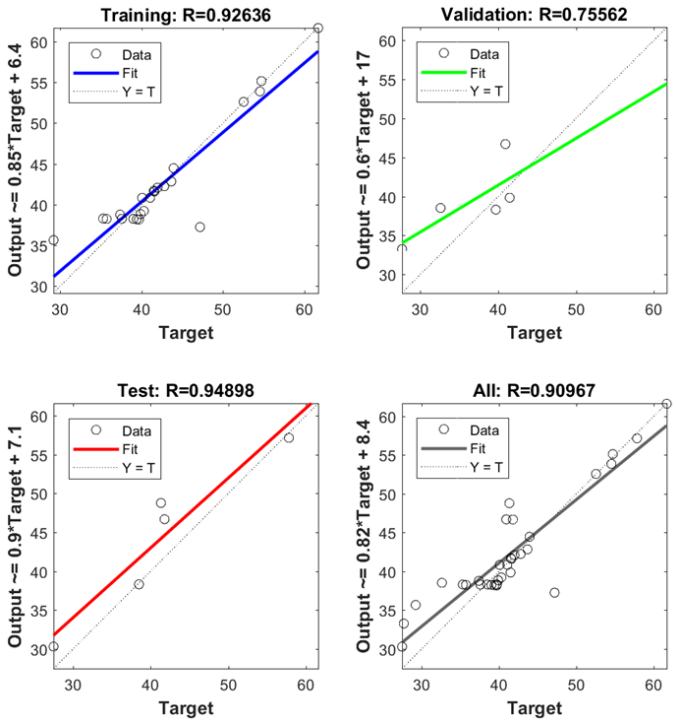


Figure 6. Correlation coefficient values of the function created for compressive strength (R).

Determining the relationship between the strength values obtained based on the model created with ANN and the actual values is essential. High determination (R^2) values representing this relationship indicate the suitability of the model function. Yıldizel et al. [24] obtained the coefficient of determination between actual and predicted values as 0.98 for compressive strength and 0.97 for flexural strength using artificial neural networks in their study on foam concrete samples containing waste car tires.

Table 2 details the relationship between the predicted compressive strength values and the target output values depending on different input parameters with artificial neural networks. Trial learning test stages were carried out to create the most appropriate model with artificial neural networks. For each of these stages, different data were used in various numbers. Finally, a connection covering all data was revealed due to the networks created and the information obtained. A relationship equation is given for each of these stages in the process of reaching the target values. These equations show the connection between actual values and predicted values. The expression y represents the compressive strength value predicted by ANN, while x represents the compressive strength values obtained from laboratory tests. Table 2 also shows the correlation and determination results, representing the closeness of the data obtained by the model to the laboratory results. The closeness of these numbers to 1 allows us to have information about the model's suitability. The coefficients of determination obtained for training, validation, and testing are 0.86, 0.57, and 0.9, respectively. The correlation coefficient of the model equation for all data was 0.91.

A model was also created to predict flexural strength values with artificial neural networks. The correlation coefficients (R) of the function reflecting this model are presented in Figure 7. Correlation

coefficients for estimating flexural strength values were determined for training, validation, testing, and all data, similar to the function created for compressive strength.

Table 2. Equations and relationship coefficients of the ANN model for output values representing compressive strength.

	Output-target correlation	R	R^2
Training	y:0.85x+6.4	0.93	0.86
Validation	y:0.6x+17	0.76	0.57
Test	y:0.9x+7.1	0.95	0.90
All	y:0.82x+8.4	0.91	0.83

The correlation coefficients of the function created for flexural strength were higher than those obtained with compressive strength. When the relationships created for all stages were examined, the highest correlation coefficient was 0.99, and the lowest value was 0.98. It was seen that it had high correlation coefficients. The relationships of the prediction results obtained for flexural strength with the target values were also obtained around the ideal curve. These high correlation coefficients showed the accuracy of the function reflecting the model. Gupta et al. [66] investigated the mechanical properties of concrete samples containing waste tyres exposed to high temperatures and obtained the correlation coefficient as 0.98 using artificial neural networks.

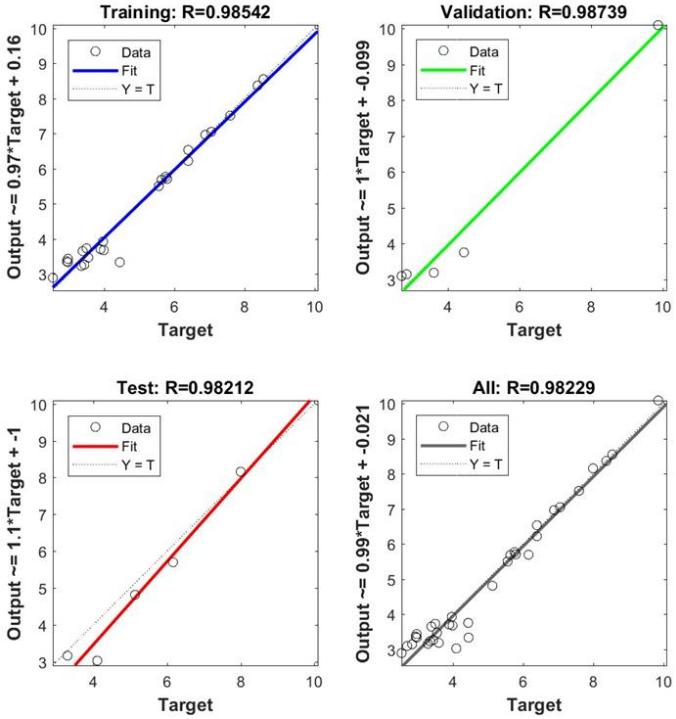


Figure 7. Correlation coefficient values of the function created for flexural strength (R).

Table 3 details the relationship between the predicted flexural strength values and the target output values depending on different input parameters with artificial neural networks. Trial learning test stages were carried out to create the most appropriate model with artificial neural networks. Various data were used in multiple numbers for each of these stages. Finally, a connection covering all data was revealed due to the networks created and the information obtained. A relationship equation is given for each of these stages in the process of reaching the target values. These equations show the connection between actual values and predicted values. The expression y represents the flexural strength value predicted by ANN, while the x represents the values obtained from laboratory tests. Table 3 also shows the correlation and determination results, representing the closeness of the data obtained by the model to the laboratory results. The closeness of these numbers to 1 allows us to have information about the model's suitability. The coefficients of determination obtained for training, validation, and testing are 0.97, 0.97, and 0.96, respectively. The correlation coefficient of the model equation for all data was 0.98.

The coefficient of determination (R^2) for compressive strength was 0.83, while for flexural strength it was 0.96. The coefficient of determination of the function for flexural strength is higher than the value obtained for compressive strength, and the coefficients of determination representing the data

relationships for both compressive and flexural strength are at acceptable levels as they are very close to 1.

Table 3. Equations and relationship coefficients of the ANN model for output values representing flexural strength.

	Output-target correlation	R	R^2
Training	y:0.97x+0.16	0.98	0.97
Validation	y:1x+0.099	0.99	0.97
Test	y:1.1x+1	0.98	0.96
All	y:0.99x+0.021	0.98	0.96

MSE explains the error relationship between the prediction values obtained with artificial neural networks and the actual values. In artificial neural networks, different numbers of iterations are performed for the model to predict the data. Thanks to the software used, it can be determined in which iteration the lowest MSE value is reached. Figure 8 shows the Mean Square Error values depending on iterations for compressive and flexural strengths. Figure 8 (a) shows that the best performance for compressive strength is obtained at the 10th iteration. It is seen that the prediction function selects the optimum point where the error is minimized among 16 iterations.

Figure 8 (b) shows that the flexural strength prediction function underwent 19 iterations, with the optimum value obtained at the 13th iteration.

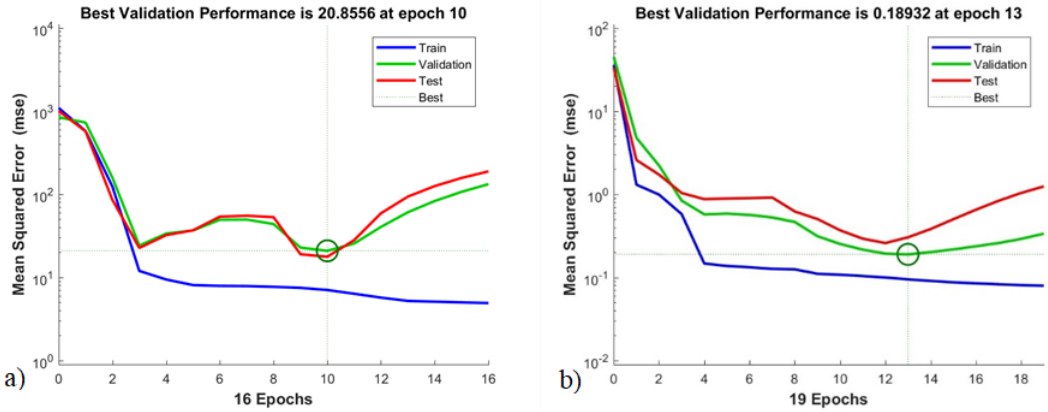


Figure 8. Mean Square Error values depending on iterations for a) compressive strength, b) flexural strength.

Figure 9 (a) illustrates the error conditions for Training, Testing, and Validation for compressive strength separately, thus presenting the error histogram for the whole dataset. Error values are obtained by numerically subtracting the output values from the target values representing the experimental results. Thanks to the program, the difference between the predicted and actual laboratory values was determined for different numbers of data. It can be seen from the figure that the error rates are low.

Figure 9 (b) shows the error histogram for flexural strength values. The error histogram includes the error rates for Training, Testing, and Validation separately. In addition, the value showing the lowest error rate is indicated as zero error. While the zero value for compressive strength values was 0.07, this result was obtained as 0.028 for flexural strength.

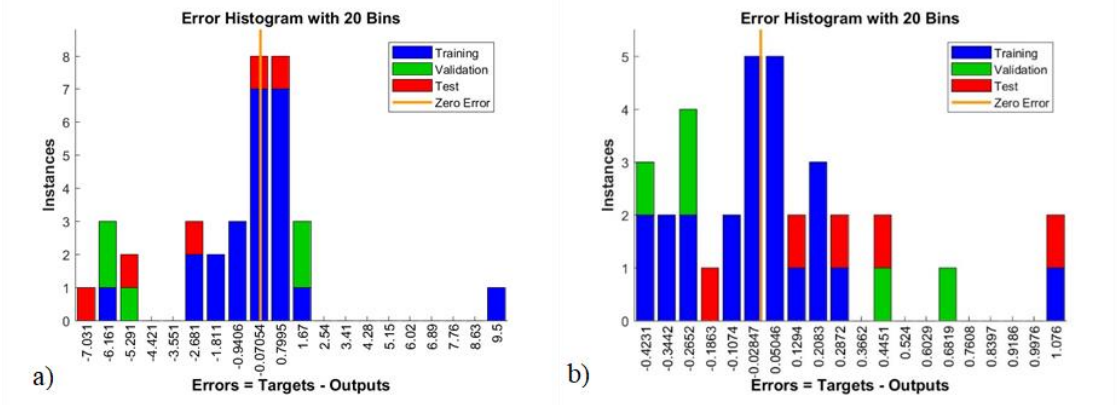


Figure 9. Error Histogram for a) compressive strength, b) flexural strength.

Figure 10 (a) presents the parameters that generate the network to predict the compressive strength results. To avoid erroneous predictions, all values governing the network's control values, convergence, and error functions are shown.

When creating a prediction model, the function that yields the closest results to the available data is generated. The parameters of the network structure used to create the function for flexural strength are presented in Figure 10(b). Since the minimum error rate is targeted during the prediction, the network's control values and all values governing the convergence and error functions are also created.

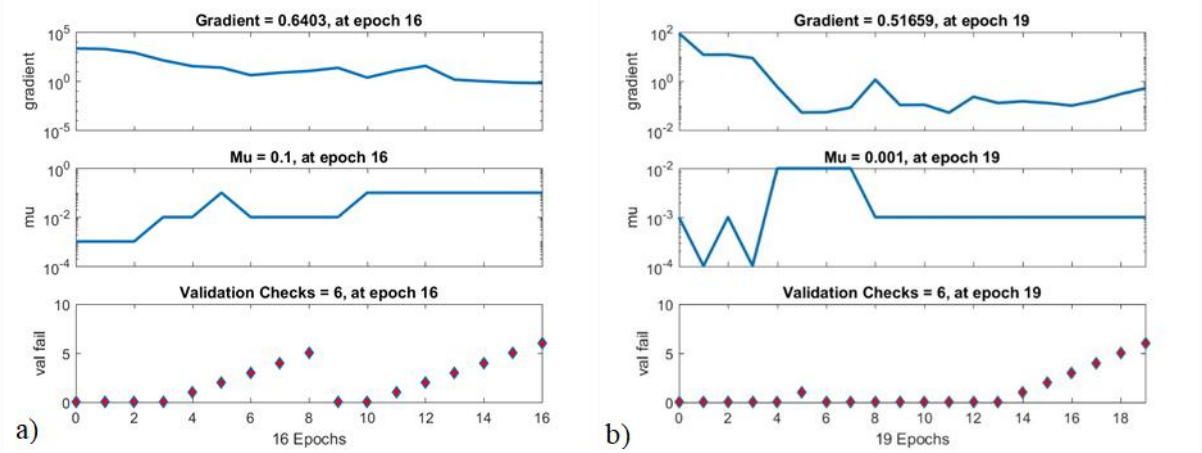


Figure 10. a) Parameters used in mesh generation for a) compressive strength, b) flexural strength.

Figure 11 shows the results obtained with ANN for compressive strength among the output parameters. Compressive strength values are given depending on the input parameters. The figure presents the compressive strength results predicted by ANN, considering the fiber properties, alongside the compressive strength values obtained from laboratory tests in the literature. Thanks to this 5-dimensional graphic design that allows all input and output parameters to be presented together, it was possible to evaluate and analyze the data. The figure shows two scales for the actual and predicted values of compressive strength results. The color change represents the ANN's values, while the round markers' size indicates the accurate laboratory data. The compressive strength values predicted by ANN varied between 30.27 MPa-61.63 MPa and increased from blue to red. The actual compressive strength values obtained from laboratory tests varied between 27.44 MPa-61.69 MPa and increased from a small to a large diameter marker.

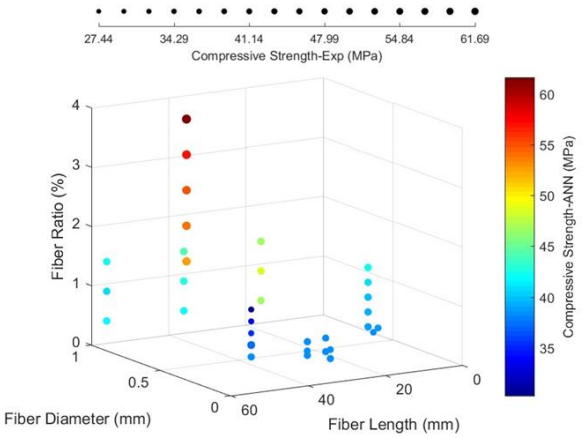


Figure 11. Values of laboratory and ANN results for compressive strength depending on recycled steel fiber properties.

Figure 12 shows Actual flexural strength values and the data obtained from the model created through ANN using inputs determined as steel fiber length, steel fiber diameter, and steel fiber usage percentage. Thanks to this 5-dimensional graphic design that allows all input and output parameters to be presented together, it was possible to evaluate and analyze the data. The color changes on two different scales represent the flexural strength values obtained by ANN. On the other hand, the actual laboratory data are represented by changes in the size of the round markers used in the representation. The flexural strength values predicted by ANN varied between 2.9

MPa-10.09 MPa and increased from blue to red. The actual flexural strength values obtained from laboratory tests varied between 2.54 MPa-9.86 MPa and increased from a small to a large diameter marker.

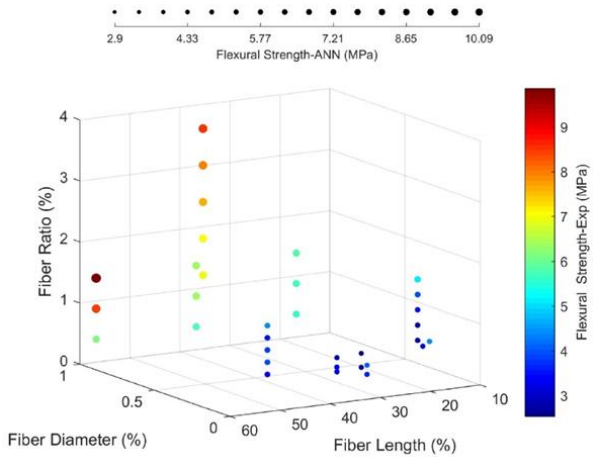


Figure 12. Values of laboratory and ANN results for flexural strength obtained depending on recycled steel fiber properties.

In Figure 13, both the actual compressive and flexural strength values obtained as a result of laboratory tests and the predicted compressive and flexural strength values obtained by the function generated by ANN are presented together. The change in strength values depending on the properties of the recycled steel fibers is shown. Compressive strength values are indicated with a + sign, while flexural strength values are marked with an asterisk. It is also seen in the graph that the strength values obtained by ANN and the laboratory results are compatible with each other.

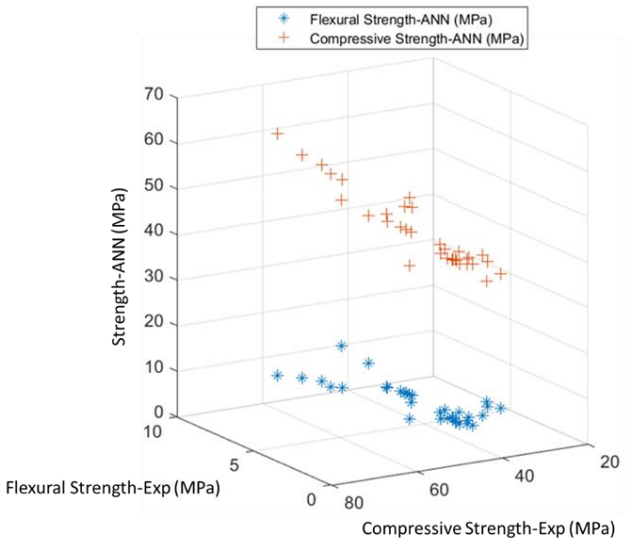


Figure 13. Laboratory and ANN strength values depending on recycled steel fiber properties.

4. Conclusion and Suggestions

In studies where steel fiber obtained from waste tires is used in concrete production, the geometric properties and amount of fiber vary. Depending on the length, diameter, and addition rate of the steel fiber used, there is expected to be a connection between the engineering properties being investigated. For this reason, establishing a good correlation between them will contribute to determining the performance results of applications with different properties. In this study, experimental data on the produced concrete were obtained from the literature to obtain a good correlation between the properties and usage rate of recycled steel fiber and the compressive and flexural strength of the produced concrete. The compressive strength and flexural strength values, two critical parameters reflecting the mechanical properties of the produced concretes, were successfully predicted by artificial neural networks depending on the length, diameter, and percentage of steel fiber used. Graphical analysis of the results obtained by artificial neural networks and the results of the experimental data obtained from the literature were analyzed graphically. Correlation coefficients between laboratory results and prediction values were determined. From the graphs obtained, it is seen that the data predictions made with artificial neural networks within the scope of this study are well approximated to the experimental data. Excellent correlation coefficients and low error rates were obtained for the compressive and flexural strengths of concretes produced depending on the properties of recycled steel fiber. The evaluations of the results of the study are given below.

- Literature research was reviewed, and a - dimensional graph showing the effect of geometrical properties and usage rate of recycled steel fiber on the compressive and flexural strength values of the concretes produced was created.
- Levenberg-Marquardt method was preferred to generate the prediction function with artificial neural networks. A single hidden layer and ten neurons were used. Training, testing, and validation data distribution is random at 70%, 15%, and 15%.
- Equations showing the relationship between the compressive strength predicted by ANN and the actual values are given. The highest correlation coefficient is 0.91. The closeness of this number to 1 indicates the accuracy of the generated function.
- The correlation coefficient of the equation showing the relationship between the target flexural strength values obtained by ANN and the laboratory results is relatively high, and it was received at 0.98.

 The compressive strength values obtained from laboratory tests varied between 27.44 MPa and 61.69 MPa, while the compressive strength values predicted by ANN varied between 30.27 MPa and 61.63 MPa.
 While the target flexural strength values varied between 2.54 MPa and 9.86 MPa, the values obtained with the ANN model varied between 2.89 MPa and 10.09 MPa.
 Depending on the input parameters of length, diameter, and percentage of use, which reflect the properties of recycled steel fiber, a 5-dimensional graph was created in which the compressive and flexural strength values predicted by artificial neural networks and the actual values were presented together. Thus, it was possible to compare the output data with the laboratory results practically.
 Based on the findings, it is concluded that the recycled steel tire-reinforced concrete parameters can be well represented by artificial neural networks, and the presented model can be used as an excellent alternative to laboratory studies for further research.
 Considering the importance of using recycled waste materials in concrete production, waste tyre was used in this study. Thanks to artificial neural networks, time, energy, and cost losses that occur as a result of laboratory experiments can be minimized. In the future, it is aimed to make the effect of using waste tires more known by using various modeling and optimization methods different from this study.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] H. Abdolpour, P. Niewiadomski, and Ł. Sadowski, "Recycling of steel fibres and spent equilibrium catalyst in ultra-high performance concrete: Literature review, research gaps, and future development," *Constr. Build. Mater*., vol. 309, no. 125147, p. 125147, 2021.

[2] J. D. Martínez, N. Puy, R. Murillo, T. García, M. V. Navarro, and A. M. Mastral, "Waste tyre pyrolysis-A review. *Renewable and sustainable energy reviews,"* vol. 23, pp. 179–213, 2013.

[3] B. S. Thomas and R. C. Gupta, "A comprehensive review on the applications of waste tire rubber in cement concrete," *Renew. Sustain. Energy Rev*., vol. 54, pp. 1323–1333, 2016.

[4] L. Xu, Y. Jiang, and R. Qiu, "Parametric study and global sensitivity analysis for co-pyrolysis of rape straw and waste tire via variance-based decomposition," Bioresour. Technol., vol. 247, pp. 545–552, 2018.

[5] C. F. Revelo, M. Correa, C. Aguilar, and H. A. Colorado, "Waste tire rubber powders based composite materials," *in REWAS Cham: Springer International Publishing*, pp. 437–445, 2019.

[6] P. T. Williams, S. Besler, D. T. Taylor, and R. P. Bottrill, "Pyrolysis of automotive tyre waste," *Journal of the Institute of Energy*, vol. 68, pp. 11–21, 1995.

[7] H. Tlemat, Steel fibres from waste tyres to concrete: testing, modelling and design (Doctoral dissertation). 2004.

[8] D. Y. C. Leung and C. L. Wang, "Kinetic study of scrap tyre pyrolysis and combustion," *J. Anal. Appl. Pyrolysis*, vol. 45, no. 2, pp. 153–169, 1998.

[9] ETRMA-European and Rubber Manufacturers' Association. http://www. etrma.org/uploads/Modules/Documentsmanager/20150706_824 etrma_trifold_ 05-15_ final_ print.pdf accesed 01 July 2019.

[10] D. Foti, "Preliminary analysis of concrete reinforced with waste bottles PET fibers," *Constr. Build. Mater.*, vol. 25, no. 4, pp. 1906–1915, 2011.

[11] S. B. Kim, N. H. Yi, H. Y. Kim, J.-H. J. Kim, and Y.-C. Song, "Material and structural performance evaluation of recycled PET fiber reinforced concrete," *Cem. Concr. Compos.*, vol. 32, no. 3, pp. 232–240, 2010.

[12] C. Frazão, J. Barros, J. A. Bogas, V. García-Cortés, and T. Valente, "Technical and environmental potentialities of recycled steel fiber reinforced concrete for structural applications*," J. Build. Eng.*, vol. 45, no. 103579, p. 103579, 2022.

[13] F. Grzymski et al., "Mechanical properties of fibre reinforced concrete with recycled fibres. Construction and Building Materials, 198," *Journal of Cleaner Production*, vol. 194, pp. 112–126, 2018.

[14] L. Lourenço, Z. Zamanzadeh, J. A. O. Barros, and M. Rezazadeh, "Shear strengthening of RC beams with thin panels of mortar reinforced with recycled steel fibres*," J. Clean. Prod.,* vol. 194, pp. 112–126, 2018.

[15] N. Lamba, R. Raj, and P. Singh, "The effects of recycled carbon fibers on the mechanical properties of high-strength concrete and its resilience to impact," *Iran. J. Sci. Technol. Trans. Civ. Eng.*, 2023.

[16] M. G. Azandariani, M. Vajdian, K. Asghari, and S. Mehrabi, "Mechanical properties of polyolefin and polypropylene fibers-reinforced concrete-An experimental study," *Composites Part C*: Open Access, vol. 12, 2023.

[17] S. Kakooei, H. M. Akil, M. Jamshidi, and J. Rouhi, "The effects of polypropylene fibers on the properties of reinforced concrete structures," *Constr. Build. Mater.*, vol. 27, no. 1, pp. 73–77, 2012.

[18] C. Devi, D. S. Vijayan, R. Nagalingam, and S. Arvindan, 'A review of the implementations of glass fiber in concrete technology, *Materials Today: Proceedings,* 62, 2010-2015, 2022.

[19] M. D. Bhat, M. U. Rehman, I. Shafi, A. Parveen, A. Fayaz, B. A. Malik, and F. Bashir, The effect of polypropylene and steel fibers on the properties of concrete at normal and elevated temperatures—a review, *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, 1-19, 2022.

[20] M. A. Yimer and T. W. Aure, "Numerical investigation of reinforced concrete and steel fiber-reinforced concrete exterior beam-column joints under cyclic loading," *Iran. J. Sci. Technol. Trans. Civ. Eng.*, vol. 46, no. 3, pp. 2249–2273, 2022.

[21] M. J. Ashraf, M. Idrees, and A. Akbar, "Performance of silica fume slurry treated recycled aggregate concrete reinforced with carbon fibers," J. Build. Eng., vol. 66, no. 105892, p. 105892, 2023.

[22] F. Shabbir, Q. uz Z. Khan, M. U. Rashid, A. Muslim, and A. Raza, "The efficiency of hybrid fibers (steel and carbon fibers) in controlling micro cracking of concrete beams," *Iran. J. Sci. Technol. Trans. Civ. Eng.*, vol. 47, no. 3, pp. 1425–1445, 2023.

[23] H. Zhong and M. Zhang, "Experimental study on engineering properties of concrete reinforced with hybrid recycled tyre steel and polypropylene fibres*," J. Clean. Prod.,* vol. 259, no. 120914, p. 120914, 2020.

[24] S. A. Yildizel, Y. O. Özkılıç, and A. Yavuz, "Optimization of waste tyre steel fiber and rubber added foam concretes using Taguchi method and artificial neural networks," *Structures*, vol. 61, no. 106098, p. 106098, 2024.

[25] M. Doğruyol, E. Ayhan, and A. Karaşin, "Effect of waste steel fiber use on concrete behavior at high temperature," *Case Studies in Construction Materials*, 20, e03051, 2024.

[26] Z. Majeed Abed, W. L. Khalil, and H. K. Ahmed, "Tire wastes and durability: the influence of crumb rubber and recycled steel fibre on properties of roller compacted concrete," *International Journal of Pavement Engineering,* 25(1), 2372628, 2024.

[27] E. Ayhan, M. Doğruyol, F. Kıpçak, and A. Karaşin, "Atık Taşıt Lastik Çelik Lifinin Beton Davranışına Etkisi.'' *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 15(2), 525-532, 2024.

[28] H. Alabduljabbar, H. Mohammadhosseini, M. Md. Tahir, and R. Alyousef, "Green and sustainable concrete production using carpet fibers waste and palm oil fuel ash," *Mater. Today,* vol. 39, pp. 929–934, 2021.

[29] K. M. Liew and A. Akbar, The recent progress of recycled steel fiber reinforced concrete. *Construction and Building Materials,* 232. 2020.

[30] M. Leone, G. Centonze, D. Colonna, F. Micelli, and M. A. Aiello, "Fiber-reinforced concrete with low content of recycled steel fiber: Shear behaviour," *Constr. Build. Mater.*, vol. 161, pp. 141–155, 2018.

[31] G. B. Golpasand, M. Farzam, and S. S. Shishvan, "Behavior of recycled steel fiber reinforced concrete under uniaxial cyclic compression and biaxial tests," *Constr. Build. Mater.*, vol. 263, no. 120664, p. 120664, 2020.

[32] M. Smrkić, D. Damjanović, and A. Baričević, ''Application of recycled steel fibres in concrete elements subjected to fatigue loading,'' *Građevinar,* 69(10.), 893-905, 2017.

[33] A. B. Ekmen, "Evaluation of SPT-N values and internal friction angle correlation using artificial intelligence methods in granular soils," *Soil Res.*, 2023.

[34] J. Dong and S. Hu, "The progress and prospects of neural network research," *Information and Control*, vol. 26, no. 5, pp. 360–368, 1997.

[35] M. Açikgenç, M. Ulaş, and K. E. Alyamaç, "Using an artificial neural network to predict mix compositions of steel fiber-reinforced concrete," *Arab. J. Sci. Eng.*, vol. 40, no. 2, pp. 407–419, 2015.

[36] M. Nasir, U. Gazder, M. Maslehuddin, O. S. Baghabra Al-Amoudi, and I. A. Syed, "Prediction of properties of concrete cured under hot weather using multivariate regression and ANN models," *Arab. J. Sci. Eng.*, vol. 45, no. 5, pp. 4111–4123, 2020.

[37] R. R. Mahesh, and D. Sathyan, Modelling the hardened properties of steel fiber reinforced concrete using ANN, *Mater. Today Proc*, vol. 49, 2021.

[38] K. Khan, M, B. A. Salami, M. N. Amin, I. Ahamd, AA. Alabdullah, ..., and F. E. Jalal. ''Estimating flexural strength of FRP reinforced beam using artificial neural network and random forest prediction model,'' *Polymers,* 14(11), 2270, 2022.

[39] N. Abunassar, M. Alas, and S. I. A. Ali, "Prediction of compressive strength in self-compacting concrete containing fly ash and silica fume using ANN and SVM," *Arab. J. Sci. Eng.*, vol. 48, no. 4, pp. 5171–5184, 2023.

[40] Afshoon, M. Miri, and S. R. Mousavi, "Using the Response Surface Method and Artificial Neural Network to Estimate the Compressive Strength of Environmentally Friendly Concretes Containing Fine Copper Slag Aggregates," *Iranian Journal of Science and Technology*, vol. 47, no. 6, pp. 3415–3429, 2023.

[41] A. B. Ekmen, "SPT-CPT İlişkisinin Yapay Zeka Desteğiyle Çeşitli Zemin Tipleri İçin Araştırılması," *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi*, vol. 39, no. 2, pp. 204–216.

[42] M. R. Akbarzadeh, H. Ghafourian, A. Anvari, R. Pourhanasa, and M. L. Nehdi, ‘’Estimating compressive strength of concrete using neural electromagnetic field optimization,’’ *Materials,* 16(11), 4200, 2023.

[43] T. F. Awolusi, D. O. Oguntayo, D. O. Oyejobi, B. D. Agboola., O. O. Akinkurolere, O. E. Babalola, and S.K. Shukla, “Performance evaluation of discontinuous coconut and steel fibers as reinforcement in concrete using the artificial neural network approach,’’ *Cogent Engineering,* 9(1), 2022.

[44] E. M. Golafshani, and A. Behnood, “Predicting themechanical propertie of sustainable concrete containing waste foundry sand using multi-objective ANN approach,’’ *Construction and Building Materials*, 291, 123314, 2021.

[45] S. Ray, M. Haque, T. Ahmed, and T.T. Nahin, “Comparison of artificial neural network (ANN) and response surface methodology (RSM) in predicting the compressive and splitting tensile strength of concrete prepared with glass waste and tin (Sn)can fiber,“ *Journal of king saud university-engineering sciences,* 35(3), 185–199, 2023.

[46] H.H. Almasaeid, A. Suleiman, and R. Alawneh, “Assessment of high-temperature damaged concrete using non-destructive tests and artificial neural network modelling,’’ *Case Studies in Construction Materials,* 16, e01080, 2022.

[47] P. G. Asteris and V. G. Mokos, “Concrete compressive strength using artificial neural networks,” *Neural Comput. Appl.,* vol. 32, no. 15, pp. 11807–11826, 2020.

[48] A. Shahmansouri, M. Yazdani, S. Ghanbari, H. Akbarzadeh Bengar, A. Jafari, and H. Farrokh Ghatte, “Artificial neural network model to predict the compressive strength of eco-friendly geopolymer concrete incorporating silica fume and natural zeolite,” *J. Clean. Prod.*, vol. 279, no. 123697, p. 123697, 2021.

[49] D. Fan et al., “Precise design and characteristics prediction of Ultra-High Performance Concrete (UHPC) based on artificial intelligence techniques,” *Cem. Concr. Compos.,* vol. 122, no. 104171, p. 104171, 2021.

[50] S. A. Yildizel, M. Uzun, M. A. Arslan, and T. Ozbakkaloglu, “The prediction and evaluation of recycled polypropylene fiber and aggregate incorporated foam concrete using Artificial Neural Networks,” *Constr. Build. Mater.,* vol. 411, no. 134646, p. 134646, 2024.

[51] S. Yadav and S. K. Tiwari, “The impact of end-of-life tires on the mechanical properties of fine-grained soil: A Review,” *Environ. Dev. Sustain.,* vol. 21, no. 2, pp. 485–568, 2019.

[52] G. Centonze, M. Leone, and M. A. Aiello, “Steel fibers from waste tires as reinforcement in concrete: A mechanical characterization,” *Constr. Build. Mater.,* vol. 36, pp. 46–57, 2012.

[53] E. Martinelli, A. Caggiano, and H. Xargay, “An experimental study on the post-cracking behaviour of Hybrid Industrial/Recycled Steel Fibre-Reinforced Concrete,” *Constr. Build. Mater.,* vol. 94, pp. 290–298, 2015.

[54] M. Leone, G. Centonze, D. Colonna, F. Micelli, and M. A. Aiello, “Experimental study on bond behavior in fiber-reinforced concrete with low content of recycled steel fiber,” *J. Mater. Civ. Eng.,* vol. 28, no. 9, p. 04016068, 2016.

[55] B. Nasir, “Steel fiber reinforced concrete made with fibers extracted from used tyres,” *Master of Science in Civil Engineering dissertation,* 2009.

[56] M. A. Aiello, F. Leuzzi, G. Centonze, and A. Maffezzoli, "Use of steel fibres recovered from waste tyres as reinforcement in concrete: pull-out behaviour, compressive and flexural strength," *Waste Manag*., vol. 29, no. 6, pp. 1960–1970, 2009.

[57] M. Pająk, M. Krystek, M. Zakrzewski, and J. Domski, ''Laboratory investigation and numerical modelling of concrete reinforced with recycled steel fibers,'' *Materials,* 14(8), 1828., 2021.

[58] D. Revuelta, P. Carballosa, J. L. García Calvo, and F.Pedrosa, ''Residual strength and drying behavior of concrete reinforced with recycled steel fiber from tires,'' *Materials,* 14(20), 6111,2021.

[59] S. Rossli and I. Ibrahim, Mechanical properties of recycled steel tire fibres in concrete, *Engineering, Materials Science, Environmental Science,* 2012.

[60] Y. Zhang and L. Gao, "Influence of tire-recycled steel fibers on strength and flexural behavior of reinforced concrete," *Adv. Mater. Sci. Eng.*, vol. 2020, pp. 1–7, 2020.

[61] X. Qin and S. Kaewunruen, "Environment-friendly recycled steel fibre reinforced concrete," *Constr. Build. Mater*., vol. 327, no. 126967, p. 126967, 2022.

[62] Y. Gao et al., "Experimental study on recycled steel fiber-reinforced concrete under repeated impact," *Rev. Adv. Mater. Sci*., vol. 62, no. 1, 2023.

[63] Y. Avci and A. B. Ekmen, "Artificial intelligence assisted optimization of rammed aggregate pier supported raft foundation systems based on parametric three-dimensional finite element analysis," *Structures,* vol. 56, no. 105031, p. 105031, 2023.

[64] A. B. Ekmen and Y. Avci, "Artificial Intelligence-Assisted Optimization of Tunnel Support Systems Based on the Multiple Three-Dimensional Finite Element Analyses Considering the Excavation Stages," *Iranian Journal of Science and Technology,* vol. 47, no. 3, pp. 1725–1747, 2023.

[65] Matlab, version R2017a, *MathWorks,* 2017.

[66] T. Gupta, K. A. Patel, S. Siddique, R. K. Sharma, and S, "Chaudhary Prediction of mechanical properties of rubberised concrete exposed to elevated temperature using ANN'', *Measurement,* 147, 106870, 2019.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1047-1066 YEAR: 2024
DOI: 10.17798/bitlisfen.1500249

Comprehensive Energy and Exergy Analysis of a Pressurized Water Reactor Driven Multi-Stage Flash Desalination Plant

Erdem AKYÜREK[1], Tayfun TANBAY[1*]

[1]*Bursa Technical University, Faculty of Engineering and Natural Sciences, Department of Mechanical Engineering, Bursa, Türkiye*
(ORCID: 0000-0002-8856-2387) (ORCID: 0000-0002-0428-3197)

Keywords: Nuclear desalination, Multi-stage flash, Pressurized water reactor, Energy analysis, Exergy analysis.

Abstract

Nuclear energy-based seawater desalination is an environmentally friendly freshwater production approach. This study introduces a novel thermodynamic model integrating a pressurized water reactor's (PWR) secondary cycle with a multi-stage flash (MSF) desalination facility to enhance freshwater production. The impacts of the design and operating conditions on thermal efficiency, utilization factor, gain output ratio, exergy efficiency, coefficient of ecological performance for cogeneration and exergy destruction factor are investigated. Results reveal that a higher live steam temperature and a reheater mass flow rate ratio is preferable for a better nuclear desalination performance. A larger freshwater production capacity is preferable for a better utilization factor, however increasing the capacity tends to decrease thermal efficiency, coefficient of ecological performance for cogeneration and exergy destruction factor. The selection of steam extraction location is important for very large scale plants, and the outlet of moisture separator is determined to be the best option. Parametric analysis shows that plant's performance can be significantly improved by adjusting the design conditions. Thermal and exergy efficiencies of an optimized plant configuration are 3.01% and 4.70% higher, respectively as compared to a base plant. It is also found that steam generator and MSF unit cause 3.2% and 82% of the total irreversibility rate of PWR's secondary cycle and MSF facility, respectively, and have the highest irreversibility rates for these sections of the plant.

1. Introduction

With the impact of industrialization, agricultural production and rapid population growth, the demand for freshwater is constantly increasing. Water pollution and arid climatic conditions reduce the per capita water supply worldwide and cause water scarcity. The water crisis has a significant negative impact on the development of countries owing to its influence on economic activities and the people's quality of life [1]. Approximately 97.5% of the water available on the Earth's surface is sea and ocean water, while 2.5% is freshwater including glaciers, groundwater, streams and lakes, and only less than 1% of freshwater is easily accessible and usable [2]. In a United Nations report [3], it is stated that the world's water stress level is about 18% in 2018. According to the report, approximately 10% of the global population reside in areas with water shortages, and it is pointed out that the requirement for water will increase in the coming years due to the impact of population density.

Seawater desalination, which is an option to meet freshwater demand, is the process of separating purified water from seawater having high salt concentration by consuming energy. The required energy for desalination processes can be provided from fossil-based, renewable or nuclear energy. Compared to alternative energy sources nuclear power plants (NPPs) have the largest capacity factor

*Corresponding author: tayfun.tanbay@btu.edu.tr
Received:12.06.2024, Accepted: 10.12.2024

[4, 5]. The high availability of nuclear energy used in desalination applications can eliminate the energy storage requirement problems of renewables. In addition, nuclear power is advantageous to fossil-fueled plants since it can completely eliminate CO_2 emissions. Fossil fuel-driven MSF and multi-effect distillation (MED) desalination plants emit approximately 22.5 kg CO_2 and 17.5 kg CO_2 per 1 m^3 of freshwater produced, respectively [6]. It is possible to significantly reduce CO_2 emissions by utilizing clean energy sources such as nuclear for the desalination of seawater [7].

In a nuclear reactor driven desalination plant, the fission energy of the reactor is converted into thermal energy, and part of this energy is used for the desalination of seawater. MSF, MED and thermal vapor compression (TVC) are used as distillation processes while reverse osmosis (RO) is utilized as a membrane process in nuclear desalination facilities. RO, MSF and MED processes account for 70%, 18% and 7% of the global installed desalination capacity, respectively [8]. MSF, MED and RO desalination processes consume total energy in the range of $18.3-28.5\ \mathrm{kWh/m^3}$, $14.2-21.6\ \mathrm{kWh/m^3}$ and $4-6\ \mathrm{kWh/m^3}$, respectively [9]. Although the energy use of MSF is high, it can be operated with high capacity and its maintenance is easier [10], and also MSF has a high water quality in the range $2-10$ ppm [9].

Seawater desalination and the integration of these processes to nuclear energy have been the topic of many technical and/or economic studies. Four different coupling alternatives of cogeneration systems were analyzed with Desalination Thermodynamic Optimization Program (DE-TOP) software developed by the International Atomic Energy Agency (IAEA) [11]. IAEA also developed the Desalination Economic Evaluation Program (DEEP) for the economic evaluation of desalination plants driven by nuclear, fossil and renewable sources [12]. An economic analysis of the MSF process coupled to nuclear and fossil fueled cogeneration plants was performed by Faibish and Ettouney [13]. The impacts of production capacity and top brine temperature (TBT) on the cost were investigated, and it was found that the nuclear-based MSF produced freshwater with a cost that is lower than fossil fuel-based distillation. In another study, the costs of water desalinated by MSF, MED and RO to meet the potable water requirement in two different site conditions in Libya were determined and compared with fossil-based desalination [14].

The integration of International Reactor Innovative and Secure plant to MSF and RO for nuclear desalination was considered by [15]. Economic analysis revealed that the plant cost of the nuclear MSF desalination plant was minimized at a TBT of $82-83$℃. Adak and Tewari [16] investigated the coupling of an MSF facility to a pressurized heavy water reactor (PHWR), and found that a water production rate of $4500\ \mathrm{m^3/day}$ with a gain output ratio (GOR) of 9 can be obtained at the expense of a 3.3 MW decrease in the net power output of the NPP. A technical and safety study on the thermal coupling aspects for isolation purposes of this plant was presented by [17]. Safety features were evaluated to prevent any radioactivity entering the product water in the integrated plant that employed two isolation heat exchangers. Tian et al. [18] compared the electricity and freshwater costs of nuclear and fossil fuel powered MSF desalination plants. The parameters considered in the assessment were the discount rate, fossil fuel price, power plant investment and desalination plant size. The analysis showed that the nuclear option was economically advantageous to fossil fuel based desalination.

The exergy analysis of a MSF desalination plant was performed by Kahraman and Cengel [19], which showed that the MSF unit cause the largest irreversibility and its exergetic performance could be improved by utilizing more flashing stages. The thermal coupling of MSF and MED facilities to a system-integrated modular advanced reactor (SMART) was studied in [20]. Three coupling techniques were compared, and it was found that the heat pipe loop was the optimum plant integration method. Thermoeconomic multiobjective optimization of a PWR coupled to an MSF plant was carried out to minimize the product water and electricity costs [21, 22]. Thermoflex simulator was used for thermodynamic modelling, and the optimization was carried out with a genetic algorithm. Alonso et al. [23] presented a comparative economic analysis by coupling MSF, MED, RO, MSF-RO and MED-RO with two different PWRs. The electricity and freshwater costs of cogeneration applications were determined for three different discount rates. MED was found to be less expensive than MSF and the coupling of RO to MSF and MED would improve the competitiveness of these thermal desalination technologies.

Yan et al. [24] examined the integration of a high-temperature gas cooled reactor (HTGR) to an MSF plant. The analysis revealed that an incrementally loaded MSF system would increase the freshwater production by 45% and use of reactor thermal power by 16% compared to the conventional MSF desalination. The techno-economic evaluation of cogeneration options for a small modular reactor (SMR) coupled to MSF, MED and RO were carried out by [25]. Three different coupling strategies of the

MSF and MED with the reactor were considered, and the relationship between electricity and freshwater outputs was investigated. The results indicated that RO causes the least cost and the SMR provided flexible operation for all desalination processes. Exergetic and thermo-economic analyses of SMART reactor-based MSF and MED-TVC desalination are performed by Priego et al. [26]. The effect of steam extraction nodes from the secondary cycle of the reactor on the cost of fresh water is assessed for three gain output ratios and four discount rates. Techno-economic analysis of two SMRs, namely SMART and CAREM (Central Argentina de Elementos Modulares), was carried out by Khan et al. [27] for nuclear desalination with MSF, MED and RO. In another techno-economic analysis, the Karachi NPP was considered for nuclear desalination with MSF, MED, RO, MSF-RO and MED-RO approaches [28]. The results indicated that MED was economically advantageous.

Polat and Dinçer [29] compared MSF, MED, RO, MSF-RO and MED-RO based nuclear desalination methods. Akkuyu NPP, which is a Water-Water Energy Reactor (WWER) under construction, is considered as the heat and/or electricity source for desalination and DEEP is used to determine the freshwater production costs. The integration of an experimental gas cooled reactor to MSF for nuclear desalination was studied by Dewita et al. [30]. The impact of steam extraction nodes on freshwater production capacity was analyzed by considering two options. Sadeghi et al. [31] investigated the integration of the Bushehr NPP with hybrid MSF-RO and MED-RO desalination processes for three coupling options and performed the techno-economic analysis of the plant. The total cost of freshwater was determined, and it was found that extracting steam from the condenser outlet was the best alternative. Finally, in a recent work [32], an exergoeconomic analysis of a PWR coupled to MSF and RO for nuclear desalination was carried out. Aspen HYSYS software was used for the modelling of the PWR's secondary cycle.

The abovementioned studies do not present a thermodynamic model that involves both the nuclear reactor's energy conversion system and MSF facility of a nuclear desalination plant. Therefore, in order to fill this research gap, this paper introduces a novel thermodynamic model integrating a PWR's secondary cycle with an MSF desalination facility to produce electricity and freshwater, and presents a comprehensive thermodynamic analysis of the cogeneration system. The novelties of this study are:

- A detailed thermodynamic model considering the components of both the PWR's secondary cycle and MSF facility is built for a comprehensive energy and exergy analysis of the plant.
- The impacts of the design and operating conditions of i) reactor thermal power, ii) live steam temperature, iii) reheater mass flow rate ratio, iv) reheater temperature, v) process steam extraction node, vi) freshwater production capacity, vii) TBT, viii) MSF unit throttling mass flow rate ratio ix) seawater temperature, and x) seawater salinity on the objective functions of i) thermal efficiency, ii) utilization factor, iii) gain output ratio, iv) exergy efficiency, v) coefficient of ecological performance, and vi) exergy destruction factor are investigated.

2. System Description

The detailed schematic representation of the PWR driven MSF desalination plant is illustrated in Figure 1. Nuclear facility is composed of two cycles. Fission energy of the nuclear reactor is transferred to liquid water circulating in the primary loop, which is coupled to the secondary loop through a steam generator (SG). Steam is obtained at the outlet of the cold side of the SG (state 4) by heat transfer from the primary loop. The majority of the steam enters the high-pressure turbine (HPT) to produce electrical power output, while the rest is directed to reheater (RH) for reheating the steam to low-pressure turbine (LPT) inlet conditions. The moisture separator (MS) transfers the liquid water to a feedwater heater (FWH) to improve the service life of the plant. Steam leaving the RH enters the LPT to produce additional power and the expanded steam then enters the condenser (C). Secondary cycle contains seven FWHs to increase the thermal efficiency of the plant and the pumps (P) are utilized to increase the pressure of water to SG inlet condition. FWHs also provide flexibility in operation of the plant by enabling the transfer of process steam to the MSF facility at different conditions. Green lines represent the possible steam extraction locations for desalination. Steam is extracted from only one of the states of 52, 53, …, 63 and returns the cycle through C.

The secondary cycle of the PWR is coupled to the MSF facility through a heat exchanger (HX). In addition to transferring the energy of the process steam to the MSF facility, the HX acts as an isolation loop that prevents the contamination of both nuclear and MSF cycles. The desalination facility includes an MSF unit consisting of heat recovery and heat rejection sections with multiple horizontal in-line evaporators, a brine heater, a throttling valve, pumps and a vacuum system. Each evaporator has a flashing chamber where seawater evaporates and a condensation section where the pure water is distilled.

Seawater entering the facility at state 0 is first directed to the condensation section of the evaporators to condense the pure water vapor and then exits through the heat rejection section of MSF unit at state 69. Part of the seawater is discharged back to the sea, while the remainder (state 70) is throttled and utilized as make-up feedwater in the MSF unit. The portion of the feedwater leaving the heat rejection section at state 74 is heated by recirculation to the evaporators in the heat recovery section. The feedwater then enters the brine heater (BH), and it is heated to the TBT with the process heat of the steam that is extracted from the PWR's secondary loop through the isolation HX. The heated feedwater enters the flash chamber of the first evaporator at state 77, and it is vaporized due to pressure drop. The pure water vapor passes through the demister, which prevents the passage of the salt, and then it condenses by transferring heat to the seawater in the condensation section and is collected on the distillation tray. This operation is repeated in each stage of the MSF unit and the product water leaves the desalination facility at state 72. The high-salinity brine water leaves the MSF unit at state 73, and it is directed back to the sea.

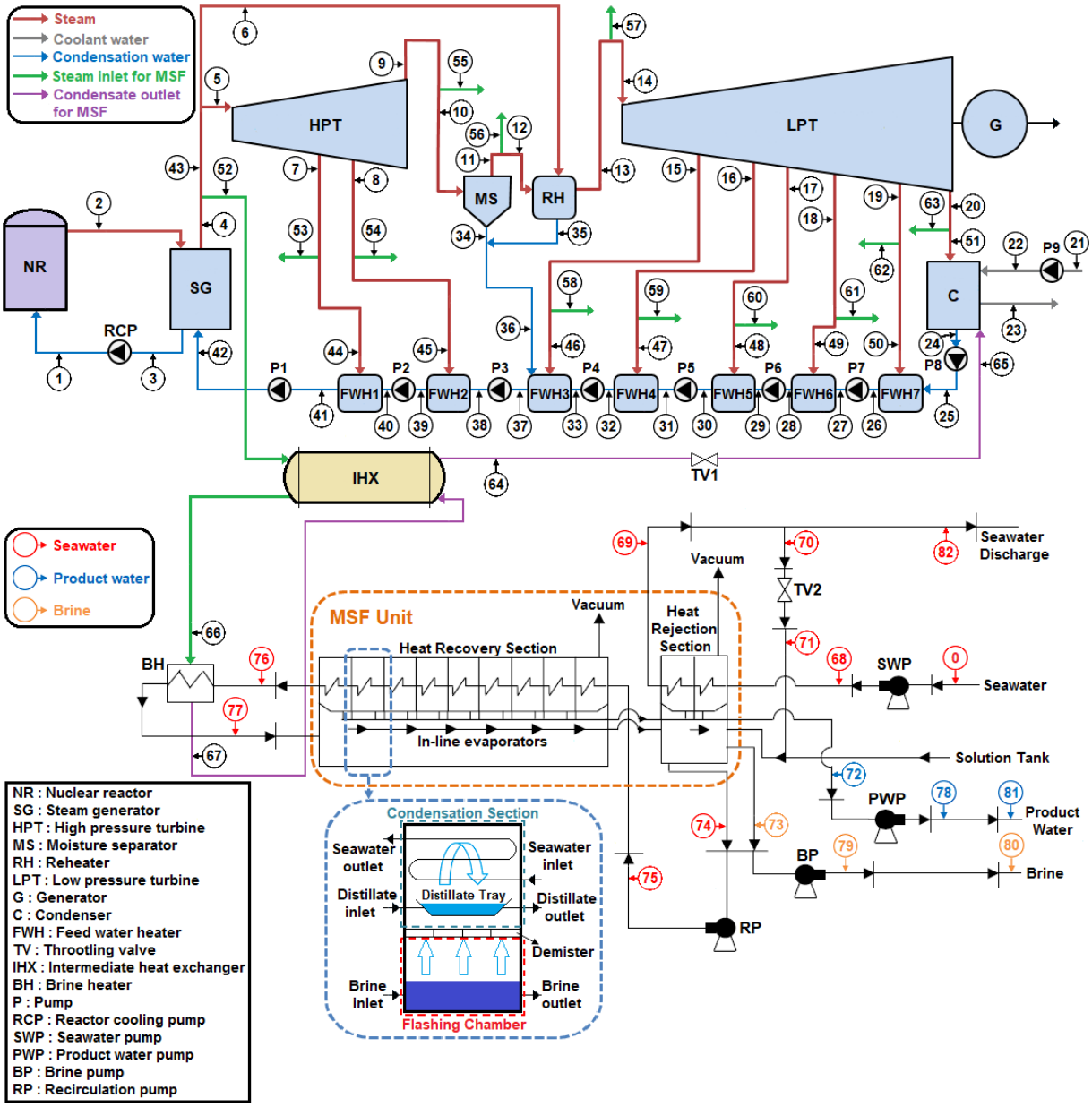


Figure 1. Illustration of the PWR driven MSF desalination plant (adapted from [11, 19]).

The plant design presented in Figure 1 has several advantages over alternative configurations. First of all, the nuclear-driven cogeneration system completely eliminates the CO_2 that would be emitted from a fossil fuel-based cogeneration plant, and hence it offers an environmental benefit. The design is also

advantageous to renewable-based plants since nuclear energy is superior to renewables in terms of capacity factor. The use of PWR instead of alternative reactor technologies such as PHWR, SMART and HTGR is another advantage of the configuration presented in Figure 1, since PWRs are the most widely utilized reactors in the nuclear industry. Therefore, the PWR technology is more accessible for nuclear desalination as compared to alternative reactors. As for the choice of MSF process for desalination, it is advantageous to alternative desalination approaches with its large production capacity and high water quality. Freshwater with a high level of purity is a valuable commodity since it can be utilized as an input for a hydrogen production facility employing electrolysis as the production method.

3. Modeling of the System

In this section, first the main assumptions employed in the modeling of the plant are given. Then, the energy-exergy analyses of the nuclear desalination plant are presented. Following the definition of saline water properties, finally, the objective functions are introduced.

3.1. Assumptions

The main assumptions for the modeling of the nuclear desalination plant are as follows:

- All processes are steady-flow and hence the parametric analyses are time-independent.
- The change in kinetic and potential energies is neglected.
- All plant components are adiabatic except the MSF unit.
- MSF unit is nonadiabatic [19, 33]. Exergy balance equations presented in Ref [19] for the MSF unit, and in Ref [33] for the heat recovery and rejection sections indicate that the MSF unit is adiabatic. However, calculations based on the thermodynamic data presented in Table 1 of Ref [19] and in Tables 4 and 5 of Ref [33] show that the component is nonadiabatic.
- There is no frictional pressure loss in the cycle.
- Isentropic efficiencies are used to factor in the irrevesibilities of turbines, pump and generator.
- Water leaves the SG and MS at state 11 as a saturated vapor, while it leaves the MS as a saturated liquid at state 34.
- Liquid water enters nuclear facility through the condenser for cooling at state 21 and the MSF facility at state 0 at ambient temperature.
- A temperature difference of 5℃ is present in both sides of C and the inlet and outlet of the cold side of C (i.e., $T_{24} = T_{23} + 5℃$ and $T_{23} = T_{22} + 5℃$) [11].
- Steam extraction from PWR's secondary loop is made through only one of the locations 52, …, 63.
- Desalination capacity is a function of the TBT, T_{77} [34]. The amount of product water, i.e., $\dot{m}_{72}$, depends on the TBT as

$$\dot{m}_{72} = \dot{m}_{74}\left[1 - (1-y)^{(TBT - T_{73})/\Delta T_{st}}\right] \quad (1)$$

where $\dot{m}_{74}$ is the recycle brine mass flow rate, y is the ratio of sensible heat to latent heat of salt-water mixture in the MSF unit, ΔT_{st} is the temperature decrease of the brine per MSF unit stage and T_{73} is the brine temperature.

- The feed seawater mass flow rate to the MSF unit, i.e., $\dot{m}_{71}$, is related to the product water mass flow rate as [34]

$$\dot{m}_{71} = \dot{m}_{72}\frac{Sal_{73}}{Sal_{73} - Sal_{71}} \quad (2)$$

where Sal_{71} and Sal_{73} are the salinities of the seawater and brine, respectively. The brine salinity is limited by environmental regulations.

- The temperature of the seawater increases by 8℃ and its pressure decreases by 30% in the heat rejection section of the MSF unit (i.e., $T_{69} = T_0 + 8℃$ and $P_{69} = 0.7P_{68}$) [19]. Also, there exists a temperature difference of 5℃ on the cold side of the BH (i.e., $T_{77} = T_{76} + 5℃$).

3.2. Energy and Exergy Analysis

The energy analysis of the plant is carried out by applying the conservation of mass and energy equations to plant components

$$\sum_{in} \dot{m} = \sum_{out} \dot{m} \quad (3)$$

$$\dot{Q}_{in} + \dot{W}_{in} + \sum_{in} \dot{m}h = \dot{Q}_{out} + \dot{W}_{out} + \sum_{out} \dot{m}h \quad (4)$$

where $\dot{m}$ is the mass flow rate, $\dot{Q}$ and $\dot{W}$ are the rates of heat and work transfer, respectively, and h represents the enthalpy. Eqs. (3) and (4) yield the heat transfer rate from the primary cycle of nuclear plant to the secondary cycle through SG ($\dot{Q}_{SG}$), the heat

transfer rate from C, the power production rates of HPT and LPT ($\dot{W}_{HPT}$ and $\dot{W}_{LPT}$), the power consumption rates of Ps ($\dot{W}_P$), heat transfer rate from the nuclear plant to MSF plant through HX ($\dot{Q}_p$) and heat loss rate of the MSF unit ($\dot{Q}_{MSF}$). The energy balance equations are presented in Table 1.

Table 1. Energy balance equations for the components of the nuclear desalination plant.

Component	Energy balance equations
	Nuclear plant
SG	$\dot{Q}_{SG} = \dot{m}_4(h_4 - h_{42})$
HPT	$\dot{W}_{HPT} = \dot{m}_5 h_5 - (\dot{m}_7 h_7 + \dot{m}_8 h_8 + \dot{m}_9 h_9)$
MS	$\dot{m}_{10} h_{10} = \dot{m}_{11} h_{11} + \dot{m}_{34} h_{34}$
RH	$\dot{m}_6(h_6 - h_{35}) = \dot{m}_{12}(h_{14} - h_{12})$
LPT	$\dot{W}_{LPT} = \dot{m}_{14} h_{14} - (\dot{m}_{15} h_{15} + \dot{m}_{16} h_{16} + \dot{m}_{17} h_{17} + \dot{m}_{18} h_{18} + \dot{m}_{19} h_{19})$
C	$\dot{m}_{22} h_{22} + \dot{m}_{51} h_{51} + \dot{m}_{65} h_{65} = \dot{m}_{23} h_{23} + \dot{m}_{24} h_{24}$
Ps	$\dot{W}_{P,j} = \dot{m}_{in,j}(h_{out,j} - h_{in,j}), \quad j = 1, \dots, 9$
FWHs	$\dot{m}_{out,j} h_{out,j} = \sum_{in,j} \dot{m} h, \quad j = 1, \dots, 7$
IHX	$\dot{Q}_p = \dot{m}_i(h_i - h_{64}) = \dot{m}_{66}(h_{66} - h_{67}), \ i = 52,53, \dots, \text{or } 63$
Flow mixing	$\dot{m}_{34} h_{34} + \dot{m}_{35} h_{35} = \dot{m}_{36} h_{36}$
TV1	$h_{65} = h_{64}$
	Desalination plant
SWP	$\dot{W}_{SWP} = \dot{m}_{68}(h_{68} - h_0)$
TV2	$h_{71} = h_{70}$
MSF unit	$\dot{Q}_{MSF} = (\dot{m}_{68} h_{68} + \dot{m}_{71} h_{71} + \dot{m}_{75} h_{75} + \dot{m}_{77} h_{77}) - (\dot{m}_{69} h_{69} + \dot{m}_{72} h_{72} + \dot{m}_{73} h_{73} + \dot{m}_{74} h_{74} + \dot{m}_{76} h_{76})$
BH	$\dot{m}_{76}(h_{77} - h_{76}) = \dot{m}_{66}(h_{66} - h_{67})$
PWP	$\dot{W}_{PWP} = \dot{m}_{72}(h_{78} - h_{72})$
BP	$\dot{W}_{BP} = \dot{m}_{73}(h_{79} - h_{73})$
RP	$\dot{W}_{RP} = \dot{m}_{74}(h_{75} - h_{74})$

The irreversibility rates of plant components are determined by applying the exergy balance equation

$$\left(1 - \frac{T_0}{T_b}\right)\dot{Q}_{in} + \dot{W}_{in} + \sum_{in} \dot{X}_i - \left(1 - \frac{T_0}{T_b}\right)\dot{Q}_{out} - \dot{W}_{out} - \sum_{out} \dot{X}_i - \dot{X}_{dest} = 0 \quad (5)$$

where $\dot{X}_{dest}$ is the exergy destruction rate, $\dot{X}_i$ is the exergy flow rate, while T_0 and T_b are the dead state and average boundary temperature, respectively. Eq. (5) can be used to determine the exergy destruction rates of nuclear desalination plant components, and the equations are presented in Table 2, where T_{SG} and T_{MSF} are the average boundary temperatures of SG and MSF unit, respectively. Exergy flow rate is defined as

$$\dot{X}_i = \dot{m}[h_i - h_0 - T_0(s_i - s_0)] \quad (6)$$

where h_0 and s_0 are the enthalpy and entropy at T_0.

Table 2. Exergy balance equations for the components of the nuclear desalination plant.

Component	Exergy balance equations
	Nuclear plant
SG	$\dot{X}_{dest,SG} = \dot{X}_{42} - \dot{X}_4 + \left(1 - \frac{T_0}{T_{SG}}\right)\dot{Q}_{SG}$
HPT	$\dot{X}_{dest,HPT} = \dot{X}_5 - \dot{X}_7 - \dot{X}_8 - \dot{X}_9 - \dot{W}_{HPT}$
MS	$\dot{X}_{dest,MS} = \dot{X}_{10} - \dot{X}_{11} - \dot{X}_{34}$
RH	$\dot{X}_{dest,RH} = \dot{X}_6 + \dot{X}_{12} - \dot{X}_{13} - \dot{X}_{35}$
LPT	$\dot{X}_{dest,LPT} = \dot{X}_{14} - \dot{X}_{15} - \dot{X}_{16} - \dot{X}_{17} - \dot{X}_{18} - \dot{X}_{19} - \dot{W}_{LPT}$
C	$\dot{X}_{dest,C} = \dot{X}_{22} + \dot{X}_{51} + \dot{X}_{65} - \dot{X}_{23} - \dot{X}_{24}$
Ps	$\dot{X}_{dest,P,j} = \dot{X}_{in,j} - \dot{X}_{out,j} + \dot{W}_{P,j}, \quad j = 1,\dots,9$
FWHs	$\dot{X}_{dest,FWH,j} = \sum_{in,j} \dot{X}_i - \dot{X}_{out,j}, \quad j = 1,\dots,7$
IHX	$\dot{X}_{dest,IHX} = \dot{X}_i + \dot{X}_{67} - \dot{X}_{64} - \dot{X}_{66}, \ i = 52,53,\dots,\text{or } 63$
Flow mixing	$\dot{X}_{dest,mix} = \dot{X}_{34} + \dot{X}_{35} - \dot{X}_{36}$
TV1	$\dot{X}_{dest,TV1} = \dot{X}_{64} - \dot{X}_{65}$
	Desalination plant
SWP	$\dot{X}_{dest,SWP} = \dot{X}_0 - \dot{X}_{68} + \dot{W}_{SWP}$
TV2	$\dot{X}_{dest,TV2} = \dot{X}_{70} - \dot{X}_{71}$
MSF unit	$\dot{X}_{dest,MSF} = \dot{X}_{68} + \dot{X}_{71} + \dot{X}_{75} + \dot{X}_{77} - \dot{X}_{69} - \dot{X}_{72} - \dot{X}_{73} - \dot{X}_{74} - \dot{X}_{76} - \left(1 - \frac{T_0}{T_{MSF}}\right)\dot{Q}_{MSF}$
BH	$\dot{X}_{dest,BH} = \dot{X}_{66} + \dot{X}_{76} - \dot{X}_{67} - \dot{X}_{77}$
PWP	$\dot{X}_{dest,PWP} = \dot{X}_{72} - \dot{X}_{78} + \dot{W}_{PWP}$
BP	$\dot{X}_{dest,BP} = \dot{X}_{73} - \dot{X}_{79} + \dot{W}_{BP}$
RP	$\dot{X}_{dest,RP} = \dot{X}_{74} - \dot{X}_{75} + \dot{W}_{RP}$
SW-Cooling	$\dot{X}_{dest,SW-Cooling} = \dot{X}_{69} - \dot{X}_{82}$
B-Cooling	$\dot{X}_{dest,B-Cooling} = \dot{X}_{73} - \dot{X}_{80}$
PW-Cooling	$\dot{X}_{dest,B-Cooling} = \dot{X}_{78} - \dot{X}_{81}$

3.3. Saline Water Properties

Desalination process alters the salinity of the feedwater and produces three streams as product, brine and discharge waters with different salinities. Therefore, the thermodynamic properties of the working fluid of the desalination plant have to be determined by considering it as a water-salt mixture. Enthalpy and entropy of such a mixture are determined by [19].

$$h_i = mf_{s,i}h_{s,i} + mf_{w,i}h_{w,i}, \tag{7}$$

$$s_i = mf_{s,i}s_{s,i} + mf_{w,i}s_{w,i} - \left(\frac{R_u}{M_{sw,i}}\right)\left(x_{s,i}\ln x_{s,i} + x_{w,i}\ln x_{w,i}\right) \tag{8}$$

where $i = 68,69,\dots,82$, mf is the mass fraction, x is the mole fraction, R_u is the universal gas constant and M is the molar mass, while the subscripts sw, s and w denote seawater, salt and water, respectively. The mass fraction of salt is expressed in terms of the salinity (Sal) of seawater having the unit ppm, and the mass fraction of water is:

$$mf_{s,i} = Sal_i(ppm) \times 10^{-6} \tag{9}$$

$$mf_{w,i} = x_{w,i}\frac{M_w}{M_{sw,i}} \tag{10}$$

The molar fractions of salt and water and the molar mass of saline water are determined by

$$x_{s,i} = \frac{M_w}{M_s\left(\frac{1}{mf_{s,i}} - 1\right) + M_w} \quad (11)$$

$$x_{w,i} = 1 - x_{s,i} \quad (12)$$

$$M_{sw,i} = x_{s,i}M_s + x_{w,i}M_w \quad (13)$$

Salt is an incompressible substance and its enthalpy and entropy are independent of pressure. On the other hand, these thermodynamic properties are functions of temperature and the temperature dependence of salt's enthalpy and entropy can be expressed as

$$h_{s,i} = h_{s,ref} + c_{p,s}\left(T_i - T_{ref}\right) \quad (14)$$

$$s_{s,i} = s_{s,ref} + c_{p,s} ln\left(\frac{T_i}{T_{ref}}\right) \quad (15)$$

where $c_{p,s}$ is the specific heat of salt, T_{ref} is a reference temperature and $h_{s,ref}$ and $s_{s,ref}$ are the corresponding enthalpy and the entropy values of salt. For $T_{ref} = 35$℃, $h_{s,ref} = 29.288\, kJ/kg$, $s_{s,ref} = 0.1009\, kJ/kgK$, while $c_{p,s} = 0.8368\, kJ/kgK$.

3.4. Objective Functions

The energetic performance of the nuclear desalination plant is investigated through three objective functions, namely thermal efficiency, utilization factor and gain output ratio defined as

$$\eta_{th} = \frac{\dot{W}_{net}}{\dot{Q}_{SG}} \quad (16)$$

$$\epsilon_u = \frac{\dot{W}_{net} + \dot{Q}_p}{\dot{Q}_{SG}} \quad (17)$$

$$GOR = \frac{\dot{m}_{distilled}}{\dot{m}_{steam}} = \frac{\dot{m}_{72}}{\dot{m}_{66}} \quad (18)$$

respectively. $\dot{Q}_{SG}$ and $\dot{Q}_p$ are given in Table 1 and $\dot{W}_{net}$ is the net electrical power output of the plant:

$$\dot{W}_{net} = \eta_{GEN}\left(\dot{W}_{HPT} + \dot{W}_{LPT}\right) - \left(\sum_{j=1}^{9} \dot{W}_{P,j} + \dot{W}_{SWP} + \dot{W}_{PWP} + \dot{W}_{BP} + \dot{W}_{RP}\right) \quad (19)$$

In Eq. (19), $\dot{m}_{72}$ and $\dot{m}_{66}$ represent the mass flow rates of the product water and steam for BH, respectively.

The exergetic performance of the plant is investigated through exergy efficiency, ecological coefficient of performance for cogeneration and exergy destruction factor

$$\eta_{ex} = \frac{\dot{W}_{net} + \dot{X}_p}{\dot{X}_{SG}} \quad (20)$$

$$ECOP_{cog} = \frac{\dot{W}_{net} + \dot{X}_p}{\dot{X}_{dest,tot}} \quad (21)$$

$$f_{ed} = \frac{\dot{X}_{dest,tot}}{\dot{X}_{SG}} \quad (22)$$

where $\dot{X}_{dest,tot}$ is the total exergy destruction rate of the plant, while $\dot{X}_p$ is the exergy of process heat and $\dot{X}_{SG}$ is the exergy of heat transferred through the SG:

$$\dot{X}_p = \dot{X}_i - \dot{X}_{64}, \quad i = 52,53, \dots, \text{or } 63 \quad (23)$$

$$\dot{X}_{SG} = \left(1 - \frac{T_0}{T_{SG}}\right)\dot{Q}_{SG} \quad (24)$$

It should be noted that although η_{ex}, $ECOP_{cog}$ and f_{ed} are exergetic performance indicators, these functions focus on different characteristics of the plant. Also, the objective functions have been normalized as follows in order to more clearly show the influence of the variables on the performance of the plant:

$$\bar{f} = \frac{f}{f_{max}} \quad (25)$$

where $f \equiv \eta_{th}, \epsilon_u, GOR, \eta_{ex}, ECOP_{cog}, f_{ed}$.

4. Results and Discussion

In this section, the comprehensive energy and exergy analysis of the PWR driven MSF desalination plant is presented. Thermal efficiency, utilization factor, gain output ratio, exergy efficiency, ecological coefficient of performance for cogeneration and exergy destruction factor are considered as objective functions to investigate the impacts of reactor thermal power ($\dot{Q}_{SG}$), live steam temperature (T_4), RH mass flow rate ratio (α_{RH}) and temperature (T_{14}), process steam extraction node (i_{ext}), freshwater production capacity ($\dot{m}_{72}$), top brine temperature (T_{77}), MSF unit throttling mass flow rate ratio (α_{TV2}), seawater temperature (T_0) and salinity (Sal_0) on the energetic and exergetic performance of the plant. An in-house code is developed with MATHEMATICA 11 to carry out the calculations, and JANAF data [35] is used for the thermophysical properties of water.

Tables 3 and 4 present the values of the design parameters and dead state properties of seawater when they are not considered as a variable in analyses, and the isentropic efficiencies of the components for PWR's secondary cycle and MSF facility, respectively. The mass flow rate ratio presented in Table 3 is defined as $\alpha_i = \dot{m}_i/\dot{m}_4$ with $\alpha_{RH} = \dot{m}_6/\dot{m}_4$, while the mass flow rate ratio of TV2 is $\alpha_{TV2} = \dot{m}_{70}/\dot{m}_{82}$.

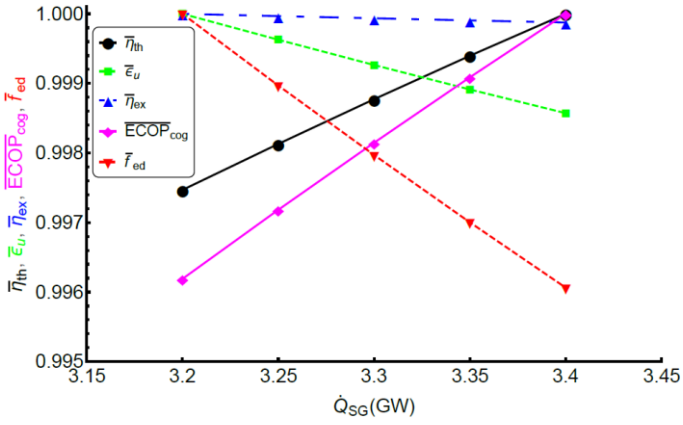
Table 3. Main design parameters of the PWR secondary cycle.

Parameter	Value	Parameter	Value
$\dot{Q}_{SG}$	3300 MW	P_7	2600 kPa
T_{SG}	346.85℃	P_8	1600 kPa
T_4	277℃	P_{14}	1120 kPa
T_{14}	255℃	P_{15}	880 kPa
α_{RH}	0.06	P_{16}	450 kPa
i_{ext}	52	P_{17}	200 kPa
α_7	0.055	P_{18}	90 kPa
α_8	0.05	P_{19}	30 kPa
α_{15}	0.04	P_{22}	200 kPa
α_{16}	0.034	η_{HPT}	0.85
α_{17}	0.034	η_{LPT}	0.83
α_{18}	0.034	η_P	0.85
α_{19}	0.026	η_{GEN}	0.98

Table 4. Main design parameters of the MSF facility.

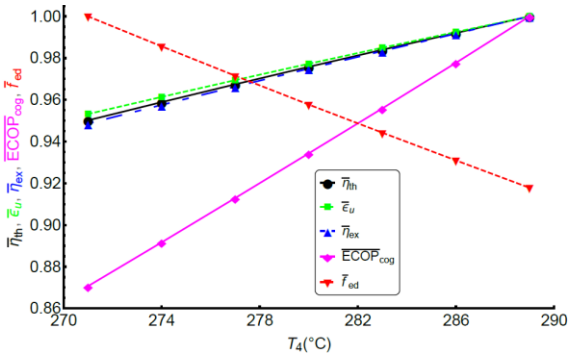
Parameter	Value	Parameter	Value
$\dot{m}_{74}$	3600 kg/s	P_{68}	168 kPa
ΔT_{st}	3℃	P_{71}	10 kPa
TBT	89.85℃	P_{75}	635 kPa
T_{72}	26℃	P_{78}	578 kPa
y	0.005	P_{79}	292 kPa
α_{TV2}	0.9	Sal_0	38000 ppm
T_0	20℃	Sal_{75}	65000 ppm
P_0	101.325 kPa	Sal_{78}	0.01 ppm
η_P	0.85	Sal_{79}	70000 ppm

Figure 2 shows the effect of reactor thermal power and live steam temperature on the objective functions related to the energetic and exergetic performance of the plant. Increasing $\dot{Q}_{SG}$ causes to improve η_{th}, $ECOP_{cog}$ and f_{ed}, and it has a negative impact on ϵ_u while the variation in η_{ex} is negligible. $\dot{Q}_{SG}$ has the largest effect on $ECOP_{cog}$ and f_{ed}. An increase of 0.2 GW changes these functions by 0.4%. A higher live steam temperature tends to increase the performance of the nuclear desalination plant. Performance indicators are more sensitive to the variation in T_4 as compared to the variation in $\dot{Q}_{SG}$. For instance, $ECOP_{cog}$ is enhanced by 13% when T_4 is increased by 18 ℃.



(a)

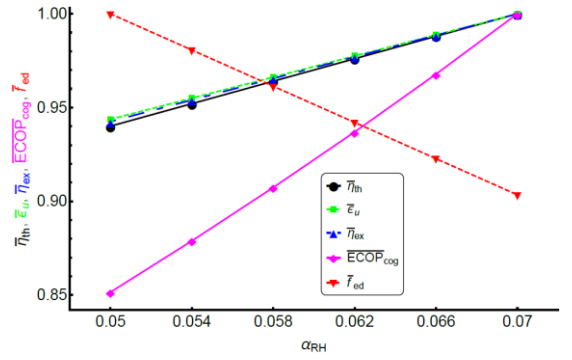
Figure 2. Impact of (a) Q ̇_SG and (b) T_4 on the objective functions



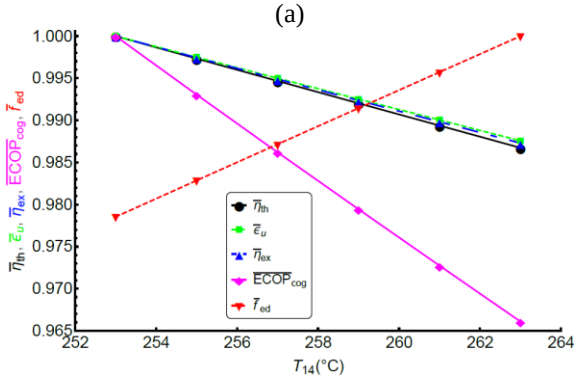
(b)

Figure 2 (Continued). Impact of (a) $\dot{Q}_{SG}$ and (b) T_4 on the objective functions.

The impact of RH mass flow rate ratio and temperature on the energetic and exergetic characteristics of the nuclear desalination plant is illustrated in Figure 3. Higher values of α_{RH} positively affect the performance of the plant. $ECOP_{cog}$ improves by 14.85% as α_{RH} is increased from 0.05 to 0.07. Although the exergy destruction rate of RH increases, FWH1 and FWH2 experience large drops in $\dot{X}_{dest}$ when a higher α_{RH} is chosen, which enhances the plant's performance. The RH temperature, T_{14}, has a relatively less important effect on the objective functions as compared to α_{RH}. Slight decrements are observed in η_{th}, ϵ_u and η_{ex}, and $ECOP_{cog}$ decreases by 3.4% with a 10℃ rise in T_{14}.



(a)



(b)

Figure 3. Variation of performance indicators with (a) α_{RH} and (b) T_{14}.

The effect of desalination capacity on the objective functions of the plant is presented in Figure 4. Increasing $\dot{m}_{72}$ is favorable in terms of ϵ_u, however, it has a negative impact on n_{th}, $ECOP_{cog}$ and f_{ed}. The decrease in thermal efficiency with increasing desalination capacity is a consequence of the drop in electricity output of the cogeneration plant, and such a trend is also shown in [25]. Although an economic analysis is not performed in this study, it should also be noted that increasing the desalination capacity is found to have a positive impact on water cost [28].

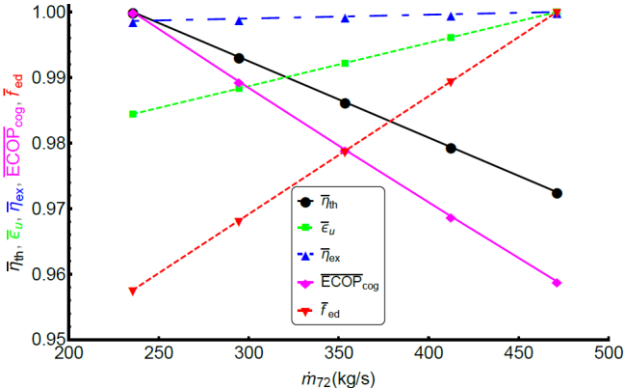


Figure 4. Effect of desalination capacity on the objective functions.

Desalination capacity also affects the plant's performance through the choice of steam extraction node. Tables 5 and 6 demonstrate the variation of the energetic and exergetic performance indicators with steam extraction node for five different desalination capacities, respectively. The desalination capacities given in Tables 5 and 6 correspond to $\dot{m}_{74}$ values of 60, 600, 2400, 3600 and 6000 kg/s. The desalination capacity of $\dot{m}_{72} = 5.9\ kg/s$ corresponds to a small scale plant, 59 kg/s is considered as a large scale plant, and the remaining three capacities are classified as very large scale desalination plants [36]. Results show that the MS outlet (i.e., $i_{ext} = 56$) is the best choice for steam extraction since it gives the optimum values for all objective functions. The selection of extraction node is especially important for very large capacity plants. For instance, based on i_{ext}, η_{th} varies in the range $0.282 - 0.327$ when the desalination capacity is $\dot{m}_{72} = 590\ kg/s$. It should be noted that the choice of steam extraction node might be influenced by economic factors. The exergoeconomic analysis made for SMART driven MSF desalination plant shows that choosing HPT inlet as the steam extraction node results with the least water cost [26].

Table 5. Impact of steam extraction node on energetic objective functions of the nuclear desalination plant for five desalination capacities.

	η_{th}				
i_{ext} \ $\dot{m}_{72}$	$5.9\ kg/s$	$59\ kg/s$	$236\ kg/s$	$353\ kg/s$	$590\ kg/s$
52	0.338030	0.335995	0.329211	0.324688	0.315643
53	0.337697	0.332659	0.315867	0.304673	0.282284
54	0.337725	0.332943	0.317002	0.306375	0.285121
55	0.338129	0.336980	0.333149	0.330596	0.325489
56	0.338146	0.337159	0.333865	0.331670	0.327279
57	0.338105	0.336741	0.332197	0.329167	0.323108
58	0.337715	0.332848	0.316624	0.305807	0.284175
59	0.337795	0.333639	0.319788	0.310554	0.292086
60	0.337817	0.333861	0.320675	0.311884	0.294302
	ϵ_u				
52	0.338392	0.339616	0.343694	0.346413	0.351851
53	0.338059	0.33628	0.33035	0.326398	0.318492
54	0.338087	0.336564	0.331485	0.328100	0.321329
55	0.338491	0.340600	0.347633	0.352321	0.361697
56	0.338509	0.340779	0.348349	0.353395	0.363487
57	0.338467	0.340362	0.346680	0.350892	0.359316
58	0.338078	0.336469	0.331107	0.327532	0.320383
59	0.338157	0.337260	0.334271	0.332279	0.328294
60	0.338179	0.337482	0.335158	0.333609	0.330510

Figure 5 illustrates the variation of the energetic and exergetic characteristics of the nuclear desalination plant with TBT and the MSF unit throttling mass flow rate ratio. Increasing TBT by 35℃ causes a slight decrement in η_{th} and η_{ex} and a more noticeable change of 0.4% in $ECOP_{cog}$ and f_{ed}. However, a higher value of TBT is advantageous in terms of GOR, since this objective function has improved significantly by 34% with a 35℃ increment in TBT. The TBT-GOR relationship presented in Figure 5a is also observed in [15]. The trends shown in Figure 5b reveal that α_{TV2} has a negligible effect on plant's performance. The BH temperature difference (ΔT_{BH}) also affects the GOR of the plant as shown in Figure 6. When ΔT_{BH} rises from 4 to 8, GOR of the facility decreases slightly, and the rate of decrement becomes faster as TBT increases.

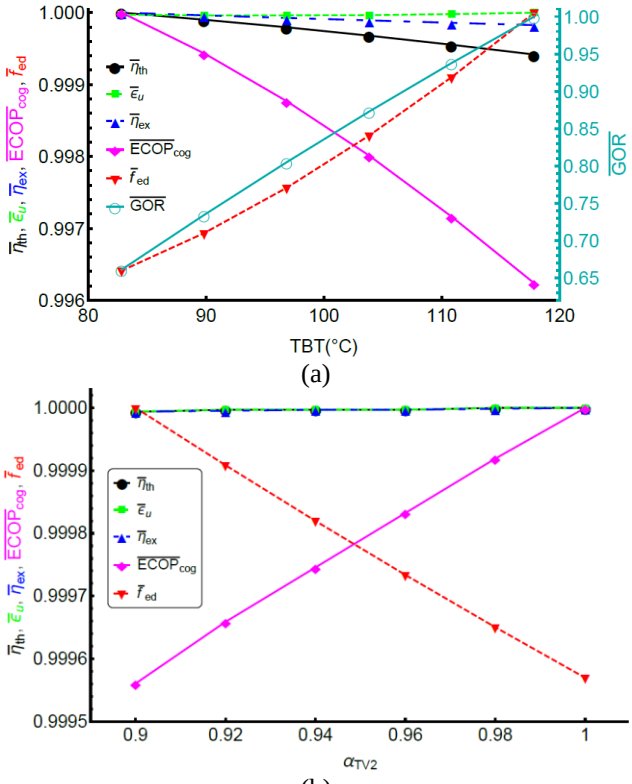


Figure 5. Effect of (a) TBT and (b) TV2 mass flow ratio on the objective functions.

Table 6. Impact of steam extraction node on exergetic objective functions of the nuclear desalination plant for five desalination capacities.

			η_{ex}		
i_{ext} \ $\dot{m}_{72}$	$5.9\ kg/s$	$59\ kg/s$	$236\ kg/s$	$353\ kg/s$	$590\ kg/s$
52	0.641658	0.641856	0.642516	0.642955	0.643834
53	0.640937	0.634639	0.613648	0.599655	0.571667
54	0.640951	0.634781	0.614216	0.600505	0.573085
55	0.641691	0.642178	0.643803	0.644886	0.647052
56	0.641718	0.642454	0.644906	0.646541	0.649810
57	0.641685	0.642123	0.643584	0.644558	0.646506
58	0.640917	0.634441	0.612857	0.598467	0.569687
59	0.640994	0.635210	0.615931	0.603078	0.577373
60	0.640992	0.635194	0.615868	0.602983	0.577214
			$ECOP_{cog}$		
52	1.84323	1.82479	1.76596	1.72887	1.65932
53	1.83784	1.77265	1.57779	1.46422	1.26825
54	1.83816	1.77569	1.58796	1.47785	1.28662
55	1.84430	1.83536	1.80627	1.78746	1.75117
56	1.84456	1.83792	1.81623	1.80214	1.77481
57	1.84405	1.83286	1.79663	1.77335	1.72875
58	1.83797	1.77386	1.58154	1.46898	1.27391
59	1.83898	1.78342	1.61426	1.51346	1.33552
60	1.83919	1.78547	1.62118	1.52278	1.34816
			f_{ed}		
52	0.348117	0.351743	0.363833	0.371892	0.388011
53	0.348744	0.358018	0.388930	0.409538	0.450754
54	0.348691	0.357484	0.386795	0.406336	0.445418
55	0.347932	0.349892	0.356427	0.360783	0.369496
56	0.347898	0.349555	0.355080	0.358763	0.366130
57	0.347976	0.350340	0.358218	0.363470	0.373974
58	0.348709	0.357662	0.387507	0.407404	0.447197
59	0.348560	0.356174	0.381556	0.398477	0.432319
60	0.348518	0.355757	0.379889	0.395976	0.428151

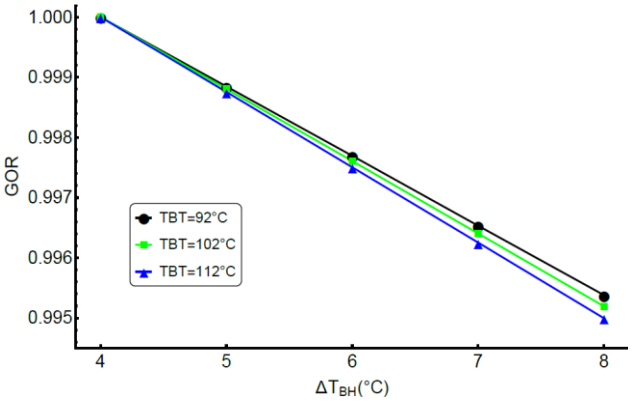


Figure 6. Effect of ΔT_{BH} on GOR of the nuclear desalination plant.

The variation of nuclear desalination plant performance with the seawater temperature and salinity is shown in Figure 7. The temperature and salinity ranges are chosen based on the data of the Mediterranean Sea [37]. The rise in T_0 from 15 ℃ to 27 ℃ has a positive effect on plant's exergetic performance, and the most significant change is observed for $ECOP_{cog}$. On the other hand, an increase in T_0 decreases η_{th} and ϵ_u of the plant relatively by 0.89% and 0.84%, respectively, and such energetic performance degradation is typical for power plants [38]. As for the impact of the seawater salinity, $ECOP_{cog}$ and f_{ed} are found to be the most sensitive objective functions. An increase of 5000 ppm in Sal leads to a relative decrement of 0.13%.

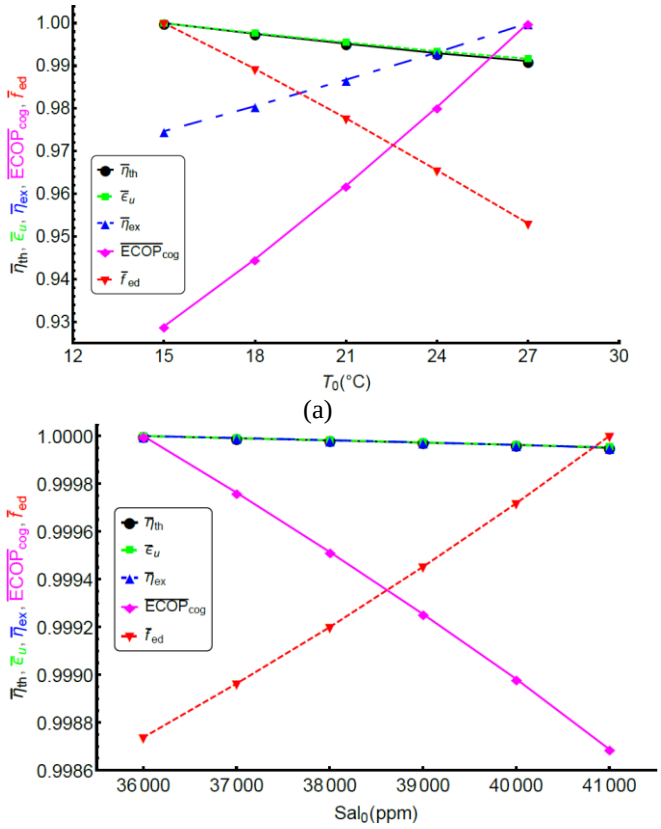


Figure 7. Impact of seawater (a) temperature and (b) salinity on the objective functions.

Based on the results presented in Figures 2-6 together with Tables 5 and 6, an optimum nuclear desalination plant configuration is built to investigate the influence of design decisions on the performance of the nuclear desalination plant by setting $\dot{Q}_{SG} = 3.4\,GW$, $T_4 = 289$℃, $\alpha_{RH} = 0.07$, $T_{14} = 253$℃, $i_{ext} = 56$ and $TBT = 82.85$℃ to produce freshwater with a capacity of $\dot{m}_{72} = 353\,kg/s$ from the seawater having a temperature of $T_0 = 20$℃ and a salinity of $38000ppm$. Thermodynamic properties of all states are presented in Table 7 and 8 for PWR's secondary cycle and desalination facility, respectively. The negative exergy flow rates seen in Table 8 are due to the higher salinity levels of the corresponding states, and such negative values were also reported in [19].

The energetic and exergetic performance indicators of the optimum nuclear desalination plant configuration are presented in Table 9. The performance of a base scenario, for which the results are obtained with the parameters given in Table 3, are also provided in Table 9. Comparison of the results reveal the importance of the parametric energy-exergy analysis of the nuclear desalination plant. All performance indicators except GOR have improved significantly with the alteration of the decision variables. η_{th}, ϵ_u, η_{ex} and $ECOP_{cog}$ increases by 3.01%, 3.15%, 4.70% and 0.46, respectively, and f_{ed} decreases by 5.7% as a result of the alteration of design parameters. The variations correspond to relative improvements of 9.27%, 8.52%, 7.31%, 26.63% and 15.27% in η_{th}, ϵ_u, η_{ex}, $ECOP_{cog}$ and f_{ed}, respectively These improvements are significant. The findings can also be interpreted from an economic perspective. For instance, the price of electricity is $0.1235\,\$/kWh$ for the Akkuyu nuclear power plant that is under construction in Mersin, Turkey for which a single reactor unit has a nominal power output of $1200MW$. A 3.01% increase in thermal efficiency can be considered as a $36MW$ improvement in the power output. Nuclear power plants have an average capacity factor of 90%. Therefore, the 3.01% increase in thermal efficiency would cause a 35.05 million $ yearly improvement in the revenue of the plant.

Table 7. Thermodynamic properties of all states for the secondary cycle of the PWR.

State	T(℃)	$P(MPa)$	$\dot{m}(kg/s)$	$h(kJ/kg)$	$s(kJ/kg\cdot K)$	$\dot{X}(MW)$
4	289.00	7.334	1899.62	2768.17	5.7910	2039.31
5	289.00	7.334	1766.64	2768.17	5.7910	1896.58
6	289.00	7.334	132.97	2768.17	5.7910	142.75
7	226.05	2.600	104.48	2606.30	5.8482	93.50
8	201.37	1.600	94.98	2535.45	5.8775	77.45
9	184.86	1.120	1567.18	2485.54	5.8999	1189.47
10	184.86	1.120	1567.18	2485.54	5.8999	1189.47
11	184.86	1.120	1335.00	2781.36	6.5457	1155.41
12	184.86	1.120	1299.15	2781.36	6.5457	1124.38
13	253.00	1.120	1299.15	2945.36	6.8804	1209.98
14	253.00	1.120	1299.15	2945.36	6.8804	1209.98
15	228.52	0.880	75.98	2899.96	6.8991	66.90
16	147.90	0.450	64.59	2706.82	6.7692	46.85

Table 7 (Continued). Thermodynamic properties of all states for the secondary cycle of the PWR.

State	$T(℃)$	$P(MPa)$	$\dot{m}(kg/s)$	$h(kJ/kg)$	$s(kJ/kg \cdot K)$	$\dot{X}(MW)$
17	120.21	0.200	64.59	2666.43	7.0257	39.39
18	96.73	0.090	64.59	2559.37	7.0942	31.18
19	69.09	0.030	49.39	2427.14	7.1905	15.91
20	30.03	0.004	980.01	2226.51	7.3661	68.74
21	20.00	0.101	99559.60	83.91	0.2965	0.00
22	20.03	0.200	99559.60	84.03	0.2968	0.60
23	25.03	0.101	99559.60	104.95	0.3676	17.97
24	30.03	0.004	1015.87	125.86	0.4372	0.70
25	30.03	0.030	1015.87	125.89	0.4373	0.70
26	55.55	0.030	1065.26	232.58	0.7751	8.91
27	55.57	0.090	1065.26	232.65	0.7753	8.92
28	87.29	0.090	1129.85	365.66	1.1614	31.84
29	87.32	0.200	1129.85	365.79	1.1618	31.86
30	116.79	0.200	1194.44	490.20	1.4932	66.24
31	116.87	0.450	1194.44	490.51	1.4940	66.33
32	143.50	0.450	1259.03	604.20	1.7754	109.20
33	143.63	0.880	1259.03	604.75	1.7767	109.41
34	184.86	0.880	232.18	784.58	2.1862	34.06
35	266.20	7.334	132.97	1165.85	2.9405	40.80
36	215.64	0.880	365.15	923.42	2.4983	70.86
37	182.87	0.880	1700.16	775.77	2.1670	243.98
38	183.09	1.600	1700.16	776.73	2.1691	244.57
39	203.88	1.600	1795.14	869.78	2.3670	321.11
40	204.18	2.600	1795.14	871.15	2.3699	322.07
41	226.05	2.600	1899.62	966.59	2.5736	408.68
42	227.48	7.334	1899.62	978.33	2.5868	423.63
43	289.00	7.334	1899.62	2768.17	5.7910	2039.31
44	226.05	2.600	104.48	2606.30	5.8482	93.50
45	201.37	1.600	94.98	2535.45	5.8775	77.45
46	228.52	0.880	75.98	2899.96	6.8991	66.90
47	147.90	0.450	64.59	2706.82	6.7692	46.85
48	120.21	0.200	64.59	2666.43	7.0257	39.39
49	96.73	0.090	64.59	2559.37	7.0942	31.18
50	69.09	0.030	49.39	2427.14	7.1905	15.91
51	30.03	0.004	980.01	2226.51	7.3661	68.74
52	289.00	7.334	0.00	2768.17	5.7910	0.00
53	226.05	2.600	0.00	2606.30	5.8482	0.00
54	201.37	1.600	0.00	2535.45	5.8775	0.00
55	184.86	1.120	0.00	2485.54	5.8999	0.00
56	184.86	1.120	35.85	2781.36	6.5457	31.03
57	253.00	1.120	0.00	2945.36	6.8804	0.00
58	228.52	0.880	0.00	2899.96	6.8991	0.00
59	147.90	0.450	0.00	2706.82	6.7692	0.00
60	120.21	0.200	0.00	2666.43	7.0257	0.00
61	96.73	0.090	0.00	2559.37	7.0942	0.00
62	69.09	0.030	0.00	2427.14	7.1905	0.00
63	30.03	0.004	0.00	2226.51	7.3661	0.00
64	184.86	1.120	35.85	784.58	2.1862	5.26
65	30.02	0.004	35.85	784.58	2.6010	0.81
66	88.85	0.067	31.33	2657.73	1.1796	472.53
67	88.85	0.067	31.33	372.20	0.3811	37.80

Table 8. Thermodynamic properties of all states for the MSF facility.

State	$T(℃)$	$P(MPa)$	$Sal(ppm)$	$\dot{m}(kg/s)$	$h(kJ/kg)$	$s(kJ/kg \cdot K)$	$\dot{X}(MW)$
68	20.00	0.168	38000	1455.88	81.99	0.3183	0.09
69	28.00	0.118	38000	1455.88	114.38	0.4275	0.66
70	28.00	0.118	38000	689.63	114.38	0.4275	0.31
71	28.00	0.010	38000	689.63	114.38	0.4275	0.30
72	26.00	0.010	0.01	315.26	109.02	0.3812	2.72
73	28.00	0.010	70000	374.37	111.68	0.4366	-1.85
74	28.00	0.010	65000	3600.00	112.09	0.4354	-15.00
75	28.00	0.635	65000	3600.00	112.62	0.4352	-12.89
76	77.85	0.635	65000	3600.00	310.29	1.0426	57.74
77	82.85	0.635	65000	3600.00	330.18	1.0988	69.96
78	26.00	0.578	0.01	315.26	109.54	0.3811	2.90
79	28.00	0.292	70000	374.37	111.92	0.4366	-1.75
80	20.00	0.101	70000	374.37	80.18	0.3303	-1.97
81	20.00	0.101	0.01	315.26	84.01	0.2965	2.67
82	20.00	0.101	38000	766.25	81.93	0.3183	0.00

Table 9. Comparison of an optimum configuration with a base configuration for the nuclear desalination plant.

Performance indicator	Base scenario	Optimum scenario
η_{th}	0.3247	0.3548
ϵ_u	0.3464	0.3759
GOR	11.1781	10.0639
η_{ex}	0.6429	0.6899
$ECOP_{cog}$	1.7289	2.1893
f_{ed}	0.3719	0.3151

The irreversibility rates of plant components are presented in Figs. 8 and 9 for PWR's secondary cycle and MSF facility, respectively. Secondary cycle of the PWR has a total irreversibility rate of $551.75 MW$ and SG has the largest contribution of 32%. The total exergy destruction rate of the MSF facility, $13.10\ MW$ is low as compared to the secondary cycle and the MSF unit is responsible for the majority of the losses. Results of [19] also show that MSF unit is the largest source of irreversibility in the desalination facility.

5. Conclusion

Access to freshwater is a challenging issue for regions with arid climate. Desalination is an essential technology for countries having water scarcity problems and nuclear energy is a promising source that can provide both thermal and electrical energy for desalination facilities with a negligible environmental impact. This study focuses on a comprehensive energy and exergy analysis of a nuclear desalination plant. A pressurized water reactor is used as the energy source to produce electricity and to supply process heat to a multi-stage flash desalination facility for seawater desalination. The effects of several design and operating parameters on the energetic and exergetic characteristics of the plant are studied in detail. The results of the analysis can be summarized as follows:

- Higher reactor thermal power has a positive impact on thermal efficiency, coefficient of ecological performance and exergy destruction factor of the plant, while it tends to decrease utilization factor.
- The overall plant performance improves as the live steam temperature and the reheater mass flow rate ratio increase, while a lower reheater temperature is preferable for better cogeneration performance.
- Increasing the freshwater production capacity improves the utilization factor of the nuclear desalination plant, on the other hand, it has a negative impact on thermal efficiency, coefficient of ecological performance and exergy destruction factor.
- The outlet of the moisture separator is the best candidate for steam extraction, and the selection of

extraction node is crucial for very large scale desalination plants.

- Gain output ratio of the plant enhances remarkably with increasing top brine temperature. Decreasing the brine heater temperature difference also has a positive impact on the gain output ratio, and the throttling valve mass flow rate ratio has a negligible effect on plant's performance.
- A higher seawater temperature tends to decrease the energetic performance of the nuclear desalination plant, however, it positively affects the exergetic objective functions. A higher seawater salinity causes a slight decrement in the exergetic performance.
- The overall plant performance can be significantly improved by carefully selecting the design parameters.
- Steam generator and multi-stage flash unit cause the largest exergy destruction rates in the secondary cycle of the reactor and multi-stage flash facility, respectively.

The findings of this study provide useful technical information for both the designers of PWR based nuclear desalination systems and policy makers, especially in countries suffering from water scarcity. With that being said, the detailed analysis presented in this paper can be further improved by considering economic factors and carrying out a multiobjective optimization that takes into account energetic, exergetic and economic objective functions. Research into performing this optimization is already in progress.

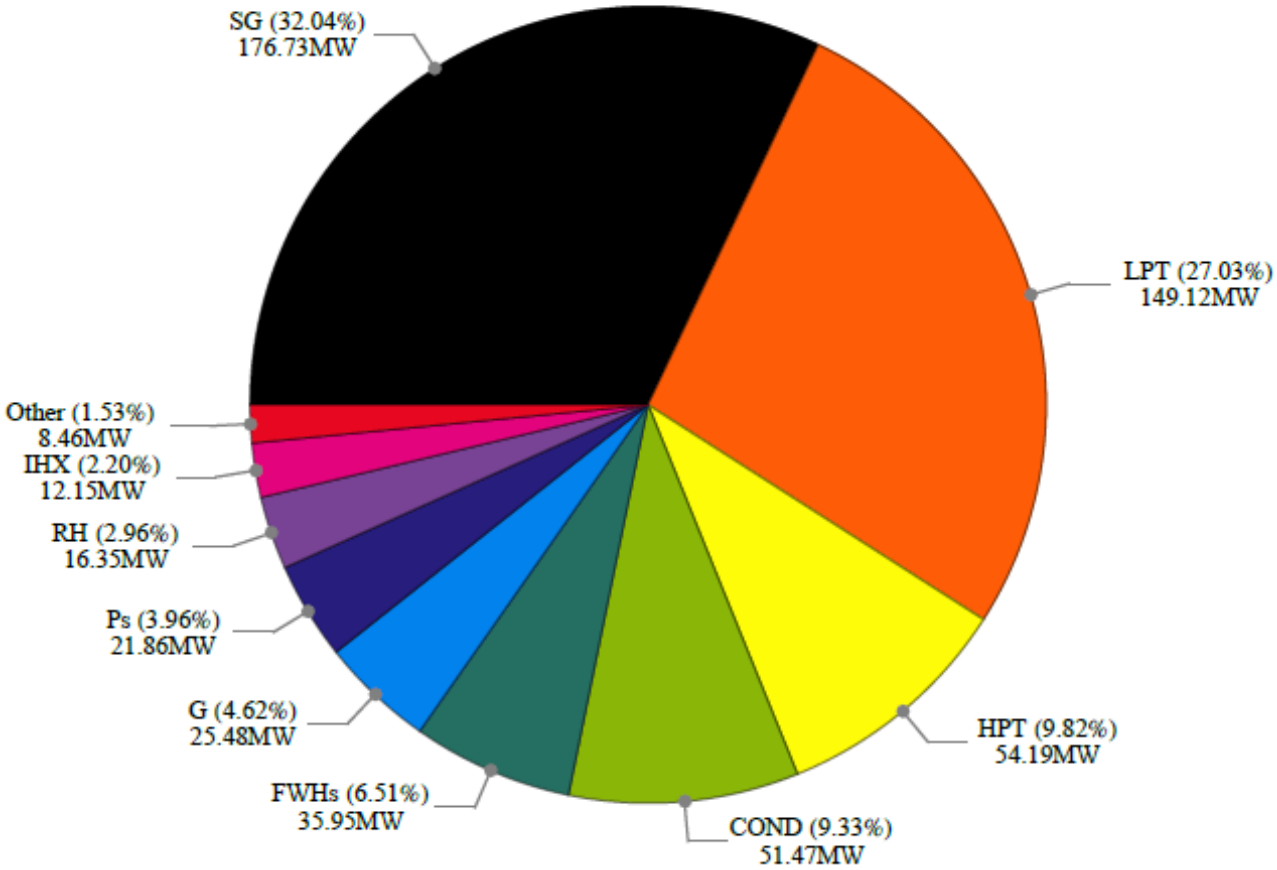


Figure 8. Exergy destruction rates of PWR's secondary cycle components.

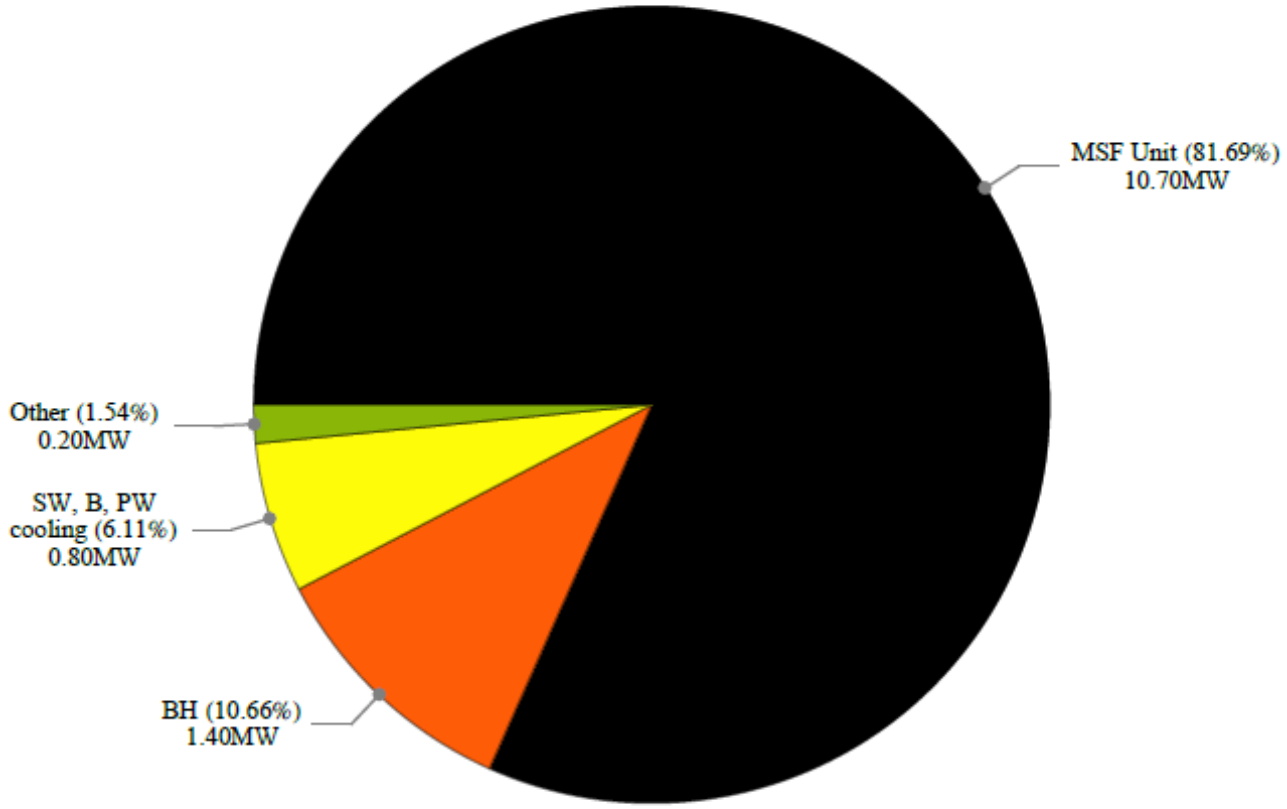


Figure 9. Exergy destruction rates of MSF facility components.

Contributions of The Authors

Akyürek: Literature review, Formal analysis, Validation, Writing – original draft
Tanbay: Conceptualization, Formal analysis, Methodology, Software development, Supervision, Visualization, Writing – review and editing

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] IAEA, "Introduction of Nuclear Desalination," International Atomic Energy Agency, Vienna, Austria, Technical Reports Series no. 400, 2000.

[2] IAEA, "Status of Nuclear Desalination in IAEA Member States," International Atomic Energy Agency, Vienna, Austria, IAEA-TECDOC-1524, 2007.

[3] FAO and UN Water, "Progress on Level of Water Stress. Global status and acceleration needs for SDG Indicator 6.4.2," United Nations, Rome, Italy, 2021.

[4] M. A. Rosen, "Nuclear Energy: Non-Electric Applications," *Eur. J. Sust. Dev. Res.*, vol. 5, no. 1, 2021, Art. no. em0147.

[5] EIA, "Electric Power Monthly," EIA. Accessed: Feb. 5, 2024. [Online]. Available: https://www.eia.gov/electricity/monthly/epm_table_grapher.php?t=epmt_6_07_b

[6] A. Panagopoulos, and K. J. Haralambous, "Environmental impacts of desalination and brine treatment - Challenges and mitigation measures," *Mar. Pollut. Bull.*, 2020, vol. 161, Dec. 2020, Art. no. 111773.

[7] R. S. El-Emam, H. Ozcan, R. Bhattacharyya, and L. Awerbuch, "Nuclear desalination: A sustainable route to water security," *Desalination*, vol. 542, Nov. 2022, Art. no. 116082.

[8] M. Ayaz, M. A. Namazi, M. Ammad ud Din, M. I. M. Ershath, A. Mansour, and E. M. Aggoune, "Sustainable seawater desalination: Current status, environmental implications and future expectations," *Desalination*, vol. 540, Oct. 2022, Art. no. 116022.

[9] H. Nassrullah, S. F. Anisa, R. Hashaikeh, and N. Hilal, "Energy for desalination: A state-of-the-art review," *Desalination*, vol. 491, Oct. 2020, Art. no. 114569.

[10] D. Curto, V. Franzitta, and A. Guercio, "A review of the water desalination technologies," *Appl. Sci.*, vol. 11, no. 2, Jan. 2021, Art. no. 670.

[11] I. G. Sánchez-Cervera, K. C. Kavvadias, and I. Khamis, "DE-TOP: A new IAEA tool for the thermodynamic evaluation of nuclear desalination," *Desalination*, vol. 321, pp. 103-109, Jul. 2013.

[12] K. C. Kavvadias, and I. Khamis, "The IAEA DEEP desalination economic model: A critical review," *Desalination*, vol. 257, pp. 150-157, Jul. 2010.

[13] R. S. Faibish, and H. Ettouney, "MSF nuclear desalination," *Desalination*, vol. 157, pp. 277-287, Aug. 2003.

[14] S. Ghurbal, and M. Ashour, "Economic competitiveness of nuclear desalination in Libya," *Desalination*, vol. 158, pp. 201-204, Aug. 2003.

[15] D. T. Ingersoll, J. L. Binder, D. Conti, and M. E. Ricotti, "Nuclear desalination option for the international reactor innovative and secure (IRIS) design," presented at 5th International Conference on Nuclear Option in Countries with Small and Medium Electricity Grids, Dubrovnik, Croatia, May 16-20, 2004.

[16] A. K. Adak, and P. K. Tewari, "Coupling aspects of an MSF desalination plant and loss of electrical power generation of a nuclear power plant: case study," *Int. J. Nucl. Desalin.*, vol. 1, no. 3, pp. 373-381, Oct. 2004.

[17] A. K. Adak, V. K. Srivastava, and P. K. Tewari, "Thermal coupling system analysis of a nuclear desalination plant," *Int. J. Nucl. Desalin.*, vol. 4, no. 2, pp. 123-133, Sep. 2010.

[18] L. Tian, J. Guo, Y. Tang, and L. Cao, "A historical opportunity: economic competitiveness of seawater desalination project between nuclear and fossil fuel while the world oil price over $50 per boe-part A: MSF," *Desalination*, vol. 183, pp. 317-325, Nov. 2005.

[19] N. Kahraman and Y. A. Cengel, "Exergy analysis of a MSF distillation plant," *Energ. Convers. Manage*, vol. 46, no. 15-16, pp. 2625-2636, Sep. 2005.

[20] M. S. Saadawy, "Optimum thermal coupling system for co-generation nuclear desalination plants," *Int. J. Nucl. Desalin.*, vol. 2, no. 1, pp. 22-43, Apr. 2006.

[21] M. H. K. Manesh, and M. Amidpour, "Multi-objective thermoeconomic optimization of coupling MSF desalination with PWR nuclear power plant through evolutionary algorithms," *Desalination*, vol. 249, no. 3, pp. 1332-1344, Dec. 2009.

[22] M. H. K. Manesh, M. Amidpour, and M. H. Hamedi, "Optimization of the coupling of pressurized water nuclear reactors and multistage flash desalination plant by evolutionary algorithms and thermoeconomic method," *Int. J. Energ. Res.*, vol. 33, no. 1, pp. 77-99, 2009.

[23] G. Alonso, S. Vargas, E. del Valle, and R. Ramirez, "Alternatives of seawater desalination using nuclear power," *Nucl. Eng. Des.*, vol. 245, pp. 39-48, Apr. 2012.

[24] X. Yan, H. Noguchi, H. Sato, Y. Tachibana, K. Kunitomi, and R. Hino, "Study of an incrementally loaded multistage flash desalination system for optimum use of sensible waste heat from nuclear power plant," *Int. J. Energ. Res.*, vol. 37, no. 14, pp. 1811-1820, 2013.

[25] D. T. Ingersoll, Z. J. Houghton, R. Bromm, and C. Desportes, "NuScale small modular reactor for Co-generation of electricity and water," *Desalination*, vol. 340, pp. 84-93, May 2014.

[26] E. Priego, G. Alonso, E. del Valle, and R. Ramirez, "Alternatives of steam extraction for desalination purposes using SMART reactor," *Desalination*, vol. 413, pp. 199-216, Jul. 2017.

[27] S. U. D. Khan, S. U. D. Khan, S. Haider, A. El-Leathy, U. A. Rana, S. N. Danish and R. Ullah, "Development and techno-economic analysis of small modular nuclear reactor and desalination system across Middle East and North Africa region," *Desalination*, vol. 406, pp. 51-59, Mar. 2017.

[28] S. U. D. Khan, and S. U. D. Khan, "Karachi Nuclear Power Plant (KANUPP): As case study for techno-economic assessment of nuclear power coupled with water desalination," *Energy*, vol. 127, pp. 372-380, May 2017.

[29] M. F. Polat, and I. Dincer, "Comparative evaluation of possible desalination options for Akkuyu Nuclear Power Plant," in *Exergetic, Energetic and Environmental Dimensions*, I. Dincer, C. O. Colpan, and O. Kizilkan, Eds., Academic Press, 2018, pp. 583-596.

[30] E. Dewita, T. Ariyanto, H. Susiati, and M. Pancoko, "Conceptual design of Indonesia experimental power reactor coupled with desalination unit," *J. Phys. Conf. Ser.*, vol. 1198, no. 2, 2019, Art. no. 022056.

[31] K. Sadeghi, S. H. Ghazaie, E. Sokolova, E. Fedorovich, and A. Shirani, "Comprehensive techno-economic analysis of integrated nuclear power plant equipped with various hybrid desalination systems," *Desalination*, vol. 493, Nov. 2020, Art. no. 114623.

[32] E. K. Redfoot, M. G. McKellar, and R. A. Borrelli, "Allocating heat and electricity in an Integrated Energy System coupled with a water purification system," *Nucl. Eng. Des.*, vol. 397, Oct. 2022, Art. no. 111902.

[33] A. Al Ghamdi, and I. Mustafa, "Exergy analysis of a MSF desalination plant in Yanbu, Saudi Arabia," *Desalination*, vol. 399, pp. 148-158, Dec. 2016.

[34] B. Najafi, A. Shirazi, M. Aminyavari, F. Rinaldi, and R. A. Taylor, "Exergetic, economic and environmental analyses and multi-objective optimization of an SOFC-gas turbine hybrid cycle coupled with an MSF desalination system," *Desalination*, vol. 334, no. 1, pp. 46-59, Feb. 2014.

[35] M. W. Chase, "NIST-JANAF thermochemical tables 2 volume set," in *Journal of physical and chemical reference data monographs*. College Park, MD: American Institute of Physics, 1998.

[36] A. N. Bdour, N. Al-Sadeq, M. Gharaibeh, A. Mendoza-Sammet, M. D. Kennedy, and S. G. Salinas-Rodriguez, "Techno-economic analysis of selected PV-BWRO desalination plants in the context of the water-energy nexus for low-medium-income countries," *Energies*, vol. 15, no. 22, Nov. 2022, Art. no. 8657.

[37] S. C. Painter, and M. N. Tsimplis, “Temperature and salinity trends in the upper waters of the Mediterranean Sea as determined from the MEDATLAS dataset,” *Cont. Shelf Res.*, vol. 23, no. 16, pp. 1507-1522, Oct. 2003.

[38] A. Durmayaz, and O. S. Sogut, “Influence of cooling water temperature on the efficiency of a pressurized-water reactor nuclear-power plant,” *Int. J. Energ. Res.*, vol. 30, no. 10, pp. 799-810, Apr. 2006.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1067-1082 YEAR: 2024
DOI: 10.17798/bitlisfen.1500558

Optic Character Recognition Using Image Processing Techniques

Halit ÇETİNER[1*], Bayram CETİŞLİ[2]

[1]*Isparta University of Applied Sciences, Vocational School of Technical Sciences,Isparta, Türkiye*
[2]*Süleyman Demirel University, Electrical Electronics Engineer, Isparta, Türkiye*
(ORCID: 0000-0001-7794-2555) (ORCID: 0000-0002-2600-3977)

Keywords: Digit Recognition, Wavelet Transform, OCR, ANN, CNN, Deep Learning.

Abstract

In pattern recognition, the automatic identification of image-based input patterns by machine is an important problem. The need to convert images taken with tools such as cameras into formats usable by computers is increasing day by day. The output and efficiency of manual data entry processes are low and error rates are high. Outsourcing these processes to companies dealing with professional information entry is not preferred due to reasons such as security and lack of continuous service quality. For these and similar reasons, a system that enables automatic recognition of optical characters is aimed. With the help of this system, it is aimed to provide the opportunity to serve more people in a shorter time. In accordance with the stated objectives, a unique dataset has been created by performing identification card segmentation. This dataset was combined with the standard OCR dataset and classification was performed with ANN and proposed CNN methods on the extended dataset. The proposed CNN method, inspired by the Inception V3 model, is a deep learning model consisting of 15 layers. The ANN model uses features obtained from different wavelet types to increase discrimination. Competitive results were obtained from both ANN and proposed CNN models. In the proposed CNN model, in the extended version of the dataset, 99.49%, 98.87%, 99.48% and 99.50% values for F1 score, recall, precision, and accuracy metrics were obtained in the same order for training. Similarly, for validation, 99.13%, 98.43%, 99.21% and 99.27% values were obtained in the same order.

1. Introduction

Optical Character Recognition (OCR) poses a serious challenge in the field of computer vision as it plays an important role in the process of converting unstructured raw data into meaningful data. In order to overcome the challenges in OCR and to support ANN and CNN based artificial intelligence applications, algorithms that include text detection and text recognition steps in the OCR process are needed [1]. Text detection involves the identification and extraction of regions with meaningful data obtained using segmentation techniques. Text recognition involves the classification of the values in the extracted text regions with high performance criteria. In recent years, unlike machine learning based OCR approaches, deep learning based OCR approaches have made significant progress. These models are reported to offer high OCR performance even for images with complex backgrounds [1].

In OCR problems, there are a wide variety of scenarios for the detection of the text region and the recognition of meaningful characters in that region. Images, videos, scene images are among the materials used for OCR process. OCR is used in many application scenarios such as scene recognition, license plate recognition, machine translation, number recognition [1]. In the OCR process, it is aimed to provide automatic detection and recognition of text regions from video images consisting of an image or images.

*Corresponding author: halitcetiner@isparta.edu.tr
Received: 13.06.2024, Accepted: 21.10.2024

Today, electronic government systems (EGS) have been spread too much. One of the most important reasons for the widespread offers great advantages. The biggest advantage of the EGS is provided with easily access to information about that person with identity. In addition to the person can obtain the information from institutions and organizations such as information services, integrated electronic services, payment transactions, short cuts to institutions and organizations, date information and announcements, message from the institutions to the citizens. This paper proposes a digit recognition system that can be integrated into systems such as EGS in the future. In this sense, both machine learning and deep learning based methods are tested for digit recognition and classification.

While the institutions employees logging to the EGS with identification card (IC), they are not entering as accurate and fast as systems of all the numbers on identity because of reasons such as eye strain, distractions. The ICRS to facilitate the use of EGS are realized. The ICRS is an optical character recognition system. Character recognition has found a wide range of applications in different fields such as recognition of card numbers belonging to various banks, directing of vehicles by reading road signs, advertising, banner, reading of market boards, automatic reading of bank checks, license plate recognition, reading of postal codes in the last century. More sophisticated methods are needed in order to meet the increasing request in this is a different application area. There are many different works related to the mentioned OCR based applications. Authors in the [2] presented number recognition system. The system uses pre-processing methods which are based on processes such thresholding, noise extraction. Then, each image of all images resized a standard size by the system. Finally, end points of numbers in thinned image are obtained by using chain code. Thus, they developed number recognition system. Authors in the [3] presented number recognition system. They use likelihood values between graphs which are created feature vectors. Then, they use KNN algorithm so as to classification of oriya numbers. Authors in the [4] presented license plate recognition system. The system consists of detection, respectively shredding according to character, recognition. Author in the [5] presented plate recognition system which uses thresholding, morphological operations, template matching. Authors in the [6] presented number recognition system which consists of segmentation and recognition. The system is a mobile application based on android operating system. Authors in the [7] presented plate recognition system which consists of thresholding, localization of plate with genetic algorithm. Authors in the [8] presented printed text recognition system which uses training with Hidden Markov model by using multiple fonts.

Authors in the [9] presented active learning methodology based on KNN. The methodology is used in the recognition of different script numbers. Author in the [10] presented character recognition system which consists of forms filled by writers, convert forms into images, creating of train and test sets, pre-processing with Otsu thresholding, segmentation with connected component analysis, feature extraction with framing feature, classification with KNN and SVM classifiers. The society began to development with information. It is difficult to control the information with increasing level of information. The circulation rate of the information is much increased according as development of computer and communications technology. This rapid growth makes necessary to make an efficient use of public and private institutions. The effective usage which can only be solved to make use of computer and internet technology. Computer and communications technology play key role in the data collection, storage, interpretation and the process of reaching the users. Advances in information technology has become facilitated access to information. The use of this information technology in public administration is suitable accordance with EGS for all private and public institutions. Thus, it will become faster and quality of public and private services. A character recognition system has been developed in Turkey due to the convenience of EGS in many different areas such as mobile electronic system integration, traffic information system, mobile information, earthquake emergency response network, lone pension operation, national judicial network. It is aimed to facilitate access to the system at a later point in the process, including facilitating the work. In Turkey, the verification process takes a lot of time as the employees of the organisation log in to the EGS with IC. In addition to this, institutional employees find the existence of individuals from EGS databases by manually logging in. During these processes, negative situations or delays may occur as a result of incorrect entries. For these reasons, ICRS has been developed.

As in the EGS system in Turkey, there are also applications in the world that automate the OCR process. In the United Kingdom, medical reports created by doctors are stored in the national health service database. Important information of these documents is identified and added to the patient's database. In this and similar ways, it is seen that studies have been carried out in the literature to

automate the manual process by reducing the cognitive load [11]. The OCR helps to digitize printed materials such as valuable information, forms, etc. that are present in printed documents.

In line with the stated aims and objectives, the main contributions of this article to the literature are listed below.

- In this paper, a new segmentation algorithm is developed that automatically detects the digits in TC IC cards and separates the digits from each other.
- A new digit dataset, which is not defined in the literature, was created using the newly proposed segmentation algorithm.
- The new dataset was merged with the publicly available standard OCR dataset to extend the dataset.
- A new deep learning-based approach is presented for the OCR process, which consists of two main components: text recognition and text detection.
- Wavelet-based features were extracted on the extended dataset (Dataset-IV) and ANN classification was performed. In the Dataset-IV, the highest values achieved with ANN were obtained using Rbio 3.1 and Bior 2.2 wavelet types. These values were found to be 100% and 99.8% accuracy for training and testing, respectively. As can be seen from this result, the performance of the Rbio 3.1 wavelet type has also improved as a result of the Dataset-IV.
- At the same time, the proposed CNN method was used to classify the characters on the Dataset-IV. In the proposed CNN model, 99.24 %, 98.96 %, 99.77%, 99.77%, and 99.24% values were obtained for F1 score, recall, precision, and accuracy metrics in training for the Dataset-I, respectively. For validation, 99.12%, 98.50%, 99.43%, and 99.02% values were obtained in the same order. In the Dataset-IV, 99.49%, 98.87%, 99.48%, and 99.50% values were obtained in the same order for training. Similarly, for validation, values of 99.13%, 98.43%, 99.21%, and 99.27% were obtained in the same order.

The following sections of the paper are organized as follows. Section 2 provides a detailed summary of machine learning and deep learning based OCR approaches in the literature. Section 3 provides information about the datasets used in the paper and the algorithms that process these datasets. Section 4 presents the performance results obtained from the ANN and proposed CNN methods used in the article. In the last section, the article is concluded.

2. Related Works

There are researches on the number recognition application areas for many languages such as Arabic, Chinese, French, German, British, Indian, Italian, Persian [12]–[19]. Unfortunately, for number recognition of TR's identity card numbers was not found in any work in the literature. A study was carried out to improve the performance of the EGS system in similar structures and to overcome this deficiency in the literature.

Extracting text from images requires overcoming many challenges. The most prominent among these difficulties are background, font size, size of images, quality and format differences. In order to overcome these difficulties and to obtain accurate and meaningful texts, modern solutions are needed. Addressing these challenging problems in the right way is critical in terms of increasing the efficiency of OCR systems. When the OCR studies carried out to overcome these difficulties are analysed, it is seen that two different approaches, machine learning based approaches and deep learning based approaches, dominate the studies.

Inspired by the power of deep learning in different problems, it is thought that deep learning-based approaches can be used in OCR systems [20]–[23]. As a result of the research conducted in this context, deep learning-based approaches that tackle OCR problems in the literature are presented below. Niranjan and Kumar developed a hybrid method using BiLSTM and CNN architectures to improve text recognition performance [24]. They use a synthetic dataset called MJSynth in their study. Yang et al. introduce a deep learning based PSENet architecture for extracting meaningful textual information from natural scene images [25]. Kirthiga et al. carried out a study on automatic recognition of Tamil scripts using deep learning based character recognition algorithms. They applied various preprocessing techniques to remove noise in Tamil scripts and to improve images. The authors reported that obtaining strong features is important in character recognition and compared the segmentation of characters, recognition rates and accuracy [26]. Batra et al. discuss the benefits of the OCR approach in the healthcare system, especially OCR applications that enable the digitisation of medical laboratory records [27]. Tesseract, EasyOCR and DocTR, three widely used OCR applications and pre-processing techniques such as image thresholding, filtering and sharpening

are examined in detail. Among these, DocTR is reported to be supported by Tensorflow 2 and Pytorch artificial intelligence libraries [27]. Low et al. evaluate the performance of pre-trained deep learning models for text detection and recognition on different datasets, especially for the detection and recognition of seven segment digits. According to different text detection algorithms, PaddleOCR's DBNet method gives a good result in text detection, while the PARSeq algorithm gives a good result in the recognition of seven segment digits [28]. Singh et al. developed an OCR model with CNN and BiLSTM based architectures using language models trained on the JFLEG dataset [29]. Gujjeti et al. develop a deep learning algorithm for detecting and classifying text in scene images [30]. In the developed approach, CNN architecture is used to extract features from each of the segmented scene images to be recognized. In the recognition and classification of characters, classification is performed with the SVM algorithm, one of the popular machine learning methods. CNN-based algorithms are reported to provide good results in extracting features of characters in scene text images. CNN architectures are very useful for recognizing characters in scene images due to their high robustness to changes in direction, scale, and noise [30]. Chandio et al. developed a CNN-based method for extracting features and combining multi-level features to recognize Urdu characters [31].

Machine learning based approaches that tackle OCR problems in the literature are presented below. Authors in the [32] presented a number recognition system by using principal component analysis and linear discriminate analysis to determine the distinct between person and machines. Thus, security communication connection between person and machines is provided by the classification and evaluation of electroencephalogram signals. Authors in the [33] convert ANN based on desktop character recognition application to mobile application with piecewise linear methodology. They have increased up to 60% the rate of speed of without any change in the rate of recognition. They have expressed to down half of the size of code in this made the system. Authors in the [34] presented text recognition system. The system uses images obtained from phone camera. Firstly, they detected regions as English texts by using DCT statistical values. Secondly, candidate text regions are merged by using morphological operations. Thirdly, thresholding is applied to increase the separation between foreground and background of text regions. Fourthly, they carried out text recognition by using Tesseract character engine. In this point, English characters obtained from fourth process are translated to Spanish characters. They search all characters of the character engine by using binary search algorithm in the translation process. Authors in the [35] presented number recognition system by using a classifier based on Euclidean and Mahalanobis distance. The system consists of thresholding, features obtained from statistical values, two classifiers. Authors in the [36] presented thresholding technique which is identified of broken Gujarati characters. Authors in the [37] presented character recognition system. The system consists of feature extraction with chain, distance and gradient based features, classification with linear discriminant, quadratic discriminant and KNN classifiers. Authors in the [38] presented number recognition system which uses six different moment features. They use multilayer perception classifier in the classification of the numbers. Authors in the [39] presented number recognition system which consists of pre-processing with DCT, matrices transform of values obtained from pre-processing, classification with SVM of these matrices.

OCR models based on deep learning are said to give better performance results than traditional methods [1]. When the studies in this field are analysed in the literature, it is seen that OCR is generally performed using ready-made datasets and deep learning models with pre-trained weights. In order to overcome this deficiency, a new OCR dataset was prepared. At the same time, a detailed comparison of the ANN method, which is a machine learning based method, and the deep learning based proposed CNN method was carried out.

3. Material and Method

3.1. Material

In this section of the study, the number dataset used in the study prepared by Çetiner in 2012 was used first [40]. In the study, this dataset is named as Dataset-I. There are 1160 images in Dataset-I, half of which are images of female and half of which are images of male TC IC cards. The images obtained within the scope of this study were obtained by preparing a special hardware setup. This setup is presented in detail in Sections 3.1.1 and 3.1.2.

3.1.1. Identification Card Recognition System (ICRS)

In this paper, ICRS is introduced. ICRS has been tested on TR IC images. TR IC is a card with a blue or pink background colour. The drawn patterns in the background make it difficult to recognise the numbers in the foreground. The TR IC usually deforms

because the level of light in environment affects the quality of the image taken from the camera. In addition to, it may be deformed due to many reasons such as identity plated, cold stamp. Figure 1 shows algorithm of the ICRS. All steps of this study are given in the Figure 1. The ICRS consists of hardware design, pre-processing, feature extraction, classification, recognition respectively.

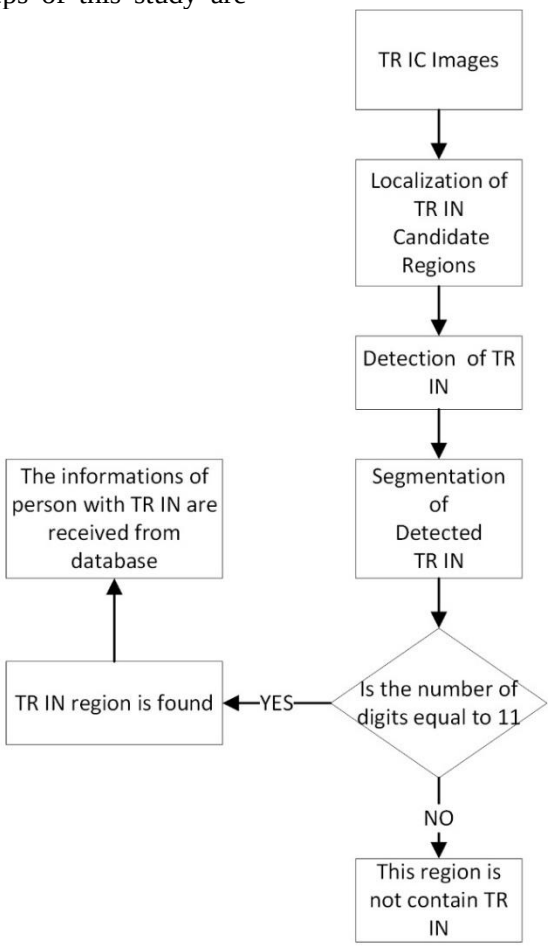


Figure 1. Identification card flow diagram [40]

3.1.2. Hardware Design

The hardware system is designed to prevent light changes. Otherwise, the images not taken under certain standards changes occur in many parameters such as quality, resolution, light level, lighting conditions. This condition makes difficult to localization of TR IN. The ICRS has been standardized to overcome all these difficulties by making closed cabinet. Figure 2 shows the designed model of cabinet. This cabinet consists of a camera, a camera lens, an illumination circuit, a computer that is connected via the USB port, an adjust power supply. An adjustable adapter is used as a power source. This adapter was used to provide power of the illumination circuit with LEDs (Light Emitting Diode).

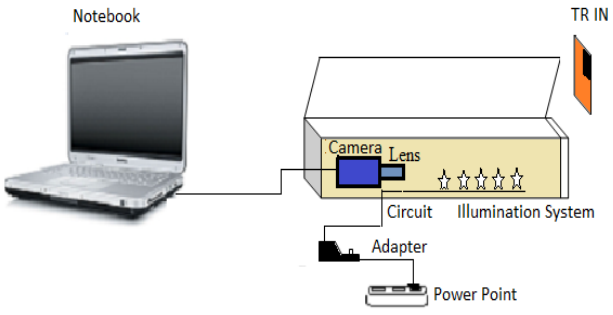


Figure 2. Test System Design [40]

The images obtained from Matrix Vision CCD BlueFox camera is send to computer via the USB port connection by placed within the cabin. There is place of IC on the other side and camera on one side of the cabin. Some statistical analysis is extracted to determine the best light values of the

image obtained from the camera. Table 1 shows that voltage, brightness, contrast and recognition test results of images obtained from designed test system in the night environment. Illumination of camera within the closed cabin is provided with white LEDs placed on the side wall of the cabin.

Table 1. The experimental results in different voltages [40]

Voltage (Volt)	Brightness Value	Contrast Value	Recognition
1.5	0.00	0	Unsuccessfully
3	0.00	0.06	Unsuccessfully
4.5	0.00	0.06	Unsuccessfully
6	2.49	2.41	Unsuccessfully
7.5	88.34	45.01	Successfully

With the hardware and algorithm presented in Sections 3.1.1 and 3.1.2, 116 images of each class were obtained using the numeric images obtained from TR ICs. These images were created separately for each of the digits between 0 and 9. Some of the sample images of the dataset used in the study are presented in Figure 3.

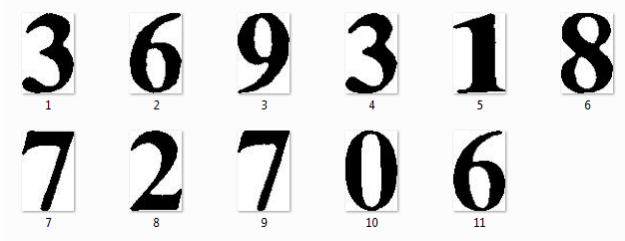


Figure 3. Images obtained with the prepared hardware [40]

In order to increase the number of images in the created dataset, images from the Standard OCR dataset were also included in the study [41]. In the study, the standard OCR dataset is named Dataset-II. Some of these images are also presented in detail in Figure 4.

Figure 4. Some of the images in the standard OCR dataset

In this way, a total of 2320 images were obtained by combined Dataset-I [40] to and the Dataset-II [41]. The dataset formed by the combined of these two datasets is named Dataset-III. When the studies on the subject in the literature are examined, it is seen that data augmentation is generally performed when the amount of data is low [42]. For this reason, the number of images was increased to 6960 by applying rotation, mirroring and inversion operations, which are the basic data augmentation operations. The new dataset obtained as a result of the data augmentation processes mentioned in the study is named Dataset-IV. The datasets used in the study are presented in detail in Table 2.

Table 2. Presented display of datasets

Datasets	Description	Number of images used
Dataset-I ([40])	The number dataset used in the study prepared by Çetiner in 2012 was used first [40].	1160
Dataset-II ([41])	Standard OCR dataset	1160
Dataset-III ([40] +[41])	In order to increase the number of images in the created dataset, images from the Standard OCR dataset were also included in the study	2320
Dataset-IV ([40] +[41] with data augmented)	Extended dataset with data augmentation technique	6960

3.2. Methods

In this study, classification performance results were obtained by using ANN and proposed CNN model on the generated dataset. For this purpose, firstly, basic preprocessing was performed on the images. In the second step, characters were detected by using horizontal and vertical histograms on the image. In the third step, features were obtained from the characters for the machine learning algorithms to be used in the study. At this stage, the features with the highest discrimination were used in the ANN

algorithm. In the last step, which is the fourth step, the classification process was performed. At this stage, ANN and proposed CNN model were applied. The steps performed in the study are presented in detail in this section.

3.2.1. Preprocessing

Localization process of TR IN in TR IC is difficult due to the fact that TR IC has two different number regions. For this reason, the ICRS determines location of TR IN early. The dynamic thresholding of image is important to in real-time systems. The brightness and contrast adjustments were used for the transaction. The TR IC images have wear at different rates, various degrees of degradation. These cards printed by the machine may have different font, width and height values. The ICRS have determined that the density changes are caused by these changes. A separate threshold value for each image is determined due to these changes. To address this problem, the ICRS tried to set the appropriate brightness and contrast. After the experiment, binary images were obtained on the desired usefulness by using the average and standard deviation of gray level image.

The weighted average of the image gray values is an indicator of the image brightness. The weighted standard deviation of the image gray values is an indicator of the image contrast. Eq. 1 calculates brightness value of image.

$$IB_i{}^N = \sum_{i=0}^{N} \mu_i - I, 1,2, \ldots N \qquad (1)$$

$IB_i{}^N$ is insufficient brightness value of image with i index. i is the index number of the image. N is the maximum index number image of the image database obtained from the ICRS with real time. Idenotes the RGB image. μ denotes the mean of image. Eq. 2 shows that increasing the insufficient brightness value of the image pixel can be removed from a fixed c value to adjust the brightness of the image.

$$Ind_i^N = IB_i^N < 50\{f_{imc}(I_{gray}) * (c - IB_i^N)\} \qquad (2)$$

Where f_{imc} is a function that calculates the complement of the I_{gray} which is grayscale image. Eq. 3 shows that increasing the insufficient brightness value of the image pixel can be multiplied from a fixed c value to adjust the brightness of the image.

$$Ind_i^N = IB_i^N >= 50|then$$
$$Th_i^N = f_{ima}\left(\mu(I_{gray})\right) + \sigma(I_{gray}) * (0.3/c) \qquad (3)$$
$$I_{gray} < Th_i^N is \ldots applied$$

Where f_{ima} is a function that adjust image intensity values of the I_{gray} by adjusting at minimum value 0.1 and maximum value 0.9. σ denotes the standard deviation of image. Th_i^N is thresholding value of image with i index. The obtained thresholding method converts the two different RGB regions to binary format according to Eqs. 1, 2 and 3. The two different number regions convert to binary format with a simple thresholding.

3.2.2. The Detection of TR IN and feature extraction

In the detection stage, there are some difficulties due to the differences as font and size in TR IC. In addition to, there is more than one candidate region. The detection of TR IN is carried out in third steps. In the first step, pixel values in the vertical axis of the binary image were collected. In the second step, intersection points with the axis of greater places than total pixel value 0 have been appointed 1 according to the sum of the pixel values. In the third step, segmented regions are obtained with logic applied in the second step according to vertical projection in the vertical axis.

The ICRS is carried out connected component of each segmented region. Then, the number of characters detected with 8 connected objects in each region are kept. Then, horizontal projection of each segmented region is obtained to ensure the accuracy of the detected region and total pixel values of the region are calculated. The ICRS is calculated parameters such as the average distance between the characters, distance between characters, the number of characters for each candidate TR IN region. At this stage, wavelet transform is used for feature extraction.

Wavelet transforms use two functions that can extract local and global features of patterns by performing localised analysis. One of these two functions is the wavelet function determined by the high pass filter. The other is the scaling function, which is the low-pass filter that produces the wavelet approximations [43]. In image processing, discrete wavelet transform (DWT) is generally employed. In DWT, the mother wavelet $\psi(t)$ is scaled and translated at different decomposition level, and then the child wavelet $\psi_{m,n}(t)$is convolved with the image. The child wavelet is given as;

$$\psi_{m,n}(t) = \frac{1}{\sqrt{m}}\psi\left(\frac{t-n}{m}\right), m > 0, n \in \Re \quad (4)$$

In Eqs. 4, 5, and 6, m and n are scaling and shifting parameters in the time domain, respectively. In generally wavelet and scaling function transform are determined with multiples of two ([44]):

$$\left.\begin{matrix}\varphi_{m,n}(t) = 2^{-m/2}\varphi(2^{-m}t - n) \\ \psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n)\end{matrix}\right\} \quad (5)$$

For the S_t time series which are containing the samples of N has an i discrete time step, the discrete wavelet transform defined as in Eq. 6:

$$W_{m,n} = \int_{-\infty}^{\infty} f(t)\,\psi_{m,n}(t)dt \quad (6)$$

These coefficients measure the variations of the field $f(t)$ about the point n with the scale given by m and in Eq. 6. Also, $W_{m,n}$ has a time value that is the wavelet coefficients. Discrete wavelet transform separates the two main components of signals. These processes are done using filters. The first of these main components is low-frequency components, thus low-pass filter outputs. Other high-frequency components, thus high-pass filter outputs.

Approaches (A) are low-frequency components of the main components and Details (D) are high-frequency components of the main components in composed of discrete wavelet transformation. $S(t)$ signal is done parsing process by subjecting to discrete wavelet transform process. It is separated to equal the number of sub-frequency component, is called $A$and high-frequency component, is called D. The detail and approximation coefficients of the S(t) signal shown in Figure 6 are high-frequency low-scale components and low-frequency high-scale values, respectively. In Figure 6 observed that decomposition of a $S(t)$ signal in third level.

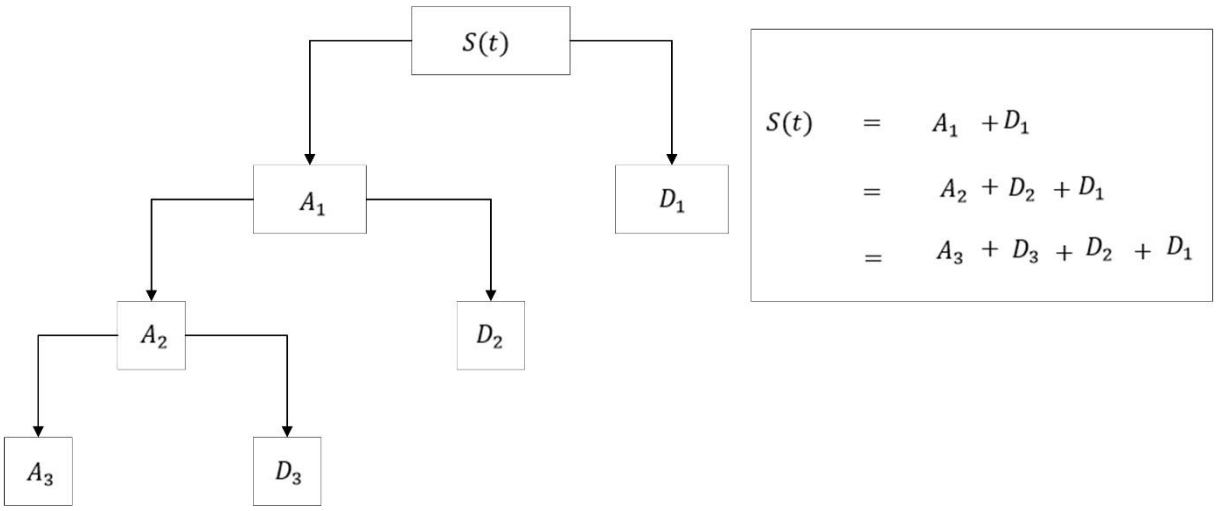


Figure 5. Filtering process in wavelet transformation [40], [43]

After leaving the sub-bands with the standard wavelet transform in Figure 5, approach component of signal is divided into sub-bands with same procedure again and this procedure continues until it reaches the desired resolution. As can be seen in Figure 5, the signal $S(t)$ subjected to discrete wavelet transformation. This signal does not any loss, only its components separated multiplied by main wavelet. In this here, A_3 approach component is basic component. Approaches are low frequency components with high weight. Details are high frequency components with low weight. As can be seen the purpose of wavelet transformation, signal are decomposed approach and detail components. The resulting approach is decomposed approach and detail components in second level again; this process can be repeated until the desired result is obtained. Signal can be separated in its components at desired resolution level by process of separating the components is sequential repeated [43]. Sequentially repeating the process is shown in Figure 5. Firstly, original $S(t)$ signal passes through high-pass $g(t)$ filter and a low-pass $h(t)$filter. This is a separation process with level and mathematically expressed as in Eq. 7 and Eq. 8:

$$D(k) = \sum_{t} S(t)\, g(2k - t) \qquad (7)$$

$$A(k) = \sum_{t} S(t)\, h(2k - t) \qquad (8)$$

Following of sub-sampling with two in Eq. 7 and Eq. 8, respectively high-pass and low-pass filter outputs. The k is the level of the filter. In this way, feature vector is obtained by TR IN images are passed through filters. After these operations, the two-dimensional discrete wavelet transform function, the approximation and other detail information of an image with a 3rd level filtering of Daubechies 'db2' type are shown in Figure 6.

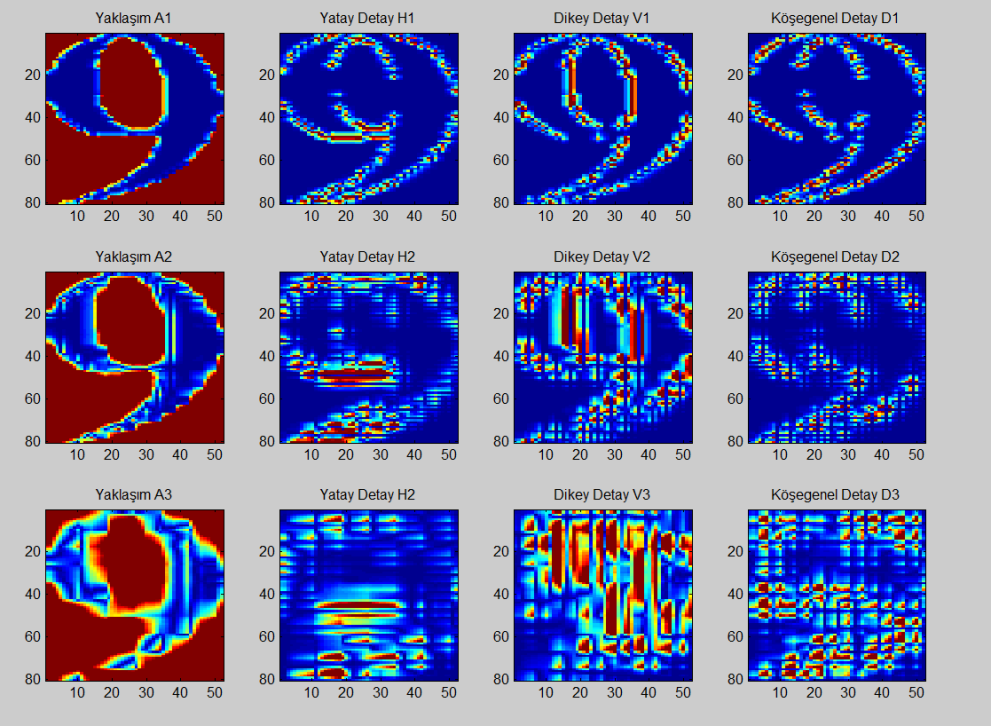


Figure 6. Third level approximation and detail features representation [40]

As shown in Figure 6, the wavelet-based feature with the most distinctive features is obtained from the 3rd level approximation coefficients. Although these features can be obtained automatically by convolution in CNN based algorithms, it is important to give highly discriminative features as input in classical machine learning based algorithms such as ANN.

3.2.3. TR IN Classification with ANN

Since the creation of human are living together with nature, and they have developed solutions for the problems faced by inspired from nature. These methods have made much progress the rapid development of numerical calculations is made using the computer. When some of the methods are developed, they are done by the example of living organism. ANN in this way is a method developed taking into account the structure of human biological neuron. In this study, the model ANN input is selected as 3480. This part is used for training purposes of 3480 feature vector. Network training is carried out by using in model ANN training of model input and output files are obtained.

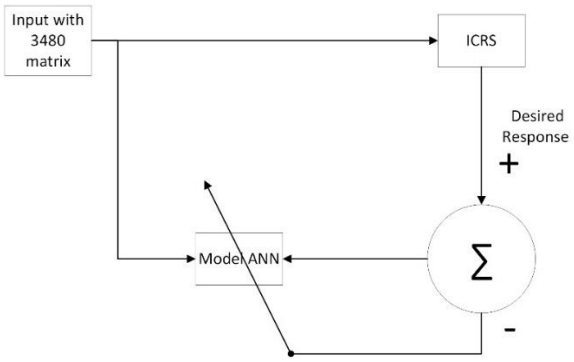


Figure 7. Training block diagram of the ANN model [40]

Each character is defined as a row matrix with 3480 inputs and 10 outputs for ANN training. The block diagram realized in this context is shown in

Figure 7. Thus, a model ANN that successfully recognizes a character given to the input has been obtained.

3.2.4. TR IN Classification with Proposed CNN

Unlike the ANN method used in Section 3.2.3, the methods commonly used in the current literature were investigated and as a result of this research, it was determined that there is a tendency towards Convolutional neural networks. For this reason, CNN-based methods were focused in this study. In this study, a new CNN model consisting of 15 layers is proposed. In the convolution layers of the proposed model, the ReLU activation function presented in Eq. 9 is used. The x value presented in Eq. 9 represents the input value.

$$f(x), = max(0, x) \qquad (9)$$

Thus, the values obtained were kept within certain limits. In the study, unlike the literature, 3 new modules connected consecutively were developed. In this proposed module, inspired by the Inception module, the 1x1 convolution process was first applied. In the second step, 3x3, 5x5, 7x7, Mixed Pooling (2x2), Max Pooling (2x2) and Soft Pooling (2x2) operations were applied to the data obtained in parallel.

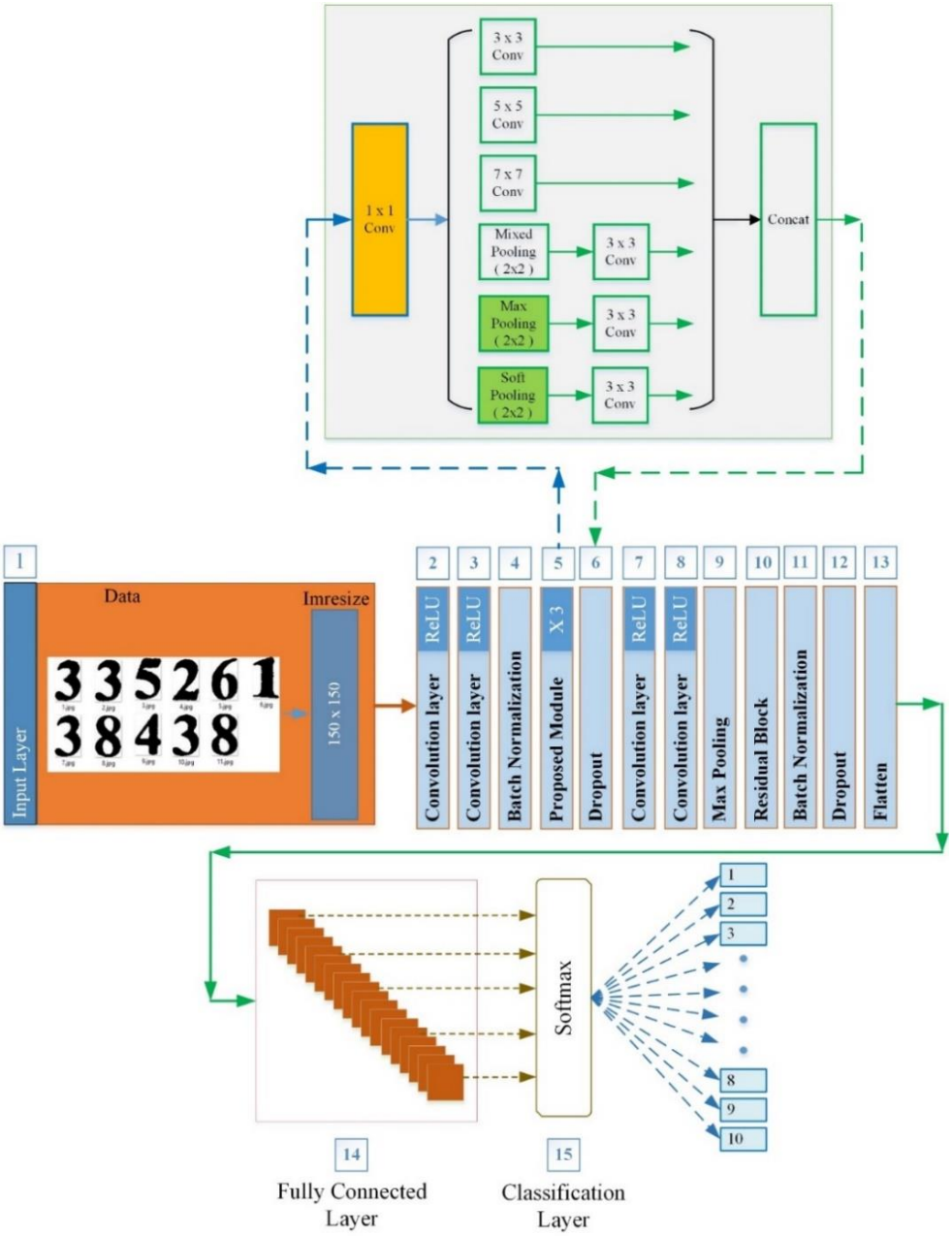


Figure 8. General structure of the proposed CNN model for OCR

The 3x3 convolution process was applied again to the data obtained from the pooling processes. As shown in Figure 8, new features were obtained by applying Concat process to the data obtained as a result of the mentioned processes. The data obtained with these methods were sent to the same module again in the next step and these processes were repeated 3 times in a row. Among the features obtained, there are very strong features as well as weak features. For this reason, Dropout layer was added to the proposed model. In the next step, convolution process was applied to the features obtained from the dropout layer. Max Pooling process was applied to the values obtained after these steps. Residual operation, which is the cornerstone of ResNet architecture, was applied to the values obtained after the pooling process. At this stage, dropout layer was applied due to too many features being obtained again and then Flatten operation was performed.

$$softmax(z_j) = \frac{e^z j}{\sum_{k=1}^{K} e^z k}, for\ j = 1, \dots, K \qquad (10)$$

In the last stages, fully connected layer and classification layer are added as in classical convolutional operations. In the classification layer, the SoftMax function presented in Eq. 10 is used. The expression *K* in Eq. 10 expresses the number of classes. The z_j value obtained as a result of these operations is the value that shows which class the data belongs to. The largest value obtained at which *K* value means that the data belongs to that class.

4. Results and Discussion

A study is conducted to analyse the success rates in the classification of number features obtained with WT of ANN algorithm. This study is carried out in MATLAB 7.14 version. Experimental results have been obtained in Windows 10 operating system which is Intel Core i5 CPU (2.7 GHz) processor with 4GB RAM.

A classifier is trained to identify differences or types of samples among separated samples in classification. Task of the classifier demonstrates whether or not represented a specific object or area of interest in image. To achieve this, prior classifier is sufficient knowledge about the image. WT is used to feature extraction from images. WT is a feature extraction algorithm. 648 features are obtained with WT.

The numbers obtained from different TR IC are given training results using different types of wavelets and the different training algorithms of ANN in the previous my study. These results show that in Table 3 and Table 4. In addition to, even if oblique TR IN, the ICRS carries out successfully localization and detection of TR IN. The neural network training algorithms such as Scaled Conjugate Gradient classification (SCG), Flexible Back Propagation (FBG), Conjugate Gradient with Powell-Beale (CGPB), Fletcher-Reeves Conjugate Gradient (FRCG), Conjugate Gradient with Polak-Rbiere (PRCG) are used in classification. These training algorithms are used features which are extracted with different wavelet types of TR IC images. Results are obtained from using these training algorithms which are given in Table 3 and Table 4. In separation of high and low frequency components of the signal, Daubechies (1, 2) which is a type of mother wavelet, Symlet (3), Biortogonal (2.2, 3.1, 3.9), Reverse Biorthogonal (3.1), Meyer wavelet are wavelet types. These types are examined with each training algorithm and training/test times are given in Table 3.

Table 3. ANN training times of features obtained by the wavelet transform in different types and grades of the figures

Method	Datasets	Training Times (second) of Dataset							
		Db1	*Db2*	*Bior 2.2*	*Bior 3.1*	*Bior 3.9*	*Sym3*	*dmey*	*Rbio 3.1*
SCG	Dataset-I	65	68	70	65	80	70	76	67
	Dataset-II	66	69	71	66	81	71	77	68
	Dataset-III	131	137	141	131	161	141	153	135
	Dataset-IV	390	480	420	456	402	390	480	420
FBG	Dataset-I	100	102	106	99	108	113	105	109
	Dataset-II	101	101	107	98	109	112	106	108
	Dataset-III	201	203	213	197	217	225	211	217
	Dataset-IV	594	648	678	630	654	594	648	678
CGPB	Dataset-I	39	41	38	43	56	40	60	40
	Dataset-II	38	42	37	44	55	41	59	41
	Dataset-III	77	83	75	87	111	81	119	81
	Dataset-IV	258	336	240	360	240	258	336	240
FRCG	Dataset-I	98	103	80	103	120	103	125	117
	Dataset-II	98	103	80	103	120	103	125	117
	Dataset-III	196	206	160	206	240	206	250	234
	Dataset-IV	618	720	618	750	702	618	720	618
PRCG	Dataset-I	99	98	76	98	110	110	120	100
	Dataset-II	100	98	75	99	110	109	121	100
	Dataset-III	199	196	151	197	220	219	241	200
	Dataset-IV	588	660	660	720	600	588	660	660
Average	Dataset-I	80.2	82.4	74	81.6	94.8	87.2	97.2	86.6
	Dataset-II	80.6	82.6	74	82	95	87.2	97.6	86.8
	Dataset-III	160.8	165	148	163.6	189.8	174.4	194.8	173.4
	Dataset-IV	489.6	568.8	523.2	583.2	519.6	489.6	568.8	523.2

Table 4. ANN training performance results of features obtained by the wavelet transform in different types and grades of the figures

Wavelet Type	Dataset-I		Dataset-II		Dataset-III		Dataset-IV	
	TrRR(%)	TsRR(%)	TrRR(%)	TsRR(%)	TrRR(%)	TsRR(%)	TrRR(%)	TsRR(%)
Db 1	100	97.7	94.3	92.2	96.4	93.4	100	99.2
Db 2	99.5	98.3	93.1	91.8	95.3	92.9	99.8	99.4
Bior 2.2	100	99.4	93.7	93.4	95.3	96.2	100	99.8
Bior 3.1	99.4	95.4	92.9	90.4	94.8	92.6	99.8	98.7
Bior 3.9	98.6	96.2	91.4	91.4	93.7	93.5	99.7	99.3
Sym 3	98.9	97.1	91.2	90.4	94.4	93.3	99.8	99.5
Dmey	99.3	94.8	92.8	89.5	95.7	93.4	99.8	98.4
Rbio 3.1	99.7	98.2	93.2	92.4	96.4	95.3	100	99.8

While TrRR in Table 4 represents the training recognition rate, TsRR represents the Test Recognition Rate. The most preferable type of wavelets is determined according to the best training and testing rates in Table 3 and Table 4. This type is wavelet type which is called Bior 2.2.

Table 5. Proposed CNN performance results for different datasets

Datasets		F1 score	Recall	Precision	Accuracy
Dataset-I	Training	0.9936	0.9896	0.9977	0.9924
Dataset-I	Validation	0.9896	0.9850	0.9943	0.9902
Dataset-II	Training	0.9511	0.9467	0.9556	0.9532
Dataset-II	Validation	0.9515	0.9435	0.9598	0.9586
Dataset-III	Training	0.9609	0.9552	0.9667	0.9688
Dataset-III	Validation	0.9607	0.9523	0.9694	0.9694
Dataset-IV	Training	0.9917	0.9887	0.9948	0.9950
Dataset-IV	Validation	0.9881	0.9843	0.9921	0.9927

In this study, CGPB training algorithm type is preferred to classify of features extracted with Bior 2.2 wavelet type, because it is completed the classification at the earliest time. Then, FDR is used in order to improvement these results obtained. Finally, selected features are classified using proposed CNN and ANN classification algorithms in real-time. The results of this classification are evaluated and compared.

When the data presented in Table 4 and Table 5 are evaluated together, it is observed that both models give good performance results. In accordance with these performance results, it is seen that both models are at a level that can in competition with the studies in the literature.

5. Conclusion and Suggestions

In the presented work, TR IN is recognised in real time by developing ICRS. The ICRS project based on real-time machine vision system to accelerate works of the institutions employees. The localization, detection and segmentation steps of the TR IN have an important role in the successful conclusion. CGPB training algorithm type is preferred to classify of features extracted with Bior 2.2 wavelet type, because it is completed the classification at the earliest time. In addition to, training recognition rate and test recognition rate with Bior 2.2. wavelet type achieves classification rate with 100% and 99.4% respectively for the Dataset-I.

In the Dataset-IV, the highest values achieved with ANN were obtained using Rbio 3.1 and Bior 2.2 wavelet types. These values were found to be 100% and 99.8% accuracy for training and testing, respectively. As can be seen from this result, the performance of the Rbio 3.1 wavelet type has also improved as a result of the expansion of the dataset. Similarly, performance improvement was also observed in other wavelet types. In the proposed CNN model, 99.24 %, 98.96 %, 99.77%, 99.77%, and 99.24% values were obtained for F1 score, recall, precision, and accuracy metrics in training for the Dataset-I, respectively. For validation, 99.12%, 98.50%, 99.43%, and 99.02% values were obtained in the same order. In the Dataset-IV, 99.49%, 98.87%, 99.48%, and 99.50% values were obtained in the same order for training. Similarly, for validation, values of 99.13%, 98.43%, 99.21%, and 99.27% were obtained in the same order.

An important limitation of the study is that max pooling, soft pooling, mixed pooling functions are used simultaneously. In the literature research, it is seen that there are different pooling functions. The contribution of the study to the literature is that soft pooling, mixed pooling and max pooling functions are used simultaneously. In future studies, different pooling functions will be tested and their effects on performance will be examined in detail.

As can be understood from these values, the performance of the proposed model in optical character recognition is promising. It is anticipated that the proposed model can also be applied in advanced embedded systems in future studies.

Contributions of the authors

Çetiner: Literature review, Validation, Conceptualization, Formal analysis, Methodology, Software development, Visualization, Writing, review and editing
Cetişli: Supervision

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

- This study is derived from the master's thesis titled Optical Character Recognition Using Image Processing Techniques, completed at Süleyman Demirel University, Graduate School of Applied and Natural Sciences, Computer Engineering Department [40].
- A small part of this work was presented as a full-text paper titled "Real time recognition of identification cards of Turkish Republic with wavelet transforms" at the 21st Signal Processing and Communications Applications Conference (SIU) in 2013 [45])

References

[1] X.-F. Wang, Z.-H. He, K. Wang, Y.-F. Wang, L. Zou, and Z.-Z. Wu, “A survey of text detection and recognition algorithms based on deep learning technology,” *Neurocomputing*, vol. 556, pp. 126702, 2023.

[2] R. Mohapatra, B. Majhi, and S. Jena, “Printed Odia Digit Recognition Using Finite Automaton,” in *Proc. 3rd Int. Conf. Adv. Comput. Netw. Informatics*, 2016, pp. 643–650.

[3] S. Ghosh, N. Das, M. Kundu, and M. Nasipuri, "Handwritten Oriya Digit Recognition Using Maximum Common Subgraph Based Similarity Measures," in *Inf. Syst. Des. Intell. Appl.*, 2016, pp. 165–173.

[4] A. Ghahnavieh and A. Raie, "A New Hierarchical Architecture Based on SVM for Persian License Plate Character Recognition," *J. Adv. Comput. Res.*, vol. 7, no. 1, pp. 49–66, 2016.

[5] S. Kaur, "An Automatic Number Plate Recognition System under Image Processing," *Int. J. Intell. Syst. Appl.*, vol. 8, no. 3, pp. 14, 2016.

[6] S. Babu and Z. Masood, "Android Based Optical Character Recognition for Noisy Document Images," *Int. J. Comput. Sci. Inf. Secur.*, vol. 14, no. 1, pp. 34, 2016.

[7] G. Joseph and T. Singh, "Registration Plate Recognition Using Dynamic Image Processing and Genetic Algorithm," in *Innovations in Computer Science and Engineering*, 2016, pp. 37–43.

[8] I. Ahmad, S. Mahmoud, and G. Fink, "Open-vocabulary recognition of machine-printed Arabic text using hidden Markov models," *Pattern Recognit.*, pp. 97–111, 2016.

[9] H. Cecotti, "Active graph based semi-supervised learning using image matching: Application to handwritten digit recognition," *Pattern Recognit. Lett.*, pp. 76–82, 2016.

[10] H. Sajedi, "Handwriting recognition of digits, signs, and numerical strings in Persian," *Comput. Electr. Eng.*, vol. 49, pp. 52–65, 2016.

[11] S. Karthikeyan, A. G. S. de Herrera, F. Doctor, and A. Mirza, "An OCR Post-Correction Approach Using Deep Learning for Processing Medical Reports," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2574–2581, 2022.

[12] S. Naz, S. Ahmed, R. Ahmad, and M. Razzak, "Arabic Script based Digit Recognition Systems," in *Int. Conf. Recent Adv. Comput. Syst.*, 2016, pp. 67-73.

[13] V. Aradhya, G. Kumar, and S. Noushath, "Robust unconstrained handwritten digit recognition using radon transform," in *Signal Process. Commun. Netw.*, 2007, pp. 626–629.

[14] H. Boveiri, "Transformation-Invariant Classification of Persian Printed Digits," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 4, no. 3, pp. 153–164, 2011.

[15] R. Azad, F. Davami, and H. Shayegh, "Recognition of Handwritten Persian/Arabic Numerals Based on Robust Feature Set and K-NN Classifier," in *2013 First Int. Conf. Comput. Inf. Technol. Digit. Media (CITADIM Proceeding-Scientific)*, 2014.

[16] M. Ghaleb, L. George, and F. Mohammed, "Printed and Handwritten Hindi/Arabic Numeral Recognition Using Centralized Moments," *Int. J. Sci. Eng. Res.*, vol. 5, no. 3, pp. 140–144, 2014.

[17] C. Liu and C. Suen, "A new benchmark on the recognition of handwritten Bangla and Farsi numeral characters," *Pattern Recognit.*, vol. 42, no. 12, 2009.

[18] I. Meymand, "Recognition of Handwritten Persian/Arabic Numerals Based on Robust Feature Set and K-NN Classifier," *Int. J. Comput. Inf. Technol.*, vol. 1, no. 3, pp. 220–230, 2013.

[19] D. Sharma, G. Lehal, and P. Kathuria, "Digit extraction and recognition from machine printed Gurmukhi documents," in *Proc. Int. Work. Multiling. OCR*, 2009.

[20] S. Metlek and H. Çetiner, "ResUNet+: A New Convolutional and Attention Block-Based Approach for Brain Tumor Segmentation," *IEEE Access*, vol. 11, pp. 69884–69902, 2023.

[21] S. Metlek, "CellSegUNet: an improved deep segmentation model for the cell segmentation based on UNet++ and residual UNet models," *Neural Comput. Appl.*, vol. 36, pp. 5799-5825, 2024.

[22] H. Çetiner and S. Metlek, "DenseUNet+: A novel hybrid segmentation approach based on multi-modality images for brain tumor segmentation," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 8, pp. 101663, 2023.

[23] S. Metlek, "A new proposal for the prediction of an aircraft engine fuel consumption: a novel CNN-BiLSTM deep neural network model," *Aircr. Eng. Aerosp. Technol.*, vol. 95, no. 5, pp. 838–848, Jan. 2023.

[24] J. N. and S. K. A. M, "A Novel Text Recognition Using Deep Learning Technique," in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, 2024, pp. 1–6.

[25] L. Yang, D. Ergu, Y. Cai, F. Liu, and B. Ma, "A review of natural scene text detection methods," *Procedia Comput. Sci.*, vol. 199, pp. 1458–1465, 2022.

[26] R. Krithiga, S. Varsini, R. G. Joshua, and C. U. O. Kumar, "Ancient Character Recognition: A Comprehensive Review," *IEEE Access*, pp. 1, 2023.

[27] P. Batra, N. Phalnikar, D. Kurmi, J. Tembhurne, P. Sahare, and T. Diwan, "OCR-MRD: performance analysis of different optical character recognition engines for medical report digitization," *Int. J. Inf. Technol.*, vol. 16, no. 1, pp. 447–455, 2024.

[28] L. M. Low, F. H. M. Salleh, Y. F. Law, and N. Z. Zakaria, "Detecting and recognizing seven segment digits using a deep learning approach," in *ITM Web of Conferences*, 2024, pp. 1007.

[29] A. Singh, S. Jangra, and G. Aggarwal, "EnvisionText: Enhancing Text Recognition Accuracy through OCR Extraction and NLP-based Correction," in *2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2024, pp. 47–52.

[30] S. Gujjeti, M. S., and V. G., "A Systematic Investigation on Different Scene Text Detection and Recognition Method using Deep Learning Techniques," in *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, 2024, pp. 1–5.

[31] A. A. Chandio, M. Asikuzzaman, and M. R. Pickering, "Cursive character recognition in natural scene images using a multilevel convolutional neural network fusion," *IEEE Access*, vol. 8, pp. 109054–109070, 2020.

[32] R. Deore, B. Gawali, and S. Mehrotra, "Digit Recognition System Using EEG Signal," in *Brain-Computer Interfaces*, Cham: Springer Int. Publ., 2015, pp. 375–416.

[33] J. Park and Y.-B. Kwon, "An Embedded OCR: A Practical Case Study of Code Porting for a Mobile Platform," in *2009 Chinese Conference on Pattern Recognition*, 2009, pp. 1–5.

[34] A. Canedo-Rodriguez, S. Kim, J. H. Kim, and Y. Blanco-Fernandez, "English to Spanish translation of signboard images from mobile phone camera," in *IEEE Southeastcon 2009*, 2009, pp. 356–361.

[35] M. Nava-Ortiz, W. Gómez, and A. Díaz-Pérez, "Digit recognition system for camera mobile phones," in *8th International Conference on Electrical Engineering, Computing Science and Automatic Control*, Merida City, Mexico, 2011, pp. 1-5.

[36] R. Shah and T. Ratanpara, “A Mean-Based Thresholding Approach for Broken Character Segmentation from Printed Gujarati Documents,” in *Proc. Second Int. Conf. Comput. Commun. Technol.*, 2016, pp. 487–496.

[37] P. Singh, A. Verma, and N. Chaudhari, “Performance Evaluation of Classifier Combination Techniques for the Handwritten Devanagari Character Recognition,” *Inf. Syst. Des. Intell. Appl.*, pp. 651–662, 2016.

[38] P. Singh, R. Sarkar, and M. Nasipuri, “A Study of Moment Based Features on Handwritten Digit Recognition,” *Appl. Comput. Intell. Soft Comput.*, vol. 2016, pp. 1-17, 2016.

[39] G. C. Batista and W. L. S. Silva, “Application of support vector machines to recognize speech patterns of numeric digits,” in *2015 11th International Conference on Natural Computation (ICNC)*, 2015, pp. 831–836.

[40] H. Çetiner, “Görüntü işleme teknikleri kullanarak optik karakter tanımlama.” Yüksek lisans tezi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği AnaBilim Dalı, Isparta, Türkiye, 2012.

[41] A. Jaiswal, “Standard OCR Dataset,” Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/preatcher/standard-ocr-dataset. [Accessed: Jan. 1, 2024].

[42] A. F. de Sousa Neto, B. L. D. Bezerra, G. C. D. de Moura, and A. H. Toselli, “Data Augmentation for Offline Handwritten Text Recognition: A Systematic Literature Review,” *SN Comput. Sci.*, vol. 5, no. 2, pp. 258, 2024.

[43] M. Misiti, Y. Misiti, G. Oppenheim, and J. Poggi, “Wavelet toolbox,” Matlab User’s Guid., 1997. [Online]. Available: https://feihu.eng.ua.edu/NSF_TUES/w7_1a.pdf.

[44] S. G. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[45] H. Çetiner and B. Cetişli, "Real time recognition of identification cards of Turkish Republic with wavelet transforms," in *2013 21st Signal Processing and Communications Applications Conference (SIU)*, Haspolat, Turkey, 2013, pp. 1-4.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1083-1098 YEAR: 2024
DOI: 10.17798/bitlisfen.1514951

Performance Comparison of Recent Metaheuristic Algorithms on Engineering Design Optimization Problems

Mümin Emre ŞENOL[1*], Tülin ÇETİN[2], Mustafa Erkan TURAN[2]

[1]*Manisa Celal Bayar University, Faculty of Engineering and Natural Sciences, Industrial Engineering Department, Manisa, Türkiye*
[2]*Manisa Celal Bayar University, Faculty of Engineering and Natural Sciences, Civil Engineering Department, Manisa, Türkiye*
(ORCID: 0000-0002-2105-6041) (ORCID: 0000-0002-1511-7338) (ORCID: 0000-0003-2501-2481)

Keywords: Engineering Design Optimization Problems, Metaheuristic Algorithms, Performance Comparison, Optimization.

Abstract

Metaheuristic algorithms have been extensively applied in a variety of complex engineering design optimization problems (EDOPs) due to their capability of yielding near-optimal solutions without excessive computational times. The aim of this study is to investigate the performance comparison among seven novel metaheuristic optimization algorithms: Artificial Hummingbird Algorithm (AHA), Artificial Protozoa Optimizer (APO), African Vultures Optimization Algorithm (AVOA), Electric Eel Foraging Optimization (EEFO), Mountain Gazelle Optimizer (MGO), Pied Kingfisher Optimizer (PKO), and Quadratic Interpolation Optimization (QIO). This comparison is performed with twelve engineering design optimization problems evaluating the best, worst, mean, and standard deviation of their results. We also use non-parametric statistical tests such as the Friedman rank test and Wilcoxon signed rank test to finally compare the performance of algorithms. The results show the merits and demerits of each algorithm, which give us clues on their suitability for different engineering design problems. According to Friedman rank test, EEFO surpasses the other algorithms in these EDOPs. In addition, it performs statistically better than AVOA and QIO according to Wilcoxon signed rank test.

1. Introduction

Optimization is a process carried out to find the values of the variables that maximize or minimize the value of a function within problem-specific restrictions. In optimization problems, the variables are called decision variables, this function is called the objective function, and the restrictions are referred to as constraints. Once the problem is defined, an optimization method is selected based on the problem's characteristics to solve it.

Many researchers have utilized the metaheuristic algorithms for solving optimization problems. As metaheuristic algorithms are stochastic by nature, they cannot guarantee the achievement of the optimal solution. Although the structures of metaheuristic algorithms do not guarantee finding the global optimum, they are used to find the global optimum value or a close value in a timely manner. New metaheuristic algorithms are continually introduced into the literature by researchers. These metaheuristic algorithms, which are quite numerous, show different performances in different optimization problems. Problem sets have been created in the literature to compare the performances of the algorithms. One of problem sets is engineering design optimization problems (EDOPs). Most EDOPs are highly complex, with many variables, constraints, and objectives. These problems are so expensive from the computational point of view that they are not easy to solve in any classical way.

There are several studies showing the effectiveness of metaheuristics in this field. For instance, the hierarchical surrogate-assisted memetic

*Corresponding author: emre.senol@cbu.edu.tr
Received: 12.07.2024, Accepted: 20.09.2024

algorithm (HSAMA) has been proven effective in reducing computational costs and efficient on Competitions on Evolutionary Computation (CEC) benchmark problems as well as a series of real-world engineering design tasks by Zhou [1]. The Artificial Bee Colony (ABC) algorithm improved handling large-scale and constrained optimization problems, which was successful in the solution of CEC benchmark functions and EDOPs [2]. Besides, rank-iMDDE, as an enhanced constrained differential evolution method, has demonstrated competitive results in enhancing solution quality and convergence rates [3]. Additionally, the IEEE CEC'2013 benchmark suite is constructed with characteristics representative of real-world problems, and hence, it aids in benchmarking most evolutionary techniques [4]. Cooperative co-evolution frameworks have effectively dealt with complex design problems that exist in concurrent engineering by decomposing them into smaller subproblems [5]. Furthermore, a Novel Differential Evolution Algorithm (NDE) using triangular mutation rules has also outperformed the list of CEC benchmark functions and EDOPs [6]. The improved Butterfly Optimization Algorithm (BOA) with the use of the cross-entropy method enhances global search abilities to become capable of successfully solving classical EDOPs and the CEC benchmark functions [7]. The Search and Rescue optimization algorithm (SAR) has been successfully applied to solve constrained EDOPs and CEC [8]. Besides, Atomic Orbital Search (AOS) algorithm has also proved effective on constrained EDOPs, including those benchmarked by CEC 2020, using principles from quantum mechanics [9]. Lastly, the convergence rates and performance of the Accelerated Arithmetic Optimization Algorithm by Cuckoo Search (AOACS) algorithm are better when tested using both CEC 2019 functions and EDOPs [10]. These are all studies collectively demonstrating the fact that metaheuristic algorithms offer robust and practical solutions to a large set of benchmarks as well as real-world problems, like CEC competitions.

The aim of this study is to compare the performance of recent metaheuristic optimization algorithms on selected twelve EDOPs. For this purpose, seven metaheuristic optimization algorithms developed in the literature in recent years namely Artificial Hummingbird Algorithm (AHA), Artificial Protozoa Optimizer (APO), African Vultures Optimization Algorithm (AVOA), Electric eel foraging optimization (EEFO), Mountain Gazelle Optimizer (MGO), Pied Kingfisher Optimizer (PKO) and Quadratic Interpolation Optimization (QIO), are used in this study. Despite these algorithms are utilized to solve some EDOPs, it is aimed to compare these algorithms' performance under the same conditions (the same number of function evaluations (FEs) and the same population size) in this study. These algorithms' performance is compared through the best, the worst, standard deviation and mean values. Also, non-parametric statistical tests (Friedman rank test and Wilcoxon signed rank test) are organized to compare the performance further.

2. Material and Method

2.1. Optimization Algorithms

Seven metaheuristic optimization algorithms were used in the study. Brief information about algorithms is given in this section. Detailed information about the algorithms and the source codes can be reached from the references and code columns given in Table 1.

Table 1. Information about optimization algorithms

Algorithm	Inspiration	Reference	Code
Artificial Hummingbird Algorithm (AHA)	Flight skills and foraging strategies of hummingbirds	[11]	[12]
Artificial Protozoa Optimizer (APO)	Foraging, dormancy, and reproductive behaviors of protozoa	[13]	[14]
African Vultures Optimization Algorithm (AVOA)	Foraging and navigation behavior of African vultures	[15]	[16]
Electric Eel Foraging Optimization (EEFO)	Intelligent group foraging behavior of electric eels	[17]	[18]
Mountain Gazelle Optimizer (MGO)	Social life and hierarchy of wild mountain gazelles	[19]	[20]
Pied Kingfisher Optimizer (PKO)	Hunting behavior and symbiotic relationships of pied kingfishers	[21]	[22]
Quadratic Interpolation Optimization (QIO)	Derived from generalized quadratic interpolation method	[23]	[24]

2.2. Engineering Optimization Problems

Twelve EDOPs were used in this study. Detailed information about the EDOPs is given in this section.

The first of the EDOPs is the Cantilever Beam, which focuses on the objective of minimizing the weight of a cantilever beam with a square cross-section. It is shown in Figure 1 that one end of the beam is fixed rigidly, and at the other end, a vertical force is applied. The beam has five hollow square blocks with uniform thickness, whereas the heights (or widths) of these blocks are taken as the decision variables, each having a fixed thickness of 2/3. A general mathematical formulation of the problem may be stated analytically as follows [25]:

Minimize:

$$f(X) = 0.0624(x_1 + x_2 + x_3 + x_4 + x_5) \quad (1)$$

Subject to:

$$g(X) = \frac{61}{x_1^3} + \frac{37}{x_2^3} + \frac{19}{x_3^3} + \frac{7}{x_4^3} + \frac{1}{x_5^3} - 1 \leq 0 \quad (2)$$

Variable Range:

$$0.01 \leq x_i \leq 100, i = 1, \ldots, 5 \quad (3)$$

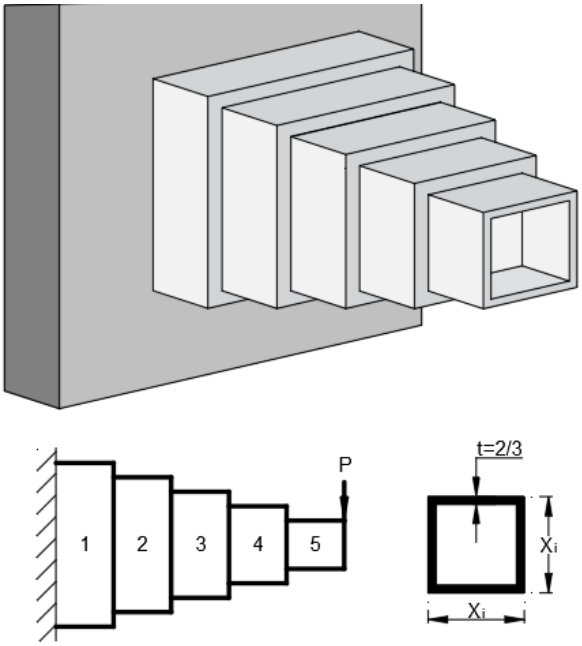


Figure 1. Schematic representation of cantilever beam design problem

Another fascinating problem in EDOPs is the I-beam design problem. It is a crucial benchmark problem; the formulation to be solved is finding the minimum vertical deflection of a beam as shown in Figure 2 considered loads, with constraints on the cross-sectional area and stress. The design variables are the flange width b (x_1), section height h (x_2), web thickness t_w (x_3), and flange thickness t_f (x_4). The maximum deflection y of the beam is given by y = PL^3 / 48EI, where L= 5200 cm is the length of the beam and E = 523.104 kN/cm² is the modulus of elasticity. This problem is stated as [26]:

Minimize:

$$f(X) = \frac{5000}{\frac{x_3(x_2-2x_4)^3}{12} + (x_1 x_4^3/6) + 2bx_4(x_2 - x_4/2)^2} \quad (4)$$

Subject to:

$$g_1(X) = 2x_1x_3 + x_3(x_2 - 2x_4) \leq 300 \quad (5)$$

$$g_2(X) = \frac{18x_2 \times 10^4}{x_3(x_2-2x_4)^3 + 2x_1x_3\left(4x_4^2 + 3x_2(x_2-2x_4)\right)} + \frac{15x_1 \times 10^3}{(x_2-2x_4)x_3^2 + 2x_3x_1^3} \leq 56 \quad (6)$$

Variable Range:

$$10 \leq x_1 \leq 50 \quad (7)$$

$$10 \leq x_2 \leq 80 \quad (8)$$

$$0.9 \leq x_3 \leq 5 \quad (9)$$

$$0.9 \leq x_4 \leq 5 \quad (10)$$

The third EDOPs aims to minimize the volume of a statically loaded three-bar truss design with stress (σ) constraints on its members as shown in Figure 3. The optimization is done to determine the cross-sectional areas A_1 (x_1) and A_2 (x_2). The constants in the formulas are l=100 cm, $= 2\ kN/cm^2$, $\sigma = 2\ kN/cm^3$. Mathematically, the formulation for this problem can be represented as [27]:

Minimize:

$$f(X) = \left(2\sqrt{2}x_1 + x_2\right) \times l \quad (11)$$

Subject to:

$$g_1(X) = \frac{\sqrt{2}x_1 + x_2}{\sqrt{2}x_1^2 + 2x_1x_2} P - \sigma \leq 0 \quad (12)$$

$$g_2(X) = \frac{x_2}{\sqrt{2}x_1^2 + 2x_1x_2} P - \sigma \leq 0 \quad (13)$$

$$g_3(X) = \frac{1}{x_1 + \sqrt{2}x_2} P - \sigma \leq 0 \quad (14)$$

Variable Range:

$$0 \leq x_1, x_2 \leq 1 \quad (15)$$

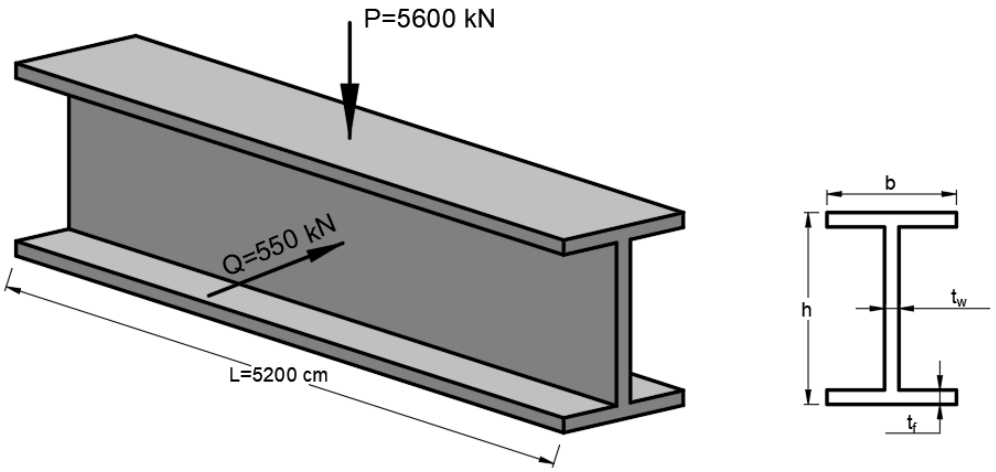


Figure 2. Schematic representation of I shaped beam design problem

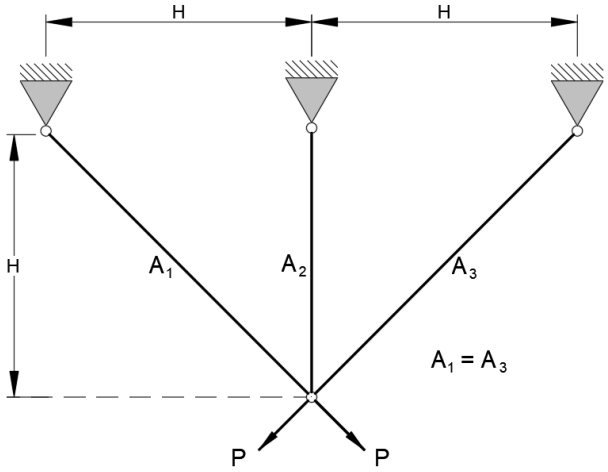


Figure 3. Schematic representation of three-bar truss design problem

Tubular column design problem aims to design a uniform column in the tubular section subjected to a compressive load with minimum cost. The two design variables considered for this problem are the mean diameter of the column, d (x_1), and the thickness of the tube, t (x_2). The cross-sectional geometry and the critical dimensions of the column are shown in Figure 4. The material used to fabricate the column has a yield stress of σ_y = 500 kgf/cm² and a modulus of elasticity of $E = 0.85\times10^6$ kgf/cm². The optimization model for this problem is formulated as follows [28]:

Minimize:

$$f(X) = 9.82x_1x_2 + 2x_1 \quad (16)$$

Subject to:

$$g_1(X) = \frac{2500}{\pi x_1 x_2} - \sigma_y \leq 0 \quad (17)$$

$$g_2(X) = \frac{2500}{\pi x_1 x_2} - \frac{\pi^2 E\left(x_1^2+x_2^2\right)}{8(250)^2} \leq 0 \quad (18)$$

$$g_3(X) = 2 - x_1 \leq 0 \quad (19)$$

$$g_4(X) = x_1 - 14 \leq 0 \quad (20)$$

$$g_5(X) = -x_2 + 0.2 \leq 0 \quad (21)$$

$$g_6(X) = x_2 - 0.8 \leq 0 \quad (22)$$

Variable Range:

$$2 \leq x_1 \leq 14 \quad (23)$$

$$0.2 \leq x_2 \leq 0.8 \quad (24)$$

The constraints g_1 and g_2 ensure that the column-induced stress is less than the buckling and yield stresses, respectively. Furthermore, other constraints namely, g_3, g_4, g_5, and g_6, confine the design variables to their allowable ranges.

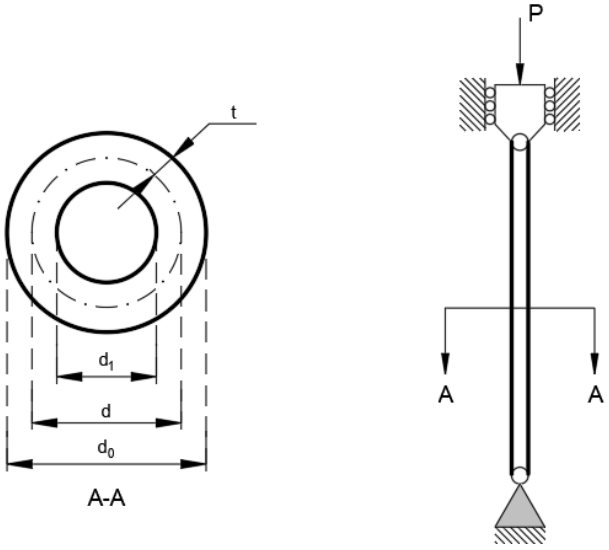


Figure 4. Schematic representation of tubular column design problem

The Welded Beam design problem is one of the EDOPs introduced by Coello [29]. A vertical force loads the beam, as shown in Figure 5. The task is to find the best possible design for the welded beam

concerning minimizing manufacturing cost under seven constraints regarding stress, deflection, welding, and geometry. The design variables are h (x_1), l (x_2), t (x_3), and b (x_4). The constants are $P = 6000\ lb,\ L = 14\ in\ lb,\ \delta_{max} = 0.25\ in,\ E = 30 \times 10^6\ psi,\quad G = 12 \times 10^6\ p,\quad \tau_{max} = 13{,}600\ psi,$ $\sigma_{max} = 30{,}000\ psi$. This objective function can mathematically be given as [29]:

Minimize:

$$f(X) = 1.10471x_1^2x_2 + 0.04811x_3x_4(14.0 + x_2) \quad (25)$$

Subject to:

$$g_1(X) = \tau(X) - \tau_{max} \le 0 \quad (26)$$
$$g_2(X) = \sigma(X) - \sigma_{max} \le 0 \quad (27)$$
$$g_3(X) = \delta(X) - \delta_{max} \le 0 \quad (28)$$
$$g_4(X) = x_1 - x_4 \le 0 \quad (29)$$
$$g_5(X) = P - P_c(X) \le 0 \quad (30)$$
$$g_6(X) = 0.125 - x_1 \le 0 \quad (31)$$
$$g_7(X) = 0.10471x_1^2 + 0.04811x_3x_4(14.0 + x_2) - 5.0 \le 0 \quad (32)$$
$$\tau(X) = \sqrt{(\tau')^2 + 2\tau'\tau''\frac{x_2}{2R} + (\tau'')^2} \quad (33)$$
$$\tau' = \frac{P}{\sqrt{2}x_1x_2} \quad (34)$$
$$\tau'' = \frac{MR}{J} \quad (35)$$
$$M = P\left(L + \frac{x_2}{2}\right) \quad (36)$$
$$R = \sqrt{\frac{x_2^2}{4} + \left(\frac{x_1 + x_3}{2}\right)^2} \quad (37)$$
$$J = 2\left\{\sqrt{2}x_1x_2\left[\frac{x_2^2}{12} + \left(\frac{x_1 + x_3}{2}\right)^2\right]\right\} \quad (38)$$
$$\sigma(X) = \frac{6PL}{x_4x_3^2} \quad (39)$$
$$\delta(X) = \frac{4PL^3}{Ex_3^3x_4} \quad (40)$$
$$P_c(X) = \frac{4.013E\sqrt{x_3^2x_4^6/36}}{L^2}\left(1 - \frac{x_3}{2L}\sqrt{\frac{E}{4G}}\right) \quad (41)$$

Variable Range:

$$0.1 \le x_1 \le 2 \quad (42)$$
$$0.1 \le x_2 \le 10 \quad (43)$$
$$0.1 \le x_3 \le 10 \quad (44)$$
$$0.1 \le x_4 \le 2 \quad (45)$$

Amir and Hasegawa [30] formulated an EDOP for designing a reinforced concrete beam, presented in Figure 6. The beam is simply supported with a span equal to 30 feet and subjected to a live loading of 2000 lbs and a dead loading of 1000 lbs, which includes the weight of the beam. The strength of the concrete is σ_c = 5 ksi and the strength of the reinforcing steel is equal to σ_y = 50 ksi. The cost for concrete is \$0.02/in²/ft, while the price for steel is \$1.0/in²/ft. The subject of EDOP is determined to satisfy the structural requirements of the ACI building code 318-77, in which the area of reinforcement A_s (x_1), the width of the beam b (x_2), and the depth of the beam h (x_3) are the design variables such that the total cost of the structure can be minimized.

Minimize:

$$f(X) = 2.9x_1 + 0.6x_2x_3 \quad (46)$$

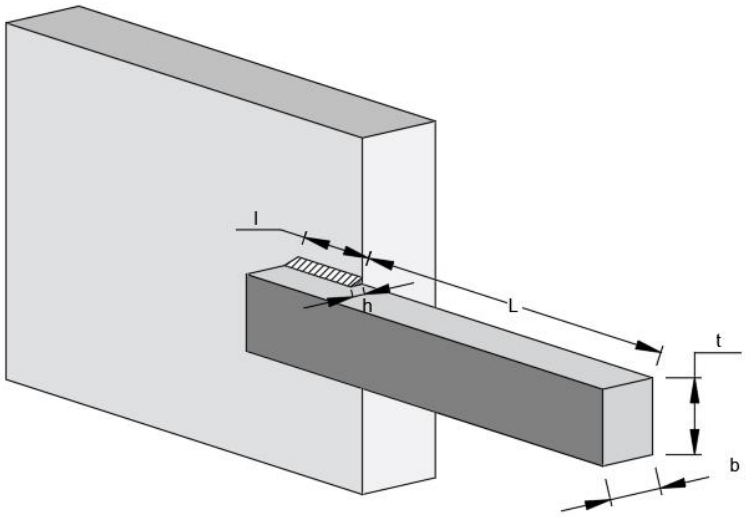


Figure 5. Schematic representation of welded beam design problem

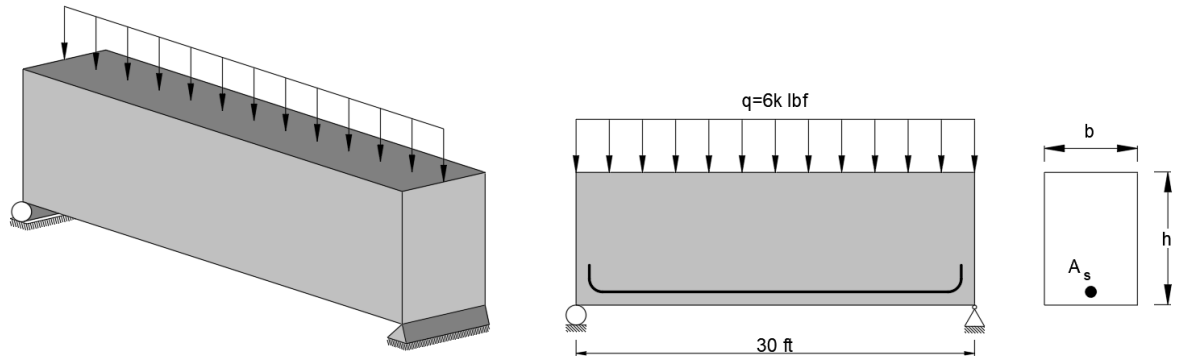


Figure 6. Schematic representation of reinforced concrete beam design problem

Subject to:

$$g_1(X) = \frac{x_2}{x_3} - 4 \leq 0 \quad (47)$$

$$g_2(X) = 180 + 7.375\frac{x_1^2}{x_3} - x_1 x_2 \leq 0 \quad (48)$$

Variable Range:

$$x_1 \in \{6,6.16,6.32,6.6,7,7.11,7.2,7.8,7.9,8,8.4\} \quad (49)$$

$$x_2 \in \{28,29,30,\ldots,40\} \quad (50)$$

$$5 \leq x_3 \leq 105 \quad (51)$$

The corrugated bulkhead EDOP shown in Figure 7, concerns the minimization of a bulkhead weight in a chemical tanker with a corrugated bulkhead, where its design variables are the width (x_1), depth (x_2), length (x_3), and plate thickness (x_4) of the bulkhead. The mathematical model for this optimization problem is described as follows [31]:

Minimize:

$$f(X) = \frac{5.885 x_4 (x_1 + x_3)}{x_1 + \sqrt{|x_3^2 - x_2^2|}} \quad (52)$$

Subject to:

$$g_1(X) = -x_4 x_2 \left(0.4x_1 + \frac{x_3}{6}\right) + 8.94\left(x_1 + \sqrt{|x_3^2 - x_2^2|}\right) \leq 0 \quad (53)$$

$$g_2(X) = -x_4 x_2^2 \left(0.2x_1 + \frac{x_3}{12}\right) + 2.2\left(8.94\left(x_1 + \sqrt{|x_3^2 - x_2^2|}\right)\right)^{4/3} \leq 0 \quad (54)$$

$$g_3(X) = -x_4 + 0.0156x_1 + 0.15 \leq 0 \quad (55)$$

$$g_4(X) = -x_4 + 0.0156x_3 + 0.15 \leq 0 \quad (56)$$

$$g_5(X) = -x_4 + 1.05 \leq 0 \quad (57)$$

$$g_6(X) = -x_3 + x_2 \leq 0 \quad (58)$$

Variable Range:

$$0 \leq x_1, x_2, x_3 \leq 100 \quad (59)$$

$$0 \leq x_4 \leq 5 \quad (60)$$

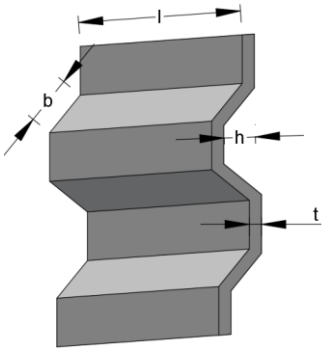


Figure 7. Schematic representation of corrugated bulkhead design problem

The tension/compression spring design problem introduced in [32] is aimed at weight minimization of a tension/compression spring, as shown in Figure 8. The design problem consists of constraints on minimum deflection, shear stress, surge frequency, outside diameter limits, and design variables, which are the wire diameter d (x_1), the mean coil diameter D (x_2), and the number of active coils N (x_3). The mathematical formulation is written as follows:

Minimize:

$$f(X) = (x_3 + 2)x_2 x_1^2 \quad (61)$$

Subject to:

$$g_1(X) = 1 - \frac{x_2^3 x_3}{71785 x_1^4} \leq 0 \quad (62)$$

$$g_2(X) = \frac{4x_2^2 - x_1 x_2}{12566(x_2 x_1^3 - x_1^4)} + \frac{1}{5108x_1^2} \leq 0 \quad (63)$$

$$g_3(X) = 1 - \frac{140.45 x_1}{71785 x_1^4} \leq 0 \quad (64)$$

$$g_4(X) = \frac{x_1 + x_2}{1.5} - 1 \leq 0 \quad (65)$$

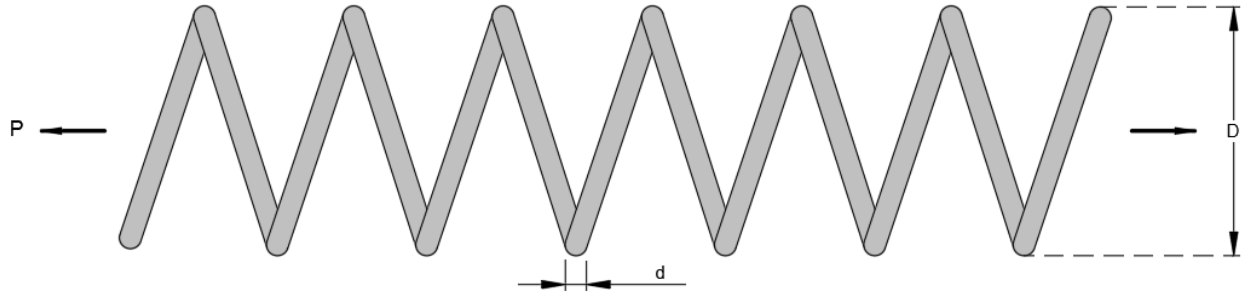


Figure 8. Schematic representation of spring design problem

Variable Range:

$$0.05 \leq x_1 \leq 2 \quad (66)$$
$$0.25 \leq x_2 \leq 1.3 \quad (67)$$
$$2 \leq x_3 \leq 15 \quad (68)$$

The optimization of the design of a cylindrical pressure vessel contained between two hemispherical heads is concerning minimum total cost that is comprising material, forming, and welding costs plotted in Figure 9. The problem consists of four design variables the thickness of the shell T_s (x_1), the thickness of the head T_h (x_2), the inner radius R (x_3), and the length of the cylindrical section of the vessel, excluding the head L (x_4). Additionally, x_1 and x_2 are integer multiples of 0.0625 inches, and other variables are continuous. The resulting optimization problem can be cast as follows [33]:

Minimize:

$$f(X) = 0.6224x_1x_3x_4 + 1.7781x_2x_3^2 + 3.1661x_1^2x_4 + 19.84x_1^2x_3 \quad (69)$$

Subject to:

$$g_1(X) = -x_1 + 0.0193x_3 \leq 0 \quad (70)$$
$$g_2(X) = -x_2 + 0.00954x_3 \leq 0 \quad (71)$$
$$g_3(X) = -\pi x_3^2x_4 + \frac{4}{3}\pi x_3^3 + 1{,}296{,}000 \leq 0 \quad (72)$$
$$g_4(X) = x_4 - 240 \leq 0 \quad (73)$$

Variable Range:

$$x_1, x_2 \in \{1 \times 0.0625, 2 \times 0.0625, 3 \times 0.0625, \ldots, 99 \times 0.0625\} \quad (74)$$
$$0 \leq x_3, x_4 \leq 200 \quad (75)$$

In mechanical systems, the speed reducer is a crucial component of the gearbox and is utilized in various applications [33]. This optimization problem which is illustrated in Figure 10 aims to minimize the weight of the speed reducer, subject to eleven constraints. The problem involves seven design variables: face width b (x_1), module of teeth m (x_2), number of teeth in the pinion z (x_3) length of the first shaft between bearings l_1 (x_4), length of the second shaft between bearings l_2 (x_5), diameter of the first shaft d_1 (x_6), and diameter of the second shaft d_2 (x_7). The mathematical formulation of this problem is as follows:

Minimize:

$$f(X) = 0.7854x_1x_2^2(3.3333x_3^2 + 14.9334x_3 - 43.0934) - 1.508x_1(x_6^2 + x_7^2) + 7.4777(x_6^3 + x_7^3) + 0.7854(x_4x_6^2 + x_5x_7^2) \quad (76)$$

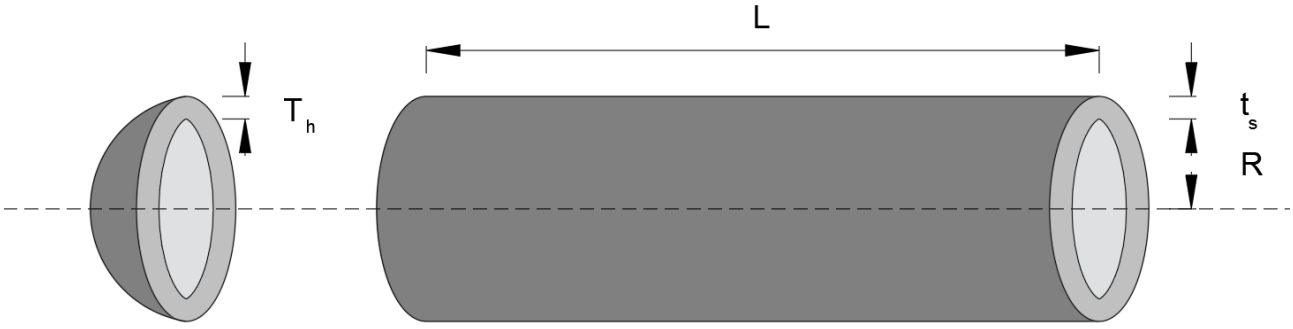


Figure 9. Schematic representation of pressure vessel design problem

Subject to:

$$g_1(X) = \frac{27}{x_1 x_2^2 x_3} - 1 \leq 0 \quad (77)$$

$$g_2(X) = \frac{397.5}{x_1 x_2^2 x_3^2} - 1 \leq 0 \quad (78)$$

$$g_3(X) = \frac{1.93 x_4^3}{x_2 x_6^4 x_3} - 1 \leq 0 \quad (79)$$

$$g_4(X) = \frac{1.93 x_5^2}{x_2 x_7^4 x_3} - 1 \leq 0 \quad (80)$$

$$g_5(X) = \frac{\sqrt{(745 x_4 / x_2 x_3)^2 + 16.9 \times 10^6}}{110 x_6^3} - 1 \leq 0 \quad (81)$$

$$g_6(X) = \frac{\sqrt{(745 x_5 / x_2 x_3)^2 + 157.5 \times 10^6}}{85 x_7^3} - 1 \leq 0 \quad (82)$$

$$g_7(X) = \frac{x_2 x_3}{40} - 1 \leq 0 \quad (83)$$

$$g_8(X) = \frac{5 x_2}{x_1} - 1 \leq 0 \quad (84)$$

$$g_9(X) = \frac{x_1}{12 x_2} - 1 \leq 0 \quad (85)$$

$$g_{10}(X) = \frac{1.5 x_6 + 1.9}{x_4} - 1 \leq 0 \quad (86)$$

$$g_{11}(X) = \frac{1.1 x_7 + 1.9}{x_5} - 1 \leq 0 \quad (87)$$

Variable Range:

$$2.6 \leq x_1 \leq 3.6 \quad (88)$$

$$0.7 \leq x_2 \leq 0.8 \quad (89)$$

$$17 \leq x_3 \leq 28 \quad (90)$$

$$7.3 \leq x_4 , x_5 \leq 8.3 \quad (91)$$

$$2.9 \leq x_6 \leq 3.9 \quad (92)$$

$$5 \leq x_7 \leq 5.5 \quad (93)$$

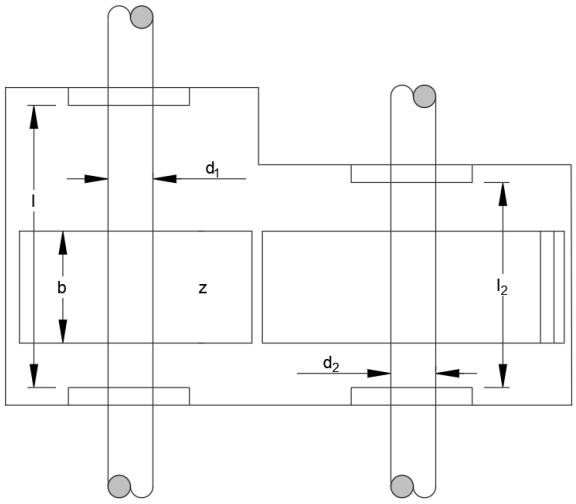


Figure 10. Schematic representation of speed reducer design problem

Minimize:

$$f(X) = 1.98 + 4.90 x_1 + 6.67 x_2 + 6.98 x_3 + 4.01 x_4 + 1.78 x_5 + 2.73 x_7 \quad (96)$$

Subject to:

$$g_1(X) = 1.16 - 0.3717 x_2 x_4 - 0.00931 x_2 x_{10} - 0.484 x_3 x_9 + 0.01343 x_6 x_{10} - 1 \leq 0 \quad (97)$$

The gear train design problem, introduced by Sandgren [34], is an unconstrained discrete EDOP in mechanical engineering. The objective of this problem is to minimize the gear ratio, defined as the ratio of the angular velocity of the output shaft to the angular velocity of the input shaft. The design variables are the number of teeth of gears n_A (x_1), n_B (x_2), n_C (x_3), and n_D (x_4). Figure 11 illustrates the model of this problem. The mathematical formulation is as follows:

Minimize:

$$f(X) = \left(\frac{1}{6.931} - \frac{x_3 x_2}{x_1 x_4}\right)^2 \quad (94)$$

Variable Range:

$$12 \leq x_1, x_2, x_3, x_4 \leq 60 \quad (95)$$

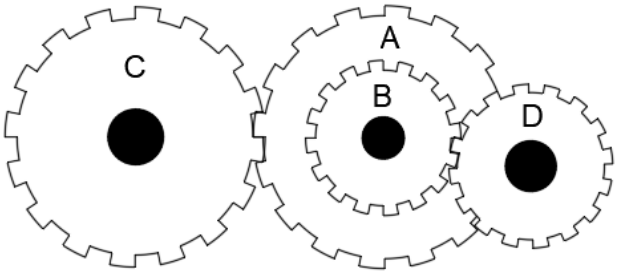


Figure 11. Schematic representation of gear train design problem

In the European Enhanced Vehicle-Safety Committee testing procedure, this EDOP relates to subjecting a car to a side impact to minimize the door's weight. The problem considers eleven decision variables which are the thicknesses of the B-pillar inner (x_1), B-pillar reinforcement (x_2), floor side inner (x_3), cross members (x_4), door beam (x_5), door beltline reinforcement (x_6), roof rail (x_7), materials of the B-pillar inner (x_8), floor side inner (x_9), barrier height (x_{10}), and hitting position (x_{11}). In this EDOP, the formulation that Youn et al. [35] proposed is used as follows:

$$g_2(X) = 46.36 - 9.9x_2 - 12.9x_1x_8 - 0.1107x_3x_{10} - 32 \leq 0 \quad (98)$$

$$g_3(X) = 33.86 + 2.95x_3 + 0.1792x_{10} - 5.057x_1x_2 - 11.0x_2x_8 - 0.0215x_5x_{10} - 9.98x_7x_8 + 22.0x_8x_9 - 32 \leq 0 \quad (99)$$

$$g_4(X) = 28.98 + 3.818x_3 - 4.2x_1x_2 - 0.0207x_5x_{10} + 6.63x_6x_9 - 7.7x_7x_8 + 0.32x_9x_{10} - 32 \leq 0 \quad (100)$$

$$g_5(X) = 0.261 - 0.0159x_1x_2 - 0.188x_1x_8 - 0.019x_2x_7 + 0.0144x_3x_5 + 0.0008757x_5x_{10} + 0.08045x_6x_9 + 0.00139x_8x_{11} + 0.00001575x_{10}x_{11} - 0.32 \leq 0 \quad (101)$$

$$g_6(X) = 0.214 + 0.00817x_5 - 0.131x_1x_8 - 0.0704x_1x_9 + 0.03099x_2x_6 + 0.018x_2x_7 + 0.0208x_3x_8 + 0.121x_3x_9 - 0.00364x_5x_6 + 0.0007715x_5x_{10} - 0.0005354x_6x_{10} + 0.00121x_8x_{11} + 0.00184x_9x_{10} - 0.02x_2^2 - 0.32 \leq 0 \quad (102)$$

$$g_7(X) = 0.74 - 0.61x_2 - 0.163x_3x_8 + 0.001232x_3x_{10} - 0.166x_7x_9 + 0.227x_2^2 - 0.32 \leq 0 \quad (103)$$

2.3. Constraint Handling Mechanism

A parameter free constraint handling approach known as the Inverse-Tangent-Constraint-Handling (ITCH) method is used to address EDOPs in this study. The ITCH technique which is initially developed by Kim et al. [36] assesses each solution vector using Equation (104):

$$f(\vec{x}) = \begin{cases} \hat{g}(\vec{x}) = g_{max}(\vec{x}) & if\ g_{max}(\vec{x}) > 0 \\ \hat{f}(\vec{x}) = atan[f(\vec{x})] - \pi/2 & otherwise \end{cases} \quad (104)$$

Here $g_{max}(\vec{x})$ is defined as max $g_i(\vec{x})$ are the constraint functions. The inverse tangent function is represented by $atan[.]$. It is important to note that $\hat{g}(\vec{x}) < 0$ for any $\vec{x}$, which ensures $\hat{f}(\vec{x}) < \hat{g}(\vec{x})$.

3. Results and Discussion

Seven recent metaheuristic algorithms performance is evaluated through twelve EDOPs in this study. The algorithms are implemented in MATLAB 2024a, and all experiments are performed on a PC with 2.3 GHz CPU and 16 GB RAM running under a Windows 11 operating system. To make a fair comparison, the number of function evaluation is determined as 50000 for all algorithms. The parameters for all compared algorithms from the literature are summarized below in Table 2.

Table 2. Algorithm parameters

Algorithm	Parameter settings
AHA	Migration coefficient M
APO	Number of neighbor pairs np=1, pfmax =0.1, maximum proportion fraction
AVOA	L1=0.8, L2=0.2, w=2.5, P1=0.6, P2=0.4, P3=0.6
EEFO	Parameter Free
MGO	Parameter Free
PKO	Beating factor (BF) =0.8, PEmax=0.5, PEmin=0
QIO	Exploration weight (W1), Exploitation weight (W2)

The best, the worst, mean and standard deviation of the results are reported for each problem in the Table 3-Table 14 respectively. Also, two non-parametric statistical tests (Friedman rank test and Wilcoxon signed rank test) are organized to evaluate algorithms performance further in Table 15.

The results of the cantilever beam problem are given in Table 3. It is determined that the performance of AHA, APO, EEFO, PKO, and QIO algorithms is indifferent since there is a negligible difference between the best values. There exist some variations between the worst values, especially with AVOA and MGO present a little bit higher value. The mean values are very close with AHA and APO in the lowest means. The standard deviations are low, showing that they are very precise and have a very little variability in their performance for this problem.

Table 3. Computational results of cantilever beam problem

Performance Metric	AHA	APO	AVOA	EEFO	MGO	PKO	QIO
Best	**1.339912**	**1.339912**	1.339924	**1.339912**	1.339921	**1.339912**	**1.339912**
Worst	1.339913	1.339913	1.340140	1.339932	1.34027	1.339936	1.339924
Mean	**1.339912**	**1.339912**	1.339966	1.339917	1.340016	1.339923	1.339914
Std. Dev	**3.06E-07**	3.01E-07	4.7E-05	5.58E-06	8.07E-05	6.95E-06	2.38E-06

The best, worst, and mean values for the results of the I-shaped beam problem given in Table 4, are amazingly identical for all algorithms and value of 0.013074. This shows that every one of the algorithms performs under the best and the uniform conditions in this problem. The extremely low standard deviations confirm the high precision and reliability of these algorithms when applied to the I-Shaped Beam Problem.

Table 4. Computational results of I-shaped beam problem

Performance Metric	AHA	APO	AVOA	EEFO	MGO	PKO	QIO
Best	**0.013074**	**0.013074**	**0.013074**	**0.013074**	**0.013074**	**0.013074**	**0.013074**
Worst	0.013074	0.013074	0.013074	0.013074	0.013074	0.013074	0.013074
Mean	0.013074	0.013074	0.013074	0.013074	0.013074	0.013074	0.013074
Std. Dev	1.47E-11	3.68E-13	1.68E-09	2.75E-09	8.06E-09	**2.73E-17**	3.44E-08

In the three bar truss problem given in Table 5, there are slight differences in the best and mean values among the algorithms; in fact, they are all very close to each other. The best and means of AHA, APO, and EEFO are almost the same, which reflects their best performance and consistency. Only the worst values are slightly more in AVOA and MGO, denoting variability. The AHA, APO, EEFO, and QIO standard deviations are very low, meaning very great precision; the AVOA and MGO standards are relatively a bit higher.

Table 5. Computational results of three bar truss problem

Performance Metric	AHA	APO	AVOA	EEFO	MGO	PKO	QIO
Best	**263.8826**	**263.8826**	263.8827	**263.8826**	263.8827	**263.8826**	263.8827
Worst	263.8826	263.8826	263.8995	263.8826	264.0332	263.8828	263.8827
Mean	**263.8826**	**263.8826**	263.8858	**263.8826**	263.8963	263.8827	263.8827
Std. Dev	**5.84E-12**	5.41E-12	0.004709	5.09E-12	0.028167	3.18E-05	1.49E-05

The results of the tubular column design problem given in Table 6, indicate that all algorithms are behaving in quite the same manner, having the best and mean values around 26.53132. The worst values show minimal differences, with QIO showing a slightly higher worst value. The very low standard deviations for all algorithms suggest that the performance of these algorithms is exact and consistent for this design problem.

Table 6. Computational results of tubular column design problem

Performance Metric	AHA	APO	AVOA	EEFO	MGO	PKO	QIO
Best	**26.53132**	**26.53132**	**26.53132**	**26.53132**	**26.53132**	**26.53132**	26.53133
Worst	26.53132	26.53132	26.53133	26.53132	26.53132	26.53132	26.53135
Mean	**26.53132**	**26.53132**	**26.53132**	**26.53132**	**26.53132**	**26.53132**	26.53134
Std. Dev	6.58E-14	6.09E-14	6.7E-07	5.95E-14	9.91E-12	4.78E-14	7.18E-06

For the welded beam design problem given in Table 7, the best values are quite the same amongst the algorithms, although the AVOA attains a slightly higher best value. The worst values present a more significant variation because both are higher for the AVOA and MGO. The mean values are very close, with the lowest means being for the AHA, APO, and EEFO. The standard deviations are low for the AHA, APO, and EEFO, which show high precision, but the AVOA and MGO show more variability in their results.

Table 7. Computational results of welded beam design problem

Performance Metric	AHA	APO	AVOA	EEFO	MGO	PKO	QIO
Best	**1.724717**	**1.724717**	1.724895	**1.724717**	1.724728	**1.724717**	1.72472
Worst	1.724717	1.724717	1.799024	1.724717	2.050502	1.724724	1.72474
Mean	**1.724717**	**1.724717**	1.744075	**1.724717**	1.746229	1.724718	1.724725
Std. Dev	1.19E-12	1.79E-09	0.021077	4.3E-11	0.059604	1.7E-06	4.18E-06

The results of the concrete beam design problem given in Table 8, show that the best and mean values are all clustered around 359.2037. The means only differ at the second decimal for some of the algorithms. The AVOA and MGO have worse values than the rest, which means that their results are more inconsistent. The AHA, APO, EEFO, PKO, and QIO have very low standard deviations, which tell about very high levels of precision. At the same time, the AVOA and MGO are much more variable in their results.

Table 8. Computational results of concrete beam design problem

Performance Metric	AHA	APO	AVOA	EEFO	MGO	PKO	QIO
Best	**359.2037**	**359.2037**	**359.2037**	**359.2037**	**359.2037**	**359.2037**	**359.2037**
Worst	359.2037	362.2494	362.6302	359.2037	362.2494	359.2037	359.2038
Mean	**359.2037**	359.3052	360.5108	**359.2037**	359.3052	**359.2037**	**359.2037**
Std. Dev	1.26E-12	0.556069	1.632746	1.14E-12	0.556069	1.01E-12	2.05E-05

For the problem of corrugated bulkhead design given in Table 9, all the algorithms have the best and mean value that is very close and has negligible variations. The worst values are near to each other with slight differences, with the AVOA and QIO being slightly higher. Standard deviations are extremely low in the AHA, APO, EEFO, and PKO, which means high precision and reliability, whereas the AVOA and QIO have higher variability in their results.

Table 9. Computational results of corrugated bulkhead design problem

Performance Metric	AHA	APO	AVOA	EEFO	MGO	PKO	QIO
Best	**6.842374**	**6.842374**	6.842375	**6.842374**	**6.842374**	**6.842374**	6.842382
Worst	6.842374	6.842374	6.848544	6.842374	6.842374	6.842374	6.842457
Mean	**6.842374**	**6.842374**	6.843027	**6.842374**	**6.842374**	**6.842374**	6.842412
Std. Dev	4.24E-15	1.34E-12	0.001603	5.5E-15	2.84E-11	5.78E-12	1.8E-05

In the spring design problem given Table 10, the best and mean values of all algorithms are close to 0.012662, with a slight difference in means for the AVOA and MGO. The worst values are very different, indicating higher values for the AVOA and MGO. Standard deviations are low for the AHA, APO, EEFO, PKO, and QIO, which presents them as high precisions compared to the other methods; the AVOA and MGO have high variability in their performance.

Table 10. Computational results of spring design problem

Performance Metric	AHA	APO	AVOA	EEFO	MGO	PKO	QIO
Best	**0.012662**	**0.012662**	**0.012662**	**0.012662**	**0.012662**	**0.012662**	**0.012662**
Worst	0.012662	0.012669	0.014593	0.012663	0.016077	0.012715	0.012663
Mean	**0.012662**	**0.012662**	0.013009	**0.012662**	0.013075	0.012677	**0.012662**
Std. Dev	1.17E-07	1.79E-06	0.000466	3.4E-07	0.000849	1.95E-05	3.07E-07

The results of the pressure vessel design problem given in Table 11, indicates that the best values slightly differ among algorithms, while a slight difference is seen; the worst values differ significantly with higher values on the AVOA. The mean varies for all; the AHA and EEFO present the lowest values for it. For the standard deviation, the AHA and EEFO have the lowest values, meaning more precision than the other algorithms. The other algorithms have shown more variability in their results.

Table 11. Computational results of pressure vessel design problem

Performance Metric	AHA	APO	AVOA	EEFO	MGO	PKO	QIO
Best	**6059.083**	**6059.083**	**6059.083**	**6059.083**	6059.084	**6059.083**	6059.09
Worst	6820.026	7332.543	7544.493	6370.352	7333.142	6820.027	6820.08
Mean	6202.341	6389.97	6366.711	**6083.951**	6600.921	6085.477	6107.482
Std. Dev	256.6121	378.3935	434.1332	78.56102	501.2787	138.8485	148.8167

In the speed reducer design problem given in Table 12, the best and mean values are found to be around 2996.274 for most algorithms, with a little difference in the means for the AVOA. Its high worst values reveal a poorer behavior of the AVOA, showing less consistency. The minor standard deviations of the AHA, APO, EEFO, PKO, and QIO reveal good precision and reliability. On the other hand, the AVOA is much more variable in its performance.

Table 12. Computational results of speed reducer design problem

Performance Metric	AHA	APO	AVOA	EEFO	MGO	PKO	QIO
Best	**2996.274**	**2996.274**	2996.275	**2996.274**	**2996.274**	**2996.274**	2996.668
Worst	2996.274	2996.274	3014.038	2996.274	2996.274	2996.274	2997.32
Mean	**2996.274**	**2996.274**	2998.756	**2996.274**	**2996.274**	**2996.274**	2996.954
Std. Dev	5.18E-08	3.06E-10	3.726336	3.44E-08	3.66E-10	2.04E-10	0.163654

The mean and best values present approximately 2.7E-12 for most of the algorithms in the Gear Train Problem given in Table 13, the means are not being wildly divergent for the AVOA. The worst values are considerably higher for the AHA and AVOA, which means that these two methods have a lower level of consistency. The corresponding standard deviations are low for the AHA, APO, EEFO, PKO, and QIO, proving that they have high precision and reliability in their performance.

Table 13. Computational results of gear train problem

Performance Metric	AHA	APO	AVOA	EEFO	MGO	PKO	QIO
Best	**2.7E-12**	**2.7E-12**	**2.7E-12**	**2.7E-12**	**2.7E-12**	**2.7E-12**	**2.7E-12**
Worst	2.36E-09	9.94E-11	2.36E-09	1.18E-09	9.92E-10	9.92E-10	9.92E-10
Mean	3.22E-10	**1.2E-11**	8.15E-10	1.68E-10	1.35E-10	3.62E-10	6.4E-11
Std. Dev	6E-10	1.9E-11	8.16E-10	3.78E-10	2.92E-10	4.33E-10	1.86E-10

The results for the car side impact design problem given in Table 14, show the best and mean values to be around 22.841, with very slight variations for most of the algorithms. The worst values show colossal difference, with higher values in the AVOA. The low values for standard deviations are revealed for the AHA, APO, EEFO, PKO, and QIO, rendering them of the high precision and reliability. In contrast, for the AVOA, the standard deviations are not relatively so low.

Table 14. Computational results of car side impact design problem

Performance Metric	AHA	APO	AVOA	EEFO	MGO	PKO	QIO
Best	22.84151	**22.84138**	22.8432	22.84148	22.84144	22.84139	22.84163
Worst	23.47903	22.84143	24.0513	23.33362	23.30264	22.8463	23.26352
Mean	23.04541	**22.84138**	23.16012	22.90698	23.00187	22.8424	22.89669
Std. Dev	0.18936	8.74E-06	0.29405	0.124723	0.156148	0.001181	0.122991

The statistical comparison given in Table 15, based on the Friedman test ranks the EEFO first, thereby showing its overall the best performance with problems of different types. The Wilcoxon signed-rank test indicates the p-values showing statistical differences between the EEFO and other algorithms. The EEFO statistically performs better than the AVOA and QIO as $p < 0.1$.

Table 15. Statistical test results

Friedman test average rankings		Wilcoxon signed-rank test between EEFO and the other algorithms	
Algorithms	Sum of Ranks	EEFO vs.	p-value
AHA	2.91667 (2)	AHA	0.70020
APO	3.00000 (3)	APO	0.85010
AVOA	6.16667 (7)	AVOA	0.00146
EEFO	2.66667 (1)	MGO	0.01221
MGO	5.41667 (6)	PKO	0.27832
PKO	3.33333 (4)	QIO	0.06396
QIO	4.50000 (5)		

4. Conclusion

In this study, it is presented that a comparative analysis of the performance assessment of seven latest metaheuristic algorithms over twelve EDOPs. The results prove that all algorithms can find close to optimal solutions, but the algorithms significantly vary in terms of consistency and accuracy. The Friedmans test showed that the Electric Eel Foraging Optimization (EEFO) algorithm outperforms all other tested optimization algorithms overall. Furthermore, the Wilcoxon signed-rank test confirms EEFO is significantly better than AVOA and QIO. Although performances vary with the type of problem, this study has concluded that these metaheuristic algorithms have good potential to deal with a wide range of complex engineering optimization problems. Future work will further develop the ability of these algorithms to adapt and solve a larger class of optimization problems and investigate hybrid approaches, which attempt to leverage multiple strengths from various algorithms.

Contributions of the authors

Mümin Emre Şenol Tülin Çetin, and Mustafa Erkan Turan contributed to the study conception and design. Mümin Emre Şenol, Tülin Çetin, and Mustafa Erkan Turan read and approved the draft/final manuscripts.

Acknowledgment

This study was supported by the Manisa Celal Bayar University Scientific Research Projects Coordination Unit. Project Number: 2024-047.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] Z. Zhou, "Memetic algorithm using multiple surrogates for complex engineering design optimization", Doctoral thesis, Nanyang Technological University, Singapore, 2008, https://doi.org/10.32657/10356/13587.

[2] B. Akay and D. Karaboğa, "Artificial bee colony algorithm for large-scale problems and engineering design optimization" *Journal of Intelligent Manufacturing*, 23, pp. 1001-1014, 2012, https://doi.org/10.1007/s10845-010-0393-4.

[3] W. Gong, Z. Cai and D. Liang, "Engineering optimization by means of an improved constrained differential evolution", *Computer Methods in Applied Mechanics and Engineering*, 268, pp. 884-904, 2014, https://doi.org/10.1016/J.CMA.2013.10.019.

[4] M. Omidvar, X. Li and K. Tang, "Designing benchmark problems for large-scale continuous optimization", *Information Sciences*, 316, pp. 419-436, 2015, https://doi.org/10.1016/j.ins.2014.12.062.

[5] X. Lu, S. Menzel, K. Tang and X. Yao, "Cooperative Co-Evolution-Based Design Optimization: A Concurrent Engineering Perspective", *IEEE Transactions on Evolutionary Computation*, 22, pp. 173-188, 2018, https://doi.org/10.1109/TEVC.2017.2713949.

[6] A. Mohamed, "A novel differential evolution algorithm for solving constrained engineering optimization problems", *Journal of Intelligent Manufacturing*, vol. 29, pp. 659-692, 2018, https://doi.org/10.1007/s10845-017-1294-6.

[7] G. Li, F. Shuang, P. Zhao and C. Le, "An Improved Butterfly Optimization Algorithm for Engineering Design Problems Using the Cross-Entropy Method", *Symmetry*, vol. 11, no. 1049, 2019, https://doi.org/10.3390/sym11081049.

[8] A. Shabani, B. Asgarian, M. Salido and S. Gharebaghi, "Search and rescue optimization algorithm: A new optimization method for solving constrained engineering optimization problems", *Expert Systems with Applications*, 161, 113698, 2020, https://doi.org/10.1016/j.eswa.2020.113698.

[9] M. Azizi, S. Talatahari and A. Giaralis, "Optimization of Engineering Design Problems Using Atomic Orbital Search Algorithm", *IEEE Access*, 9, pp.102497-102519, 2021, https://doi.org/10.1109/ACCESS.2021.3096726.

[10] M. Hijjawi, M. Alshinwan, O. Khashan, M. Alshdaifat, W. Almanaseer, W. Alomoush, H. Garg, and L. Abualigah, "Accelerated Arithmetic Optimization Algorithm by Cuckoo Search for Solving Engineering Design Problems", *Processes* 11, no.5, 1380, 2023, https://doi.org/10.3390/pr11051380.

[11] W. Zhao, L. Wang and S. Mirjalili, "Artificial hummingbird algorithm: A new bio-inspired optimizer with its engineering applications", *Computer Methods in Applied Mechanics and Engineering*, vol. 388, no. 114194, 2022, https://doi.org/10.1016/j.cma.2021.114194.

[12] W. Zhao, "Artificial Hummingbird Algorithm", 2024, [Online] Available: https://www.mathworks.com/matlabcentral/fileexchange/101133-artificial-hummingbird-algorithm, MATLAB Central File Exchange. [Accessed: May, 9, 2024].

[13] X. Wang, V. Snášel, S. Mirjalili, J.-S. Pan, L. Kong, and A. S. Hisham, "Artificial Protozoa Optimizer (APO): A novel bio-inspired metaheuristic algorithm for engineering optimization", *Knowledge-Based Systems*, Volume 295, 111737, 2024, https://doi.org/10.1016/j.knosys.2024.111737.

[14] W. Xiaopeng, "Artificial Protozoa Optimizer", [Online] Available: https://www.mathworks.com/matlabcentral/fileexchange/162656-artificial-protozoa-optimizer , MATLAB Central File Exchange. [Accessed: May, 9, 2024].

[15] B. Abdollahzadeh, F. S. Gharehchopogh, and S. Mirjalili, "African vultures optimization algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems", *Computers & Industrial Engineering*, Volume 158, 107408, 2021, https://doi.org/10.1016/j.cie.2021.107408.

[16] B. Abdollahzadeh, F. S. Gharehchopogh, and S. Mirjalili, "Artificial Protozoa Optimizer", [Online] Available: https://www.mathworks.com/matlabcentral/fileexchange/94820-african-vultures-optimization-algorithm, MATLAB Central File Exchange. [Accessed: May, 9, 2024].

[17] W. Zhao, L. Wang, Z. Zhang, H. Fan, J. Zhang, S. Mirjalili, N. Khodadadi and Q. Cao, "Electric eel foraging optimization: A new bio-inspired optimizer for engineering applications", *Expert Systems with Applications*, 238, 122200, 2024, https://doi.org/10.1016/j.eswa.2023.122200.

[18] W. Zhao, L. Wang, Z. Zhang, H. Fan, J. Zhang, S. Mirjalili, N. Khodadadi and Q. Cao, "Electric Eel Foraging Optimization", [Online] Available: https://ww2.mathworks.cn/matlabcentral/fileexchange/153461-electric-eel-foraging-optimization-eefo, MATLAB Central File Exchange. [Accessed: May, 9, 2024].

[19] B. Abdollahzadeh, F. S. Gharehchopogh, N. Khodadadi and S. Mirjalili, "Mountain Gazelle Optimizer: A new Nature-inspired Metaheuristic Algorithm for Global Optimization Problems", Advances in Engineering Software, Volume 174, 103282, 2022, https://doi.org/10.1016/j.advengsoft.2022.103282.

[20] W B. Abdollahzadeh, "Mountain Gazelle Optimizer", [Online] Available: https://www.mathworks.com/matlabcentral/fileexchange/118680-mountain-gazelle-optimizer, MATLAB Central File Exchange. [Accessed: May, 9, 2024].

[21] A. Bouaouda, F. A. Hashim, Y. Sayouti and A. G. Hussien, "Pied kingfisher optimizer: a new bio-inspired algorithm for solving numerical optimization and industrial engineering problems", *Neural Computing and Applications*, 2024, https://doi.org/10.1007/s00521-024-09879-5

[22] A. Hussien, "Pied Kingfisher Optimizer (PKO) [Online] Available: (https://www.mathworks.com/matlabcentral/fileexchange/160043-pied-kingfisher-optimizer-pko), MATLAB Central File Exchange. [Accessed: May, 9, 2024].

[23] W. Zhao, L. Wang, Z. Zhang, S. Mirjalili, N. Khodadadi and Q. Ge, "Quadratic Interpolation Optimization (QIO): A new optimization algorithm based on generalized quadratic interpolation and its applications to real-world engineering problems", *Computer Methods in Applied Mechanics and Engineering*, 116446, 2023, https://doi.org/10.1016/j.cma.2023.116446.

[24] W. Zhao, "Quadratic Interpolation Optimization (QIO) [Online] Available: (https://ww2.mathworks.cn/matlabcentral/fileexchange/135627-quadratic-interpolation-optimization-qio), MATLAB Central File Exchange. [Accessed: May, 9, 2024].

[25] H. Chickermane and H. C. Gea, "Structural optimization using a new local approximation method", *International Journal for Numerical Methods in Engineering*, vol. 39, no. 5, pp. 829–846, 1996.

[26] S. Gold and S. Krishnamurty, "Trade-offs in Robust Engineering Design", *Proceedings of the 1997 ASME Design Engineering Technical Conferences*, DETC97/DAC3757, September 14-17, Saramento, California, 1997.

[27] T. Ray and P. Saini, "Engineering design optimization using a swarm with an intelligent information sharing among individuals", *Engineering Optimization*, vol. 33, no. 6, pp. 735-748, 2001.

[28] S. S. Rao, *Engineering optimization: Theory and practice*, 5th ed. Nashville, TN: John Wiley & Sons, 2019.

[29] C. A. C. Coello, "Use of a self-adaptive penalty approach for engineering optimization problems", *Computers in Industry*, vol. 41, no. 2, pp. 113-127, 2000.

[30] H. M. Amir and T. Hasegawa, "Nonlinear mixed-discrete structural optimization"," *Journal of Structural Engineering"*, vol. 115, no. 3, pp. 626–646, 1989

[31] A. H. Gandomi, X. S. Yang and A. H. Alavi, "Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems", *Engineering with Computers*, 29, pp. 17-35, 2013.

[32] J. S. Arora, *Introduction to Optimum Design*, 2nd ed. San Diego, CA: Academic Press, 2004.

[33] D. Sattar and R. Salim, "A smart metaheuristic algorithm for solving engineering problems," *Eng. Comput.*, vol. 37, no. 3, pp. 2389–2417, 2021.

[34] E. Sandgren, "Nonlinear integer and discrete programming in mechanical design optimization," *J. Mech. Des. N. Y.*, vol. 112, no. 2, pp. 223–229, 1990. https://doi.org/10.1115/1.2912596

[35] B. D. Youn and K. K. Choi, "A new response surface methodology for reliability-based design optimization," *Comput. Struct.*, vol. 82, no. 2–3, pp. 241–256, 2004.

[36] T.-H. Kim, I. Maruta, and T. Sugie, "A simple and efficient constrained particle swarm optimization and its application to engineering design problems," *Proc Inst Mech Eng Part C*, vol. 224, no. 2, pp. 389–400, 2010.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1099-1110 YEAR: 2024
DOI: 10.17798/bitlisfen.1517262

Green Synthesis of Cosmetic Soaps Obtained from Laurel Nobilis

Hatice KARAER YAĞMUR[1*], Kübra BAYKARA[1], Seyithan SÖNMEZ[1]

[1] *Dicle University, Faculty of Science, Department of Chemistry, Diyarbakir, Türkiye*

(ORCID: 0000-0002-3728-1825) (ORCID: 0009-0005-7889-634X) (ORCID: 0009-0003-1249-7640)

Keywords: Laurel nobilis oil, activated carbon, soap, saponification value, peroxide value.

Abstract

In this study, Laurel nobilis oil was obtained from Laurel nobilis fruits. Qualitative analysis of Laurel fixed oil was performed by GC-MS. Additionally, saponification value (SV), acid value (AV), free fatty acid value (% FFA) and peroxide value (PV) values of bay Laurel oil were determined as 140.0 mg KOH/g, 26.5 mg KOH/g, 13.25 mg KOH/g, and 10.28 meq O_2/kg, respectively. Activated carbon was obtained as a result of carbonization of the Laurel fruit seeds, which were waste as a result of oil extraction, at 500 ^oC. Activated carbon was characterized by FT-IR, BET, SEM and TGA. Soap is made of different fatty acid salts, either sodium or potassium. The obtained Laurel fruit oil and activated carbon were used in the preparation of solid soaps with peeling effect by the cold process method. pH, foam test, total alkaline test, and total fatty matter determination (TFM) of the prepared soaps were determined. Thus, the synthesis of activated carbon with peeling effect, which has good cleaning and foaming properties, was carried out.

1. Introduction

Natural cosmetics with bioactive phytochemical substances have a significant pharmacological impact and aesthetic value while posing less of a risk to the environment and users. In addition to other negative effects, many artificial and synthetic ingredients in skin and hair cosmetics typically cause dry skin, irritation, and damaged skin. The primary byproduct of the chemical interaction between lye solution (sodium hydroxide) and triglyceride, or fixed oil from seeds, is soap. The method is known as saponification (Scheme 1) [1]. Soaps are chemical mixtures of Na^+ or K^+ ions with fatty acids. The components of fats and oils utilized in the production of soap are called fatty acids. They are divided into two groups: saturated and unsaturated. However, the most prevalent unsaturated fatty acids are oleic and linoleic acids, while the most prevalent saturated fatty acids are palmitic and stearic acids [2]. The types of oils used, the degree of saponification, the age of the soaps, and the strength and purity of the alkali all affect the chemical properties of soaps. Total fatty acids (TFM), pH, free alkali, and foam stability are examples of these physico-chemical properties [3].

Laurel nobilis L. (Lauraceae), also referred to as sweet bay, bay laurel, roman laurel, or real laurel, is an evergreen tree that may grow up to 10 meters in height. In Mediterranean countries including Turkey, Algeria, Morocco, Portugal, Spain, and Italy, bay laurel is farmed economically for its aromatic leaves and oils [4]. In addition to its special aroma, it is utilized globally as a medicine. Certain elements of this plant, like its organic acids and essential oils, have demonstrated potent antibacterial properties [5]. The industrially significant plant L. nobilis is utilized in medications, food products, and cosmetics. The food sector uses a lot of dried leaves and essential oils to season meat items, soups, and fish [6]. Of the fixed oil in bay laurel trees (Figure 1), 24–30% is found in their fruits (Figure 1). Typically, fixed oil made from the

*Corresponding author: profhatice23@hotmail.com
Received: 16.07.2024, Accepted: 14.10.2024

fruit's kernel has a high lauryl acid content. About 70% of the entire fruit is made up of the fruit's kernel. [4]. The fruits' volatile and fixed oils are mostly utilized in the production of soap. Laurel nobilis L. essential oil is widely produced and used at home in Hatay (a province of Turkey). Laurel nobilis L. Trees (Figure 1) can be found growing wild in Antakya, Defne, Yayladagi, and Samandagi (districts of Hatay). There are differences in the climate and geography of these three towns.

Natural soaps are becoming more and more popular as consumers start to explore alternatives to commercial soaps with all of their ingredients. Therefore, the tendency towards activated carbon soaps has increased. The phrase "activated carbon" refers to a group of highly porous carbonaceous compounds. Because of their large surface area, activated carbons are powerful adsorbents. Due to their lower cost and renewable nature, agricultural wastes are regarded as a very important feedstock for the synthesis of activated carbon [7].

Because of its high surface area and pores, activated charcoal has gained popularity as an ingredient in cosmetic products. It can pull bacteria and pollutants from the skin by absorbing substances through these pores. This makes it possible for it to absorb chemicals and pollutants off the skin's surface, allowing it to be rinsed or peeled off to remove any clogged pores and detoxify the skin. Activated charcoal, which has the function of adsorbing oil, dark spots and pollutants that stick to our skin, is used as an additive in cosmetic products. It is utilized in a variety of beauty products, including face masks, cleansers, and even soaps, because of its adsorbing qualities [8].

In our country, bay leaves are generally used as spices in food. In the literature, bay leaves and activated carbon produced from these leaves have been used in adsorption studies. Bay oil obtained from bay fruits is generally used in soap making. As a result of Laurel oil production, tons of Laurel fruit seeds are in waste condition and do not contribute to our country's economy. On the contrary, they cause environmental pollution. Activated carbon is an expensive product. In this study, it was aimed to obtain oil from Laurel fruits and to produce activated carbon from waste seeds. The obtained activated carbon was used as an additive in the production of Laurel soap. Thus, it was aimed at produce a high value cosmetic product from a waste material. No study was found in the literature on obtaining such a product from bay laurel fruits for cosmetic purposes. It is thought that this study will contribute to the literature in this respect.

2. Material and Method

2.1. Materials

Laurel nobilis fruits were supplied from the Defne district of Hatay (in November 2023). Sodium hydroxide (NaOH), potassium hydroxide (KOH), sulfuric acid (H_2SO_4, %98), phenolphthalein ($C_{20}H_{14}O_4$, 3,3-bis(4-hydroxyphenyl)-1(3H)-isobenzofuranone) and hydrochloric acid (HCl, %37.5) were supplied from Merck.

2.2. Obtaining Laurel nobilis oil from Laurel nobilis fruits

Laurel nobilis fruit oil was obtained according to the literature [9]. Hot water flotation is a traditional method for extracting Laurel oil in Hatay. First of all, bay fruits were washed with pure water to remove dust. 200 g of L. nobilis fruit was taken into a 500 mL beaker. Distilled water was added to cover it. The mixtures were boiled at 80-90°C for 5 hours. After cooling, the upper oil layer was separated and extracted with hexane. Approximately 100 mL of oil was obtained from approximately half a kilo of bay fruit. L. nobilis oil was weighed and stored at 4°C for use. The Laurel oil obtained was characterized by GC-MS (Figure 4).

Figure 1. Laurel nobilis tree and Laurel nobilis fruit.

2.3. Determination of physicochemical parameters of Laurel oil

2.3.1.Saponification value of oil (SV)

2 g of the oil sample was added to a vial containing 30 mL of ethanolic KOH/NaOH and then refluxed for 30 minutes to ensure complete dissolution of the sample. After the sample cooled, 1 mL phenolphthalein was added and titrated with 0.5 M HCl until it gave a pink color [1], [2].

$$SV = \frac{56.1*M(V2-V1)}{W} = \frac{mgKOH}{g} \; or \; \frac{mgNaOH}{g} \qquad (1)$$

where W=weight of sample(g);V1= Volume of hydrochloric acid used in test; V2= Volume of hydrochloric acid used in the blank; M=Molarity of hydrochloric acid; Molecular weight of KOH (g/mole) = 56.1

2.3.2. Acid value (A.V.)

1 g of oil was taken into a conical flask containing 25 mL of methanol, and 3 drops of phenolphthalein indicator were added. The mixture was heated in a water bath for 5 minutes and titrated with 0.1 M KOH until a pink color appeared. The acid value was calculated using the following equation [1].

$$AV = \frac{VKOH*M*56.1\left(\frac{g}{mol}\right)}{W} = mg\frac{KOH}{g} \qquad (2)$$

M:molarity of KOH, W: weıght of sample (g)

2.3.3.Determination of Free Fatty Acids (FFA)

The FFA value of the oil was calculated according to the literature [10].

$$Free\ Fatty\ Acids\ (FFA) = Acid\ Value\ (AV)/2 \qquad (3)$$

FFA of Laurel nobilis oil: 13.25 mg KOH/g

2.3.4. Peroxide value (PV)

The PV value of the oil was calculated according to the literature [1].

2.4. Synthesis of activated carbon from Laurel nobilis fruit seeds

L. nobilis fruits were used as a starting material in experimental studies to produce activated carbon. After obtaining bay oil from bay fruits by hydroflotation, bay fruit seeds were separated from the pulp, washed with distilled water and dried in an oven at 105 ^{o}C for 48 hours. After the drying process, bay fruit seeds were ground in a grinder and stored in a desiccator to be used. The carbonization process is carried out by placing the 7 cm diameter and 100 cm long stainless steel tube reactor in the furnace (Protherm PTF 12/20/400) and passing nitrogen (N_2) gas at a flow rate of 150 mL/min for 1 hour at 500 ^{o}C at a speed of 15 ^{o}C/min. Heating was provided. After the carbonized sample was cooled to room conditions under a nitrogen atmosphere, HCl (0.1 M) was refluxed for 1 hour to purify the sample from impurities and reduce the ash content. Then, the sample was filtered and washed with hot distilled water until it did not react with chloride. A Chloride test (1%) was performed with $AgNO_3$ [11]. Experimental stages of obtaining Laurel oil and activated carbon from Laurel fruits are shown in Figure 2. The resulting activated carbon was coded as AC (Figure 2).

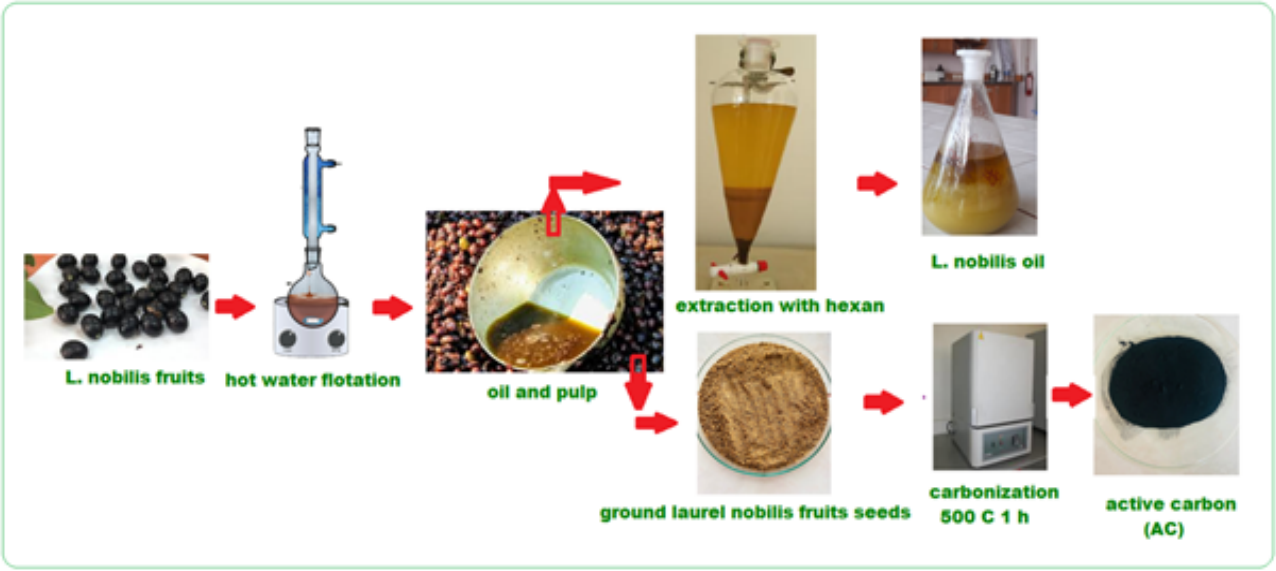


Figure 2. Obtaining bay oil and activated carbon from bay fruits

2.5.Characterization

The fatty acid composition of the Laurel nobilis was analyzed using a Rtx-5MS (30 m x 0.25 mm inner diameter x 0.25 µm film thickness) column on a Shimadzu GC-MS/QP2010. Carrier gas is helium (2.32 mL/min.), split 1. Ion source temperature was applied as 300 ºC for qualitative analyses. It was diluted with methanol for qualitative analysis of the oil. The Brunauer-Emmett-Teller (BET) surface analyzer (a TriStar 3000 surface analyzer.) was used to measure the surface area and pore size distribution of the AC at a temperature of -196 °C and N_2 gas. Conversely, using an FTIR spectrometer (Model 2000, Perkin-Elmer, USA), Fourier transform infrared (FTIR) spectra were acquired throughout the range of 4000-400 cm^{-1}. TG measurement of activated carbon was performed between 20 and 1000 °C (200 mL min^{-1}, in N_2 atmosphere 10 °C min^{-1}, Perkin Elmer Diamond Thermal Analysis).

2.6. Preparation of Soap

One of the most significant components of fats, fatty acids (FAs) and free fatty acids (FFAs) are non-volatile aliphatic monocarboxylic molecules that can also be used as a soap base. One of the most often used for personal hygiene is soap. It is created when FA triglycerides undergo the saponification reaction (hydrolysis) with a strong base (often potassium or sodium hydroxides), resulting in the production of glycerol and soap (potassium or sodium salts of FAs). Although the primary ingredient in soap is salts of FAs, it also contains glycerol, FFAs, and unsaponifiable matter, which includes sterols, higher aliphatic alcohols, pigments, hydrocarbons, and other substances. Free-form FFAs also have a significant impact on the quality of soap. FFA-enriched soap has higher fattening and moisturizing qualities, as well as improved foaming. These attributes serve as markers of high-quality soap [12].

Using a cold procedure, the extracted oil was saponified. Soaps were prepared using the soap formulations in Table 1. The saponification value of olive oil and coconut oil was determined according to previous studies [13]. The main ingredient of the formula consists of the following components as a percentage by mass: 200 g of olive oil, 200 g of coconut oil, 200 g of laurel oil and 37% of the total oil amount consists of distilled water. Activated carbon (1-10%) can be used as an additive [14]. In our study, 10% of the total amount of oil was used. In active carbon soap studies, a teaspoon of active carbon was generally used. If the amount of oil used is not too much, using more than a teaspoon of activated carbon will result in gray foam [15]. Figure 3 shows the preparation of the soap. After following conventional technique, a thick, semi-solid mass of soap was obtained and allowed to cool and set for six weeks. The saponified product's pH, foaming capacity, hardness, cleaning efficacy, and total alkalinity were characterized.

Table 1. Formulation of soap

	Saponification Value (SAP)(gr NaOH/gr)	activated carbon free soap	with activated carbon soap
		Amount of oil(g)	Amount of oil(g)
L.nobilisoil	0.140	200 g	200 g
Olive oil	0.135	200 gr	200 gr
Coconut oil	0.183	200 gr	200 gr
Amount of water		%37 (a according to the total amount of fat)	%37 (a according to the total amount of fat)
Activated carbon		-	6 gr
Colorant		Not used	
Soap fragrance		Since Laurel oil is scented, no essence was used.	

2.7. Determination of physiochemical parameters of soaps

2.7.1. Determination of pH of soaps

pH values of soaps were determined using a pH meter according to the literature. 1 g of soap with and without activated carbon was weighed separately. The soaps were dissolved in 10 mL of distilled water. Make up to 100 mL with distilled water to obtain a 1% homogeneous soap solution.

2.7.2. Foaming ability tests

Each of the soaps, 0.2 g, was placed in a 100 mL conical flask, and 10 mL of pure water was added. The mixture was shaken vigorously for 2 minutes to form foam. After shaking the mixture, it was left for 10 minutes. Foam height was measured.

2.7.3.Total Alkaline

100 mL of ethanol and 5 mL of 1 N H_2SO_4 solution were added to 10 g of soap. The mixture was heated until the soap sample was completely dissolved and then titrated with 1 N NaOH using phenolphthalein. Total alkali was obtained by the following formula:

$$\%Total\ alkali = \left(\frac{Vacid-Vbase}{m}\right) x3.1 \quad (4)$$

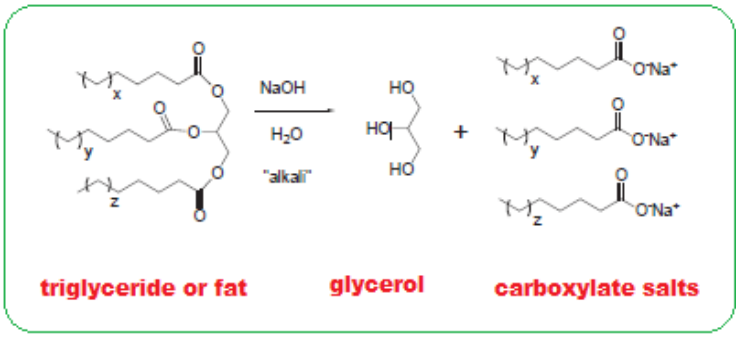
m: mass of soap

2.7.4. Total Fatty Matter Determination (TFM)

10 g of the finished soap was weighed, 150 mL of distilled water was added, and it was heated slightly in a water bath. The soap was dissolved in 20 mL of 15% H_2SO_4 by heating until a clear solution was obtained. The fatty acids on the surface of the resulting solution were solidified by adding 7 g of beeswax and reheated. The apparatus was allowed to cool to form a solid cake-like mold. The oily patch on the surface was removed and allowed to dry for several days and weighed to obtain the total oily matter using the following formula (5):

$$\%\ TFM = \left(\frac{A-X}{w}\right) x100 \quad (5)$$

A: mass of wax+oil
X: mass of wax
W: mass of soap



Scheme 1. Saponification Reaction

Figure 3. Preparation of soap

3. Results and Discussion

Saponification value (SV) can be expressed as a measurement of the alkali-reactive groups in fats and oil or as the milligrams of KOH/NaOH required to saponify one gram of oil. Additionally, the SV shows the oil's average molecular weight. In other words, SV describes the molecular weights of the fatty acids in oil, which are inversely related to the fatty acids' average molecular weights or chain lengths. It has been stated that this is one of the key factors in the industrial methods used to make soap. According to studies, oils with high SVs are good raw materials for the soap industry, but oils with low SVs are not acceptable for use in the soap industry. The saponification value of Laurel nobilis oil was calculated as 0.1925 g KOH/g (192.5 mg KOH/g) or 0.140 g NaOH/g (140.0 mg NaOH/g).

These values are above the specification range of (175 mg KOH/g-187 mg KOH/g) recommended for oils by ASTM (the American Society for Testing and Materials). Vermaak et al., state that high SV indicates the predominance of low molecular weight fatty acids [16]. Even if the SV obtained is greater than the ASTM standard values, the values from this study are lower than those from the most common oils, such as palm oil (200 mg KOH/g), and coconut oil (257 mg KOH/g) reported by Nchimbi [2].

The acid value (AV) of Laurel nobilis oil was calculated as 0.0264 g KOH/g or 26.5 mg KOH/g. For oils and fats, AV is a standard parameter used in specification and quality control. It also serves as a physicochemical property indicator for oil, indicating its age, quality, edibility, and suitability for usage in various sectors. There could be a greater quantity of FFA in the oil, which would explain the increased AVs seen. These raise the possibility of oxidative deterioration, which lowers the material's suitability for industrial uses. Therefore, in order to be suitable for industrial uses such as soap production, the resulting oil must be refined to lower acidity to the prescribed requirements [2]. The acidity of an oil indicates how much lipase action has broken down the glycerides. Heat and light typically speed up the breakdown process. Amino acids, free fatty acids, and acid phosphates are the acids that are often produced. Compared to other forms of acids, free fatty acids are generated more quickly. The oils' low acid levels also suggest that they might be kept in storage for a long period without degrading [17].

FFA of Laurel nobilis oil: 0.01325 g KOH/g (13.25 mg KOH/g). Fatty acids, or FFAs, are created in any stage of the process by hydrolytic reactions from the triglycerides of oils. One of the most often used metrics to evaluate the quality of oil during production, storage, and marketing is its FFA content. It is also used to categorize the oils.

It has been stated that a high percentage of FFAs in oils indicates low quality; this could be the consequence of impurity contamination, which could cause the ester linkage to hydrolyze, raising the percentage of FFAs. On the other hand, low FFA levels indicate high quality and prevent the oil from going rancid or smelling bad, which makes it more suitable for usage in the soap manufacturing sector [2].

The peroxide value (PV) of Laurel nobilis oil was calculated as 10.28 meq O_2/kg. The peroxide value (PV) is one of the most commonly assessed quality metrics during the processing, storage, and marketing of oil. It is typically is an important indicator of the degree of oxidation of fatty acids [18], [19]. The conventional method for determining PV is based on the reaction between potassium iodide and hydroperoxides found in oils, which causes iodide oxidation to produce iodine [20]. The highest level of peroxide value (PV) permitted for edible oils and cold press oils in accordance with Iranian national standards can be divided into four ranges: $0 < PV \leq 1$, $1 < PV \leq 2$, $2 < PV \leq 5$, and $5 < PV \leq 10$ meq/kg for edible oils at the time of production, free at the port of shipment, and consumable limit, in that order. Additionally, the maximum PV levels for refined and cold press oils are $PV < 10$ and $PV \leq 15$ meq/Kg, respectively, according to the Codex standards [21].

The indicator of oil's oxidative rancidity is the peroxide value. When an enzyme or specific chemical substances are present, unsaturated fatty acids' double bonds might become more oxidatively rancid. The release of short chain carboxylic acids is what gives rancidity its characteristic flavor and odor. A greater rate of rancidity is correlated with elevated peroxide levels. The oils' low peroxide values suggest that, at room temperature, oxidative rancidity will not affect them as much [17]. Antioxidants added to oils may result in low peroxide levels. Antioxidants can be added so as to prevent rancidity and increase shelf-life.

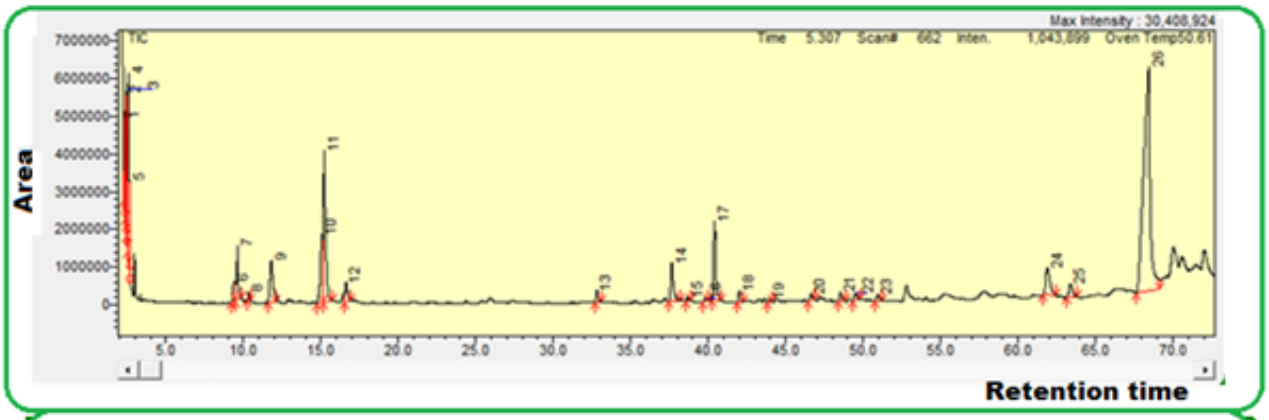
The chromatogram of GC-MS analysis and shows fatty acid compositions of Laurel nobilis oil were given in Figure 4. 26 components of the total oil content, were identified by GC-MS analysis of the hydrodistilled fixed oil of the fruits. 1.8-cineole was identified as the main component. According to the literature, it has been determined that the main component of Laurel oil obtained from Turkey is 1.8-cineole [22].

Soap was characterized according to the literature [1]. The pH values of soaps with and without activated carbon were determined as 8.75 and 8.85, respectively. pH is an important parameter that indicates whether the soap is alkaline or acidic, and soap that is too alkaline can irritate the skin [23]. Literature has shown that soaps will not harm skin or fabric when the standard pH value of soap is between 7 and 10 [22]. Additionally, the natural acidic state of the skin can be neutralized by high alkaline soap [25].

The foam test was performed according to the literature [1]. While it was 5.3 cm for soap with activated carbon, it was measured as 4.8 cm for soap without activated carbon.

Total alkaline value was calculated according to the literature [26]. The % total alkali values were defined as 0.31 and 0.25 for activated carbon and without activated carbon respectively. Total alkaline content is one of the parameters that determine the corrosiveness of any soap [27]. According to the Bureau of Indian Standards (BIS), the alkaline content of quality soaps should be less than 5%, while according to ISO, quality soaps should contain only less than 2% alkali [27, 28].

Total Fatty Matter Determination (TFM) was determined according to the literature [3]. The %TFM values were defined as 73.23 and 72.52 for activated carbon and without activated carbon respectively. A lower TFM value is generally associated with hardness and lower quality. In some European countries, soaps with a minimum of 75% TFM are considered Grade 1. If the TFM value is 65%, this is considered a Grade 2. Soap with a higher TFM gives more suds, lasts longer, and most importantly, cleans your skin better and more gently. Grade 3 quality soap must have at least 60% TFM [28].



Peak#	Ret.Time	Area%	Height%	A/H	Name
1	2.355	2.75	6.76	4.28	Acetic acid, ethyl ester (CAS)
2	2.446	4.32	9.55	4.75	Pentanoic acid, 3-methyl-4-oxo-
3	2.476	3.43	10.22	3.52	Acetic acid, ethyl ester (CAS)
4	2.559	4.87	12.35	4.13	1,2,3-Propanetriol (CAS)
5	2.673	1.39	6.52	2.24	Acetic acid, ethyl ester (CAS)
6	9.386	1.14	1.36	8.80	.alpha.-Pinene
7	9.612	2.61	3.63	7.55	cis-Ocimene
8	10.335	0.23	0.53	4.46	Camphene
9	11.805	3.66	2.78	13.83	2-.BETA.-PINENE
10	15.068	4.04	4.63	9.15	Benzene, methyl(1-methylethyl)- (CAS)
11	15.215	10.22	10.25	10.47	1,8-Cineole
12	16.611	1.40	1.30	11.34	.beta.-Ocimene
13	32.817	0.73	0.83	9.33	Bicyclo[2.2.1]heptan-2-ol, 1,7,7-trimethyl-, acetate, (1S-endo)- (CAS)
14	37.654	2.95	2.58	12.04	.ALPHA.-TERPINENYL ACETATE
15	38.678	0.19	0.25	7.92	.alpha.-ylangene
16	39.818	0.18	0.22	8.30	Cyclohexane, 1-ethenyl-1-methyl-2,4-bis(1-methylethenyl)-, [1S-(1.alpha.,2.beta.,4.beta.)]-
17	40.418	5.25	5.40	10.20	(-)-.beta.-Elemene
18	42.020	0.72	0.79	9.62	trans-4,11,11-Trimethyl-8-methylenebicyclo[7.2.0]undeca-4-ene
19	43.963	0.20	0.23	9.02	Alloaromadendrene
20	46.639	0.50	0.45	11.68	Naphthalene, decahydro-4a-methyl-1-methylene-7-(1-methylethenyl)-, [4aR-(4a.alpha.,7.alpha.,8a.beta.)]-
21	48.546	0.61	0.61	10.61	.gamma.-Muurolene
22	49.513	0.69	0.54	13.40	6-epi-shyobunol
23	50.939	0.57	0.45	13.20	1,3-Bis-(2-cyclopropyl,2-methylcyclopropyl)-but-2-en-1-one
24	61.902	3.38	1.85	19.19	Caryophyllene oxide
25	63.381	1.35	0.88	16.03	n-Propyl 5,8,11,14,17-eicosapentaenoate
26	68.441	42.62	15.04	29.73	2(3H)-Benzofuranone, 6-ethenylhexahydro-6-methyl-3-methylene-7-(1-methylethenyl)-, [3aS-(3a.alpha.,6.alpha.,7.beta.,7a.beta.)]-

Figure 4. GC-MS chromatogram of Laurel nobilis oil and fatty acid compositions of Laurel nobilis oil.

The FT-IR method, the primary uses of which are illustrated in Figure 5, was utilized to characterize the surface functional groups. The C = C stretching of aromatic rings is responsible for the bands with centers of 2320 and 2103 cm^{-1}, respectively [29] . The bands within the range of 1800–1500 cm^{-1} correspond to the C=O stretching vibrations of the lignin's keto-carbonyl groups and the C=C of its aromatic rings (1660 and 1545 cm^{-1}). Alcohol, carboxyl, and phenol all have C–Ostretching vibrations, which are responsible for the bands between 1300 and 1000 cm^{-1}. The existence of aromatic benzene rings is indicated by the bands about 950, 747, and 613 cm^{-1}. These bands are attributed to the aromatic C–H out-of-plane bending [30], [31]. The absorption peak is at 700-400 cm^{-1} for the C-C stretching vibration. [32].

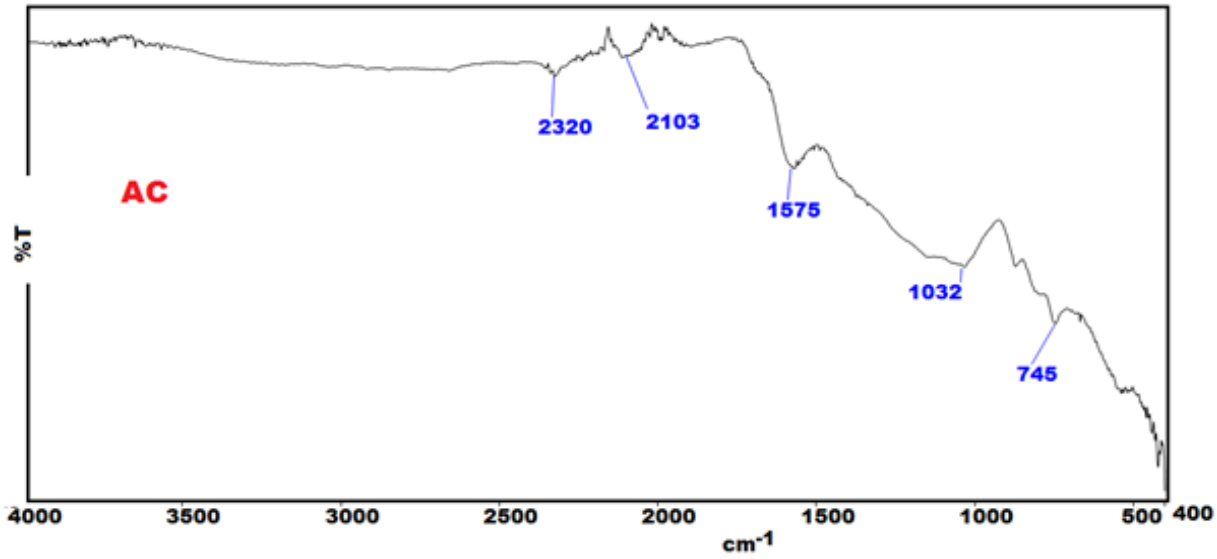


Figure 5. FT-IR spectrum of AC

The nitrogen adsorption-desorption isotherm of the activated carbon was measured in order to determine their physical characteristics, including specific surface area and total pore volume. The BET surface area was determined using the Brunauer-Emmett-Teller (BET) equation. The specific BET surface area and average pore diameter of the obtained activated carbon were determined as 783.58 m^2/g and 3.58 nm. Also, the total pore volume is 0.70 cm^3 g^{-1}. According to IUPAC, pores are divided into 3; micropore (<2 nm), mesopore (2-50 nm), and macropore (>50 nm) [30]. Table 2, which displays the BET surface areas of the AC activated at 500 °C, demonstrates the large surface area of the activated charcoal. The majority of the activated carbon generated in this work has mesoporous pores. The activated carbon adsorption and desorption isotherm generated at 500 °C was displayed in Figure 6.

The TGA curve of the resulting AC is given in Figure 7. Initial decomposition temperature (Ton) 276 °C. While the temperature at which 5% mass loss occurred was 481 °C, the remaining mass loss at 1000 °C was determined as 70.92.

The SEM image of AC was given in Figure 8. When you look at the images, you can see a non-flat, heterogeneous structure consisting of different shapes.

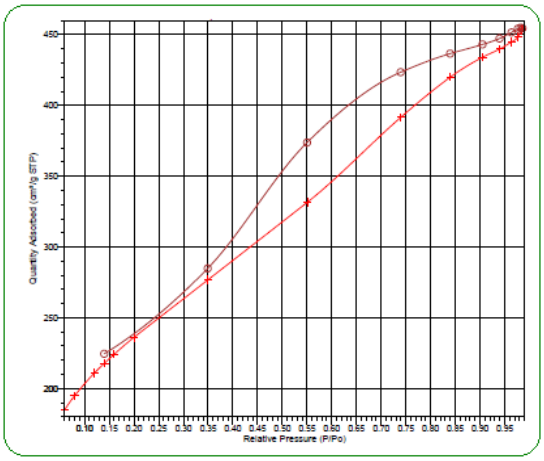


Figure 6. Adsorption–desorption isotherm of N_2 at 77 K of the activated carbon at 500°C.

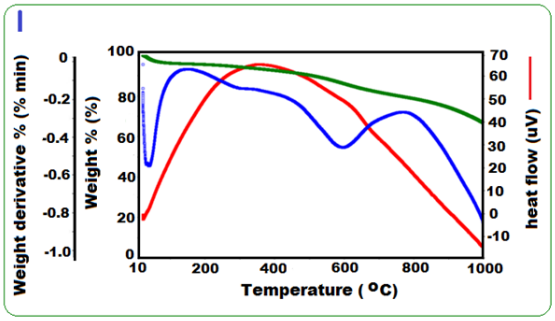


Figure 7. TGA analysis of AC

Table 2. BET results of activated carbon

	SBET (m^2/g)	micro volume (V_{micro}) (cm^3 g^{-1})	meso volume (V_{meso}) (cm^3 g^{-1})	total volume (Vt) (cm^3 g^{-1})	Average pore diameter (nm)
AC	783.58	0.039	0.661	0.70	3.58

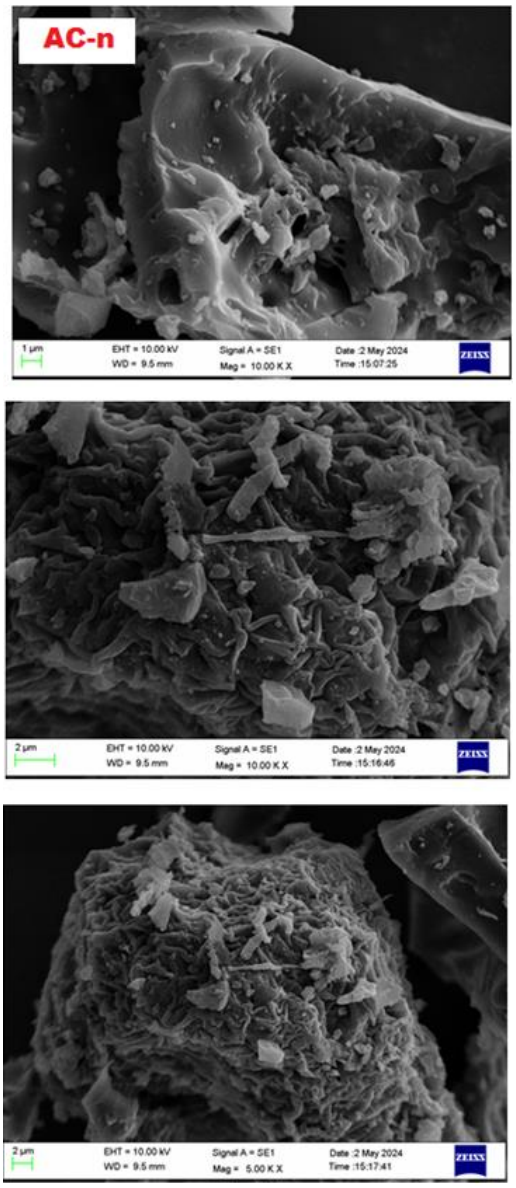


Figure 8. SEM images of AC

4. Conclusion and suggestions

L. nobilis oil was obtained from Laurel fruits brought from Hatay's Defne district in November 2023 by the hydroflotation method. Activated carbon was obtained from Laurel fruit seeds, which were waste as a result of oil extraction. The obtained activated carbon was characterized by FTIR, TGA and BET. Solid soap was obtained by the cold process method using oil obtained from bay fruits and activated carbon. Physicochemical parameters of the soap were determined. Activated carbon soaps with peeling effect and good foaming and cleaning properties were obtained.

Contributions of the authors

Hatice Karaer Yağmur: Methodology, Formal analysis, visualization, data curation, writing & editing and resources.
Kübra Baykara and Seyithan Sönmez: Investigation and resources.

Acknowledgment

The authors thank the Dicle University Scientific Research Project commission (DÜBAP) for support with the project number (Project Number: FEN.23.024).

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] O. Atolani, E. T. Olabiyi, A. A. Issa, H. T. Azeez, E. G. Onoja, S. O. Ibrahim, M. F. Zubair, O. S. Oguntoye, and G. A. Olatunji, "Green synthesis and characterisation of natural antiseptic soaps from the oils of underutilised tropical seed," *Sust. Chem. Pharm.*, vol. 4, pp. 32–39, 2016.

[2] H. Y. Nchimbi, "Quantitative and qualitative assessment on the suitability of seed oil from water plant (Trichilia emetica) for soap making," *Saudi J. Biol. Sci.*, vol. 27, no. 11, pp. 3161–3168, 2020.

[3] K. O. Boadu, M. A. Anang, and S. K. Kyei, "Chemical characterization of shea butter oil soap (Butyrospermum parkii G. Don)," *Int. J. Dev. Sustain.*, vol. 6, pp. 1282–1292, 2017.

[4] Y. B. Yurtlu, E. Yesiloglu, and F. Arslanoglu, "Physical properties of bay laurel seeds," *Int. Agrophys.*, vol. 24, pp. 325–328, 2010.

[5] B. Sırıken, C. Yavuz, and A. Güler, "Antibacterial Activity of *Laurus nobilis*: A review of literature," *Med. Sci. Discov.*, vol. 5, no. 11, pp. 374–379, 2018.

[6] O. Elkıran, E. Akbaba, and E. Bagcı, "Constituents of essential oils from leaves and seeds of *Laurus nobilis* L.: A chemotaxonomic approach," *Bangl. J. Bot.*, vol. 47, no. 4, pp. 893–901, 2018.

[7] E. Jayasathya and A. U. Santhoskumar, "Manufacture of activated carbon soap from areca nut shells," *Int. J. Eng. Appl. Sci. Technol.*, vol. 4, no. 12, pp. 261–264, 2020.

[8] M. Sajjad, R. Sarwar, T. Ali, L. Khan, and S. U. R. Mahmood, "Cosmetic uses of activated charcoal," *Int. J. Community Med. Public Health*, vol. 8, no. 9, pp. 4572–4574, 2021.

[9] A. A. Warra, "Sesame (*Sesamum indicum* L.) seed oil: Methods of extraction and its prospects in cosmetic industry—a review," *Bayero J. Pure Appl. Sci.*, vol. 4, no. 2, pp. 164–168, 2011.

[10] J. K. C. Nduka, P. O. Omozuwa, and O. E. Imanah, "Effect of heating time on the physicochemical properties of selected vegetable oils," *Arab. J. Chem.*, vol. 14, pp. 103063, 2021.

[11] H. K. Yağmur and İ. Kaya, "Synthesis and characterization of magnetic $ZnCl_2$-activated carbon produced from coconut shell for the adsorption of methylene blue," *J. Mol. Struct.*, vol. 1232, p. 130071, 2021.

[12] S. Berneckė, "Analysis of free fatty acids in soap samples by means of gas chromatography-mass spectrometry," *Chemija*, vol. 24, no. 4, pp. 307–311, 2013.

[13] "Saponification Process," *From Nature With Love*. [Online]. Available: https://www.fromnaturewithlove.com/resources/sapon.asp. [Accessed: Dec. 26, 2024].

[14] Z. Guoliang, "Bamboo charcoal handmade soap and preparation technology thereof," U.S. Patent, 2016.

[15] "Charcoal Soap," *Soapy Friends*. [Online]. Available: https://soapyfriends.com/charcoal-soap/. [Accessed: Dec. 26, 2024].

[16] I. Vermaak, G. P. P. Kamatou, B. Komane-Mofokeng, A. M. Viljoen, and K. Beckett, "African seed oils of commercial importance—Cosmetic applications," *South Afr. J. Bot.*, vol. 77, pp. 920–933, 2011.

[17] A. B. Asen, B. A. Ikyenge, D. T. Nyiatagher, and J. T. Ageh, "Chemical analysis of *Citrullus lanatus* (Thunb.), *Cucumeropsis mannii* (Naud.), and *Telfairia occidentalis* (Hook. f.) seeds oils," *J. Appl. Sci. Res.*, vol. 6, no. 3, pp. 265–268, 2010.

[18] X. Yu, Q. Li, D. Sun, X. Dong, and T. Wang, "Determination of the peroxide value of edible oils by FTIR spectroscopy using polyethylene films," *Anal. Methods*, vol. 7, no. 5, 2014.

[19] Z. Wu, J. Wei, T. Jiao, Q. Chen, M. Oyama, Q. Chen, and X. Chen, "A lead-based room-temperature phosphorescent metal–organic framework sensor for assessing the peroxide value of edible oils," *Food Chem.*, vol. 385, p. 132710, 2022.

[20] A. C. S. A. Anconi, N. C. S. Brito, and C. A. Nunes, "Determination of peroxide value in edible oils based on digital image colorimetry," *J. Food Compos. Anal.*, vol. 113, p. 104724, 2022.

[21] E. Ghohestani, J. Tashkhourian, and B. Hemmateenejad, "Colorimetric determination of peroxide value in vegetable oils using a paper-based analytical device," *Food Chem.*, vol. 403, p. 134345, 2023.

[22] E. Derwich, Z. Benziane, and A. Boukir, "Chemical composition and antibacterial activity of leaves essential oil of *Laurus nobilis* from Morocco," *Aust. J. Basic Appl. Sci.*, vol. 3, no. 4, pp. 3818–3824, 2009.

[23] S. Pakdeechot and L. Kaewsichan, "Tamanu and coconut oil blends for soap making from extraction of Tamanu kernel with coconut milk," *Eng. Appl. Sci. Res.*, vol. 47, no. 4, pp. 414–421, Oct.-Dec. 2020.

[24] I. Owoicho, "Quality assessment of soaps produced from bleached palm oil and *Moringa oleifera* seed oil," *Glob. J. Eng. Technol. Adv.*, vol. 7, no. 1, pp. 001–005, 2021.

[25] M. F. Zubair, O. Atolani, S. O. Ibrahim, O. S. Oguntoye, H. A. Abdulrahim, R. A. Oyegoke, and G. A. Olatunji, "Chemical and biological evaluations of potent antiseptic cosmetic products obtained from *Momordica charantia* seed oil," *Sust. Chem. Pharm.*, vol. 9, pp. 35–41, 2018.

[26] B. Antonić and D. Dordević, "Physicochemical characterization of home-made soap from waste-used frying oils," *Processes*, vol. 8, p. 1219, 2020.

[27] J. Abba, I. Y. Magaji, H. O. Opara, and D. I. Onyemachi, "Determination of alkali content and total fatty matter of some soaps used within Zaria, Nigeria," *Niger. Res. J. Chem. Sci.*, vol. 9, no. 2, 2021.

[28] K. J. Betsy, Mary Jilu, R. Fathima, and J. T. Varkey, "Determination of alkali content & total fatty matter in cleansing agents," *Asian J. Sci. Appl. Technol.*, vol. 2, no. 1, pp. 8–12, 2013.

[29] M. Kamaraj, T. G. Nithya, S. Shyamalagowri, J. Aravind, and R. Mythili, "Activated carbon derived from almond tree dry leaves waste for enhanced multi dye removal from aqueous solutions," *Mater. Lett.*, vol. 308, p. 131216, 2022.

[30] B. Belhamdi, H. Laksaci, C. Belabed, Z. Merzougui, S. Boudia, R. Tir, and M. Trari, "Synthesis of highly porous activated carbon derived from kernel oil treatment by-products of *Argania spinosa* as a recyclable adsorbent for amoxicillin removal from real wastewater," *Biomass Convers. Biorefinery*, vol. 13, no. 3, pp. 1–15, 2021.

[31] S. J. Rajasekaran and V. Raghavan, "Facile synthesis of activated carbon derived from *Eucalyptus globulus* seed as efficient electrode material for supercapacitors," *Diamond Relat. Mater.*, vol. 109, p. 108038, 2020.

[32] A. Ikhtiar Bakti and P. L. Gareso, "Characterization of activated carbon prepared from coconut shells using FTIR, XRD and SEM techniques," *J. Ilmiah Pendidikan Fisika Al-BiRuNi*, vol. 7, no. 1, pp. 33–39, 2018.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1111-1121 YEAR: 2024
DOI: 10.17798/bitlisfen.1518498

Face Recognition Using the Subspace and Deep Learning Algorithms for Cases of Sufficient and Insufficient Data

Serkan KESER[1*]

[1]*Kirsehir Ahi Evran University, Department of Electrical and Electronics Engineering, Kirsehir, Türkiye*

(ORCID: 0000-0001-8435-0507)

Keywords: Face recognition, Fisherface-KNN, DCVA, CNN. .

Abstract

In face recognition, the distance criterion significantly influences the recognition rate. Misclassified test signals can be accurately reassigned to the correct class using various distance measures and the nearest neighbor algorithm. This study uniquely explores the recognition performance of DCVA, Fisherface subspace classifiers, and Convolutional Neural Network (CNN) in face recognition, an aspect not thoroughly explored in the literature. Accordingly, this study introduces a Discriminative Common Vector-based (DCVA) algorithm utilizing various distance measures for face recognition for the first time. Additionally, the Fisherface-based algorithm uses different distance measures and nearest neighbors. Experiments were conducted on three different face databases. The images were downsampled to simulate both sufficient and insufficient data conditions. Experimental results indicate that the Correlation distance measure generally outperforms the Euclidean distance for the DCVA and Fisherface-KNN algorithms under both data conditions. The Fisherface-KNN algorithm surpasses the classical Fisherface in performance for various distance measures and nearest neighbor numbers and yields better recognition rates than the DCVA algorithm in sufficient data conditions. Moreover, while DCVA and Fisherface-KNN achieved superior results for two smaller face databases, CNN demonstrated better performance for larger databases.

1. Introduction

The application of speech, face, and object recognition systems is rapidly expanding across various fields. The primary objectives of these systems are to perform the recognition process with minimal delay while ensuring robustness, user security, scalability, and high recognition rates. It is crucial for these systems to provide accurate results under various environmental conditions and noise levels to maintain robustness. Additionally, they need to be scalable, meaning they should handle increasing data volumes or a larger number of users without significant performance degradation. Moreover, facial recognition systems must securely store personal data. A crucial condition for achieving high recognition rates is the appropriate classifier selection. Numerous classifiers such as CNN, ANN, KNN, SVM, Fisherface, and CVA are utilized for face recognition [1-7]. Among these, CVA is prominent as a subspace classifier with a high recognition rate and low computational complexity. It has been applied in various fields, including voice recognition, face recognition, image denoising, speaker recognition, and feature selection [8-13]. The number of samples and sample sizes in the training set classes are critical for CVA. If the size of a sample in the training set exceeds the total number of samples in all classes, an insufficient data case occurs; otherwise, a sufficient data case occurs. In the sufficient data case, the difference and indifference subspaces are intermixed and therefore the recognition rates may decrease. DCVA, a classifier mainly used in face recognition applications, is based

*Corresponding author: skeser@ahievran.edu.tr
Received: 18.07.2024, Accepted: 01.10.2024

on the effective subspace classifier CVA [7]. The principal characteristic of DCVA is its ability to find a vector representing common properties for each class during training, termed the common vector. During the testing phase, the feature vectors of the test signal are acquired by projecting the test signal into the indifference subspace. Subsequently, the test signal is assigned to the class with the smallest Euclidean distance between the common vectors from the training phase and the feature vectors of the test signal. Fisherface is another effective subspace method employed in face recognition [9]. In the Fisherface method, training signals are initially projected using optimal basis vectors that maximize class separation during training. Then, the test vector is projected using these basis vectors, and this projected test vector is assigned to the class with the smallest Euclidean distance between the projected test vector and the projected training vectors across all classes [9]. A primary disadvantage of the DCVA and Fisherface methods is their reliance solely on Euclidean distance criteria. Conversely, a test signal misclassified according to Euclidean distance can be correctly classified using different distance measures or varying the number of nearest neighbors. Thus, more advantageous distance measures than Euclidean can be identified in the recognition process by employing different distance measures. Similarly, the literature includes studies comparing recognition performances based on different distance measures [14-19,42]. In [42], it was demonstrated that cosine and correlation distance outperformed the Euclidean distance of the original SIFT for the ORL face database. This paper also shows that the correlation distance measure provides better results than the Euclidean distance. The primary reason for this superior performance is the correlation distance's robustness to variations in lighting conditions, face positions, and facial expressions, contributing to improved accuracy in face recognition algorithms.

This study proposes the DCVA algorithm, which utilizes different distance measures, and the Fisherface-KNN algorithm, which classifies based on various numbers of nearest neighbors and distance measures. Each image in the face database is downsampled to evaluate scenarios with both sufficient and insufficient data. The focus of the study is an in-depth performance comparison of DCVA, Fisherface-KNN, and CNN, using three different face databases. An examination of the average recognition results obtained by the proposed DCVA and Fisherface-KNN algorithms reveals that the correlation distance generally outperforms the Euclidean distance. Additionally, the Fisherface-KNN algorithm typically delivers better results than the classical Fisherface across various distance measures and numbers of nearest neighbors. In cases with sufficient data, Fisherface-KNN achieves higher recognition rates than DCVA.

A deep learning method, previously used as an expressive classifier in literature, was also employed through CNN [31-34]. The author in [37] introduced a face recognition method based on principal component analysis (PCA). However, PCA has limitations when handling variations in lighting conditions and facial expressions, which results in lower recognition rates. Another widely used approach is the application of support vector machines (SVMs) for face recognition [38]. SVMs effectively capture complex patterns in face images, though their computation time increases significantly with large datasets. Recent advances in deep learning, especially with CNNs, have achieved considerable success in face recognition. In [39], the well-known pre-trained AlexNet architecture was proposed. Following AlexNet, many CNN-based approaches for face recognition, such as FaceNet [40] and VGGFace [41], were developed.

Thus, the recognition performances of face databases of three different sizes were examined using DCVA, Fisherface-KNN, and CNN. The rest of the paper is structured as follows: Section 2 introduces the classifiers DCVA, Fisherface-KNN, and CNN, as well as the cases of sufficient and insufficient data. In Section 3, the face databases are presented, and the experimental results are discussed. Finally, Section 4 provides the conclusions.

2. Proposed methods

2.1. Proposed Subspace Methods

This study conducted tests using different distance measures for DCVA and Fisherface-KNN. For DCVA, the nearest neighbor algorithm cannot be applied because it identifies a common vector for each class. In contrast, an algorithm using KNN decision criteria is proposed for Fisherface. The proposed Fisherface algorithm is based on the KNN structure and incorporates different distance measures, such as Cityblock, Correlation, Euclidean, and Spearman. The KNN algorithm predicts the class of a test vector based on the number of nearest neighbors, where the value of K, representing the number of nearest neighbors, directly influences the classification outcome and must be determined experimentally [14].

Table 1 presents the computational complexity of the classifiers along with a description of their parameters. As shown in Table 1, DCVA has low, Fisherface+KNN medium, and CNN high memory and computational complexity values.

Table 1. Comparison of computational complexity and resource usage of the DCVA, Fisherface+KNN, and CNN.

Classifier	Computational Complexity	Resource Usage (Memory and CPU)	Explanation
DCVA	$O(T \cdot K \cdot n^2)$	Low	T: Number of classes K: Number of samples per class n: Data dimension
Fisherface+KNN	$O(n^3) + O(K \cdot n)$	Medium	n^3: Complexity of PCA and LDA operations K: Number of nearest neighbors
CNN	$O(l \cdot m \cdot f^2 \cdot c)$	High	l: Number of layers m: Number of filters f: Filter size c: Input channels (e.g., 3 for RGB images)

2.1.1. The Proposed DCVA Algorithm

DCVA is superior to other subspace classifiers, especially in terms of low computational complexity. This advantage comes from using a discriminative common vector belonging to a class. [7]. If there are *T* different classes in the training set, where each class contains *K* samples, there will be a total of *M=KT* samples. If we denote the sample of the r*th* signal with class *i* as $\boldsymbol{x_r^i}$ in *n*-dimensional space, the within-class scatter matrix $\boldsymbol{S_w}$ is found as follows:

$$\boldsymbol{S_w} = \sum_{i=1}^{T} \sum_{r=1}^{K} \left(\left(\mathbf{x}_r^i - \boldsymbol{\mu_i} \right) \left(\boldsymbol{x_r^i} - \boldsymbol{\mu_i} \right)^T \right) \qquad (1)$$

where $\boldsymbol{\mu_i}$ denotes the mean vector of the ith class. The projection matrix is found by the eigenvectors corresponding to $\boldsymbol{S_w}$'s smallest eigenvalues, spanning the indifference subspace [7].

$$\mathbf{P} = \mathbf{V}\mathbf{V}^T \qquad (2)$$

The common vectors of the classes are obtained as follows:

$$\boldsymbol{x_{com}^i} = \boldsymbol{P x_r^i}, \ r=1, 2, \ldots, K, i=1, 2,.., T \qquad (3)$$

The $\mathbf{S_{com}}$ is found using the following equation:

$$\boldsymbol{S_{com}} = \sum_{i=1}^{T} \left(\boldsymbol{x_{com}^i} - \boldsymbol{\mu_{com}} \right) \left(\boldsymbol{x_{com}^i} - \boldsymbol{\mu_{com}} \right)^T, \boldsymbol{i} = \mathbf{1}, \ldots, \boldsymbol{T} \qquad (4)$$

The eigenvectors associated with the nonzero eigenvalues of the $\boldsymbol{S_{com}}$ matrix provide the projection vectors for the DCVA:

$$\boldsymbol{J}(\mathbf{W_{opt}}) = \boldsymbol{argmax} \left| \mathbf{W}^T \mathbf{S_{com}} \mathbf{W} \right|. \qquad (5)$$

The feature vectors can be written as follows:

$$\boldsymbol{\Omega_i} = \left[< \boldsymbol{x_r^i}, \boldsymbol{w_1} > \cdots < \boldsymbol{x_r^i}, \boldsymbol{w_{C-1}} > \right] \qquad (6)$$

where $\boldsymbol{\Omega_i}$ are the discriminative common vectors. Then, the feature vectors of test images are found by:

$$\boldsymbol{\Omega_{test}} = \boldsymbol{W_{opt}^T x_{test}} \qquad (7)$$

where $\boldsymbol{W_{opt}^T} = [\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_{T-1}}]^T$ and $\boldsymbol{\Omega_{test}} \in \mathbb{R}^{(\boldsymbol{T-1})\times \mathbf{1}}$. The operations described above were performed for the insufficient data case ($\boldsymbol{M} < \boldsymbol{n}$). In the case of sufficient data ($\boldsymbol{M} > \boldsymbol{n}$), the common vector of the ith class is obtained from the projection of the mean feature vector ($\boldsymbol{\mu_i}$) of that class onto the indifference subspace:

$$\boldsymbol{x_{com}^i} = \mathbf{P}\boldsymbol{\mu_i} \qquad (8)$$

In case of sufficient data, the smallest *k* eigenvalues are selected to correspond to a certain *z* percentage of the sum of the eigenvalues [23],

$$\boldsymbol{z} = \frac{\sum_{i=1}^{k} \boldsymbol{\lambda_i}}{\sum_{i=1}^{n} \boldsymbol{\lambda_i}} \qquad (9)$$

where *n* represents the total number of eigenvalues and *k* indicates the count of the smallest selected eigenvalues. The DCVA algorithm's block diagram is illustrated in Figure 1. The abbreviation DCV stands for Discriminative Common Vectors.

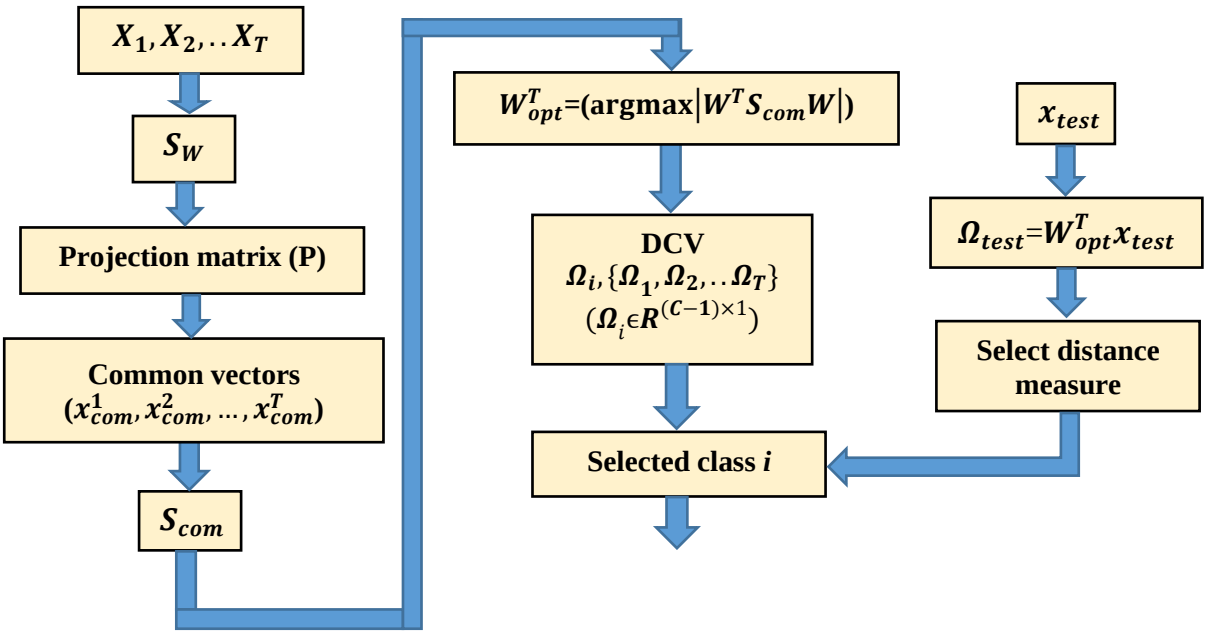


Figure 1. Block diagram of the proposed DCVA algorithm.

2.1.2. The Proposed Fisherface -KNN Algorithm

Fisherface is one of the important algorithms based on Linear Discriminant Analysis for face recognition due to its effort to maximize the discrimination between classes in the training process [9, 20]. The S_W is found using Eq. 1 mentioned above, and the between-class scatter matrix S_B is calculated by:

$$S_B = \sum_{i=1}^{T} N(\mu_i - \mu)(\mu_i - \mu)^T \qquad (10)$$

where N represents the number of samples in a class, μ_i is the mean of the *ith* class, and μ is the mean of all classes. The optimal basis vectors (U_{opt}) are determined as follows [9]:

$$U_{opt} = \underset{U}{argmax} \frac{|U^T S_B U|}{|U^T S_W U|}, \qquad (11)$$

In the Fisherface method, the above equations are valid for the sufficient data case ($M > n$). In the face recognition problem, in the case of insufficient data ($M < n$), the within-class scatter matrix S_W is usually singular. To overcome this problem, the Fisherface method projects the feature space dimension onto a lower-dimensional space using Principal Component Analysis (PCA). Using U_{opt}, feature vectors are found for each class as follows:

$$\Omega_i = U^T_{opt} X_i, \quad i = 1, 2, \dots, T \qquad (12)$$

where $\Omega_i \in \mathbb{R}^{(T-1)\times K}$ and K is the number of samples in the ith class. For classification, the test signal is first projected using U_{opt}, and then Ω_{test} ($\Omega_{test} \in \mathbb{R}^{(T-1)\times 1}$) is found by:

$$\Omega_{test} = \mathrm{U}^T_{\text{opt}} x_{test} \qquad (13)$$

The block diagram of the Fisherface-KNN algorithm is given in Figure 2 below. The S_W and S_B scatter matrices are first found for the face images in the training set. The training process is performed using the KNN algorithm for the feature vectors (Ω_i). Then, the test signal (Ω_{test}) is found and assigned to the most appropriate class using the KNN algorithm.

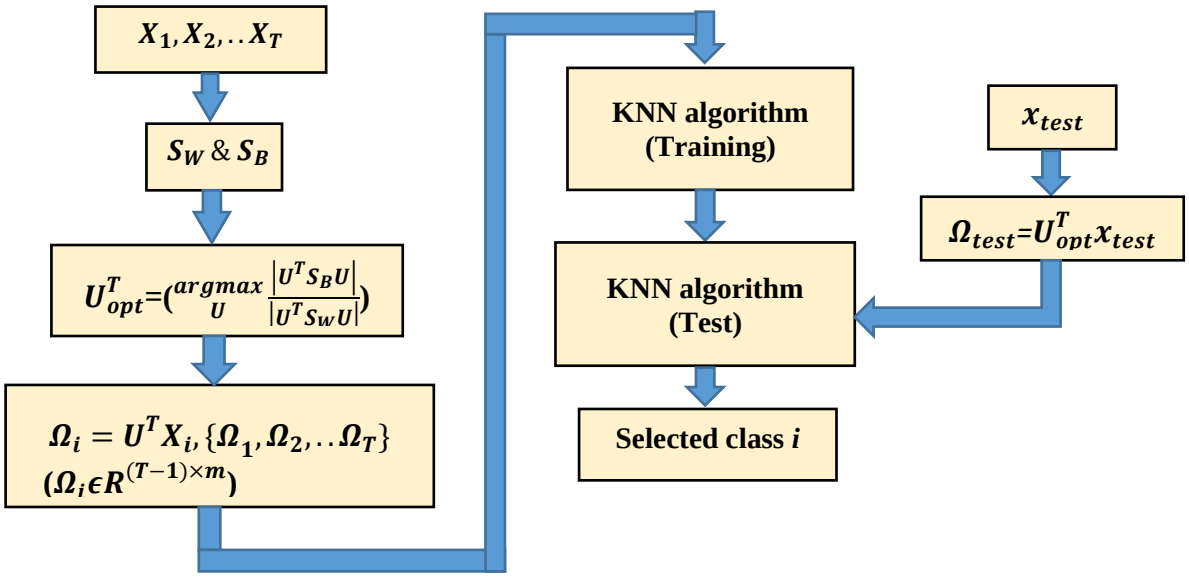


Figure 2. Block diagram of the proposed Fisherface-KNN algorithm.

2.1.3. Proposed CNN Model

In deep learning, a Convolutional Neural Network (CNN) is a class of artificial neural network, most commonly used for applications such as image and video recognition [26, 27], image classification, image segmentation [28, 29], and brain-computer interfaces [30]. The CNN model in Figure 3 was used in the study.

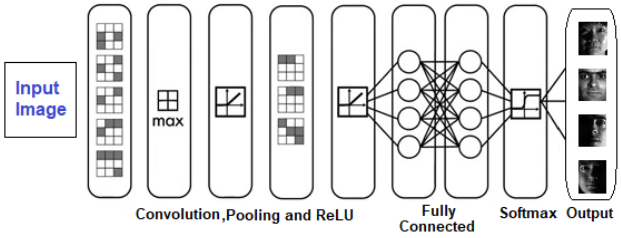


Figure 3. The proposed CNN method.

The layers of the proposed CNN model are listed in Table 2.

Table 2. The layers of the proposed CNN model

Layer level	Layers
1	Image Input (NxMx1 images)
2	Convolution (3x3 convolutions, 8 filters)
3	Batch Normalization
4	ReLU
5	Max Pooling (2x2 max pooling)
6	Convolution (3x3 convolutions,16 filters)
7	Batch Normalization
8	ReLU
9	Max Pooling (2x2 max pooling)
10	Convolution (3x3 convolutions,32 filters)
11	Batch Normalization
12	ReLU
13	Fully Connected
14	Softmax
15	Classification Output

The proposed CNN architecture includes three convolutional layers, two max-pooling layers, and three regularization layers. The training was conducted over 12 epochs with eight iterations, using a learning rate of 0.01. The network training employed the stochastic gradient descent with momentum (SGDM) optimizer.

2.2. The Cases of Sufficient and Insufficient Data

For DCVA and Fisherface, the difference and indifference spaces can be completely separated in the insufficient data case, while these spaces cannot be distinguished in the sufficient data case. To investigate the effect of the sufficient data case for DCVA and Fisherface-KNN, the sizes of the images were reduced using the downsampling method. The image is downsampled using the mathematical expressions below, resulting in sub-images.

$$I_s = I(k_i, k_j) \qquad i=1, 2 \text{ and } j=1,2, \qquad (14)$$

where I represents test images, $\mathbf{k_1}$ $(k_1 = 1, 3, 5, \dots, N-1)$ and k_2 $(k_2 = 2, 4, 6, \dots, N)$ represent the pixel indices of the rows and columns of images and I_s represents the sub-images. Four sub-images are found as follows for the pth level,

$$I_{p1} = I(k_i, k_j), \qquad i=1 \text{ and } j=1,$$

$$I_{p2} = I(k_i, k_j), \qquad i=1 \text{ and } j=2,$$

$$I_{p3} = I(k_i, k_j), \qquad i=2 \text{ and } j=1,$$

$$I_{p4} = I(k_i, k_j), \qquad i=2 \text{ and } j=2,$$

The average of sub-images is found, and an **N/(2ᵖ) ×N/(2ᵖ)**-dimensional matrix is obtained for the *p*th level. Figure 4 below shows how the downsampling process is done, and DS indicates the downsampling process.

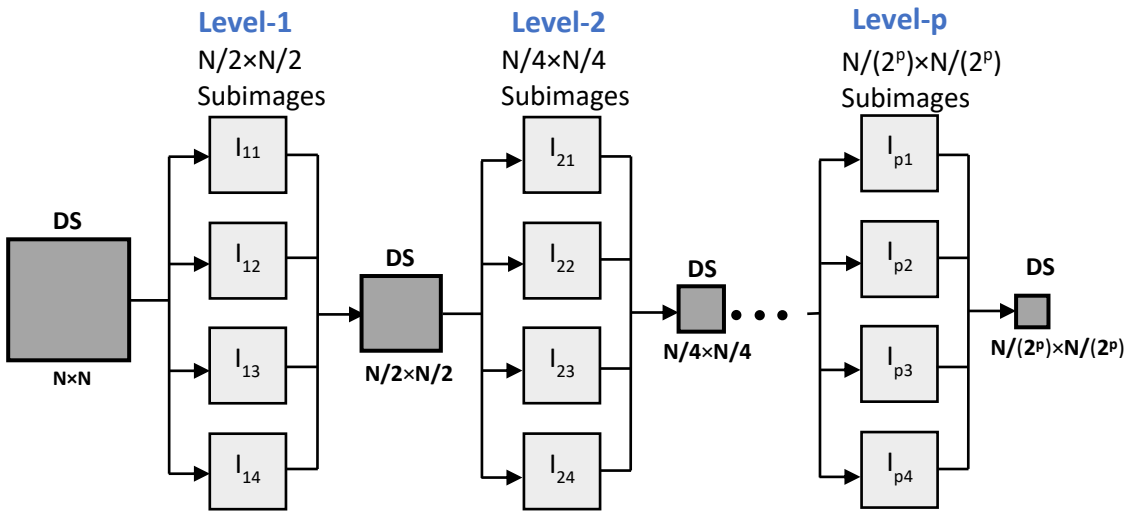


Figure 4. The downsampling process of an image for *p* level.

3. Experimental Studies

The experiments were conducted on a desktop PC with a 3 GHz (i5-7400) processor and 8 GB RAM under Windows 10. The ORL, YALE, and Cropped YALE (C-YALE) face databases were used in the studies. The ORL face database includes 400 images of size 112 × 92, with face records of 40 individuals taken at ten different poses. Each face image was manually reduced to 64 × 64 size. The YALE face database contains face images of 15 individuals, with ten different images per person. The C-YALE database consists of 2280 images for 38 individuals, aligned, cropped, and resized to 168 × 192. In this database, 40 images were used for training and 20 for testing, with 3-fold cross-validation. In the test phase, test signals were classified using different distances such as Correlation, Cityblock, Euclidean, and Spearman. The recognition process was performed according to the numbers of nearest neighbors (K=1, K=3, K=5). In the experimental studies, 10-fold cross-validation was used to evaluate the performance of the proposed algorithms for the ORL and YALE face databases. Some image sizes obtained for different *p*-values are shown in Figure 5 below.

Figure 5. Some images of the C-YALE database for p=0, p=1, p=2 and p=3.

All experimental results are given in the tables below. The symbol * indicates the sufficient data case, and the letter "p" shows the downsampling levels. CB, COR, EUC, and SP represent Cityblock, Correlation, Euclidean, and Spearman distances, respectively. As shown in Table 3, the best results for the proposed DCVA algorithm were generally obtained using the Correlation distance for sufficient and insufficient data cases across the three databases.

Table 3. Mean accuracy rates of the proposed DCVA algorithm for the ORL and YALE

	Mean accuracy rates (%)					
	Measures	**p=0**	**p=1**	**p=2**	**p=3**	**p=4***
YALE	**CB**	97.67	97.34	97.34	98	92
	COR	99.34	99.34	99.34	99.34	92
	EUC	98.34	98.34	98.34	98	93.34
	SP	96.34	96	96	96	86
		p=0	**p=1**	**p=2***	**p=3***	
ORL	**CB**	99.25	99.25	97.5	88.75	
	COR	100	100	98.25	90.25	
	EUC	99.75	99.75	97.5	90.5	
	SP	97.25	95.75	95.5	68.25	
		p=0	**p=1**	**p=2***	**p=3***	
C-YALE	**CB**	91.11	90.87	89.61	87.42	
	COR	91.81	91.42	90.25	88.22	
	EUC	91.62	91.38	90.13	87.91	
	SP	89.21	88.82	87.53	86.21	

Recognition rates for DCVA decreased significantly for face databases in the case of sufficient data. The average recognition rates of the Fisherface-KNN algorithm were found according to the number of nearest neighbors (K=1, K=3, K=5) and different distance measures. For insufficient data, the Correlation distance gave the best results for K=1, K=3, and K=5. In the case of sufficient data, Euclidean and correlation distance measures gave similar results for the YALE database. However, the best results were obtained using the Correlation distance for the C-YALE database. Regarding sufficient data, Euclidean and correlation distances gave similar results for the YALE database in Table 4. However, the best results were obtained using the correlation distance for the C-YALE database, as shown in Table 5.

Table 4. Mean accuracy rates of the proposed Fisherface-KNN (K=1, K=3, K=5) algorithm for the YALE database

	Mean accuracy rates (%)					
	Measures	**p=0**	**p=1**	**p=2**	**p=3**	**p=4***
K=1	**CB**	97.33	97.33	98.00	98.67	98.00
	COR	99.34	99.34	98.67	98.67	98.34
	EUC	99.00	98.67	98.67	98.67	98.34
	SP	96.00	96.00	97.34	92.67	92.67
K=3	**CB**	98.67	98.67	98.00	98.67	98.34
	COR	99.67	99.67	99.34	99.34	98.67
	EUC	99.34	99.34	98.67	99.34	98.67
	SP	98.67	98.67	98.00	93.34	94.67
K=5	**CB**	98.34	99.67	99.34	98.67	98.67
	COR	99.67	99.34	98.67	98.00	99.34
	EUC	98.67	97.67	98.67	99.34	99.34
	SP	98.34	98.00	98.67	98.67	94.67

Table 5. Mean accuracy rates of the proposed Fisherface-KNN algorithm for the C-YALE database

	Mean accuracy rates (%)				
	Measures	**p=0**	**p=1**	**p=2**	**p=3***
K=1	**CB**	89.32	88.82	88.02	86.97
	COR	90.21	89.84	89.24	88.96
	EUC	90.06	89.76	88.72	88.62
	SP	86.89	86.65	86.13	85.95
K=3	**CB**	90.08	89.62	88.81	87.77
	COR	90.96	90.53	90.18	90.01
	EUC	90.87	90.26	89.75	89.07
	SP	87.52	87.21	87.02	86.73
K=5	**CB**	89.93	89.52	88.74	87.58
	COR	90.68	90.23	89.93	89.65
	EUC	90.66	90.12	89.21	88.95
	SP	87.50	87.20	86.92	86.65

Table 6 provides the average recognition rates of the Fisherface-KNN algorithm for the ORL database. The Correlation distance generally gave the best results for insufficient data. while Euclidean performed better in the sufficient data case.

Table 6. Mean accuracy rates of the proposed Fisherface-KNN (K=1, K=3, K=5) algorithm for the ORL

	Mean accuracy rates (%)				
	Measures	**p=0**	**p=1**	**p=2***	**p=3***
K=1	**CB**	98.50	98.50	99.25	98.50
	COR	99.00	99.25	99.25	98.75
	EUC	99.00	99.00	99.25	98.75
	SP	96.25	97.75	98.25	93.25
K=3	**CB**	99.00	98.50	98.50	98.50
	COR	99.00	99.25	99.00	99.00
	EUC	99.25	99.00	99.25	99.25
	SP	97.50	98.75	95.75	92.00
K=5	**CB**	98.25	98.00	98.00	98.00
	COR	98.50	99.00	98.50	98.25
	EUC	98.75	98.50	98.50	98.75
	SP	96.75	98.50	92.50	90.25

Table 7 presents the average recognition rates for the CNN model under both sufficient and insufficient data cases for the three face databases.

Table 7. Mean accuracy rates of the CNN for the three databases

Databases	**Mean accuracy rates (%)**				
	p=0	**p=1**	**p=2**	**p=3**	**p=4***
YALE	93.34	92	88.67	88.67	87.34
	p=0	**p=1**	**p=2***	**p=3***	
ORL	95.25	92.5	86.25	85	-
	p=0	**p=1**	**p=2**	**p=3***	
C-YALE	93.97	93.48	92.89	91.62	-

Table 8 compares the highest recognition rates of the DCVA, Fisherface-KNN, and CNN algorithms. The best results were generally achieved using the Correlation distance for subspace algorithms. The DCVA algorithm yielded higher average recognition rates than Fisherface-KNN for the ORL database, but Fisherface-KNN outperformed DCVA for the YALE database. Additionally, for ORL and YALE databases, the average recognition rates of CNN were lower than those of DCVA and Fisherface-KNN. However, CNN achieved higher recognition rates for the C-YALE database than DCVA and Fisherface-KNN. The reason for CNN's lower recognition rates on small databases is that CNN requires a large amount of data to learn the features of faces, and when working with a small dataset, there is not enough variety of examples to learn these features. Due to insufficient data, the model cannot fully learn the necessary features to distinguish between different faces, leading to low recognition rates.

Table 8. The highest accuracy rates of the DCVA and Fisherface-KNN algorithms

Mean accuracy rates (%)								
DCVA			**Fisherface-KNN**			**CNN**		
ORL	**YALE**	**C-YALE**	**ORL**	**YALE**	**C-YALE**	**ORL**	**YALE**	**C-YALE**
100 (COR)	99.34 (COR)	91.81 (COR)	99.25 (COR&EUC)	**99.67** (COR)	90.96 (COR)	95.25	93.34	**93.97**

4. Conclusion

This study developed DCVA and Fisherface-based algorithms and explored the effects of different distance measures and the number of nearest neighbors on recognition rates. Additionally, a Convolutional Neural Network (CNN) was employed. The performances of the two subspace classifiers and CNN were examined under both sufficient and insufficient data cases. For the subspace classifiers, the Correlation distance generally provided higher recognition rates compared to Euclidean distance. In the sufficient data case, Fisherface-KNN outperformed DCVA, suggesting that Fisherface-KNN delivers better results when difference and indifference subspaces cannot be easily distinguished.

The findings show that the choice of distance measures in subspace algorithms and the number of nearest neighbors in Fisherface-KNN can lead to better results than Euclidean distance. For the ORL and YALE databases, CNN produced lower average recognition rates compared to DCVA and Fisherface-KNN. However, for the C-YALE database, CNN achieved higher recognition rates, highlighting that subspace methods are more effective for smaller databases, while CNN excels with larger datasets. The lower recognition rates of CNN for small databases are attributed to insufficient data, which prevents the model from fully learning the distinguishing features between different faces.

The Correlation distance, which measures the direction and correlation between two vectors, is particularly robust to changes in lighting and brightness in face images. While brightness variations negatively affect other distance measures, such as Euclidean distance, Correlation distance remains stable against these variations. Consequently, the experiments demonstrated higher accuracy rates when using Correlation distance.

This study focuses on three classifiers and small databases. However, future research will aim to include more classifiers and larger databases, as well as investigate recognition performance using hybrid classifiers.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] S. Jain and D. Bhati, "Face recognition using ANN with reduce feature by PCA in wavelet domain," *International Journal of Scientific Engineering and Technology*, vol. 2, no. 6, pp. 595–599, 2013.

[2] M. A. Abuzneid and A. Mahmood, "Enhanced human face recognition using LBPH descriptor, multi-KNN, and back-propagation neural network," *IEEE Access*, vol. 6, pp. 20641–20651, 2018.

[3] H. S. Dadi and G. M. Pillutla, "Improved face recognition rate using HOG features and SVM classifier," *IOSR Journal of Electronics and Communication Engineering*, vol. 11, no. 4, pp. 34–44, 2016.

[4] M. Anggo and L. Arapu, "Face recognition using fisherface method," in *Journal of Physics: Conference Series*, vol. 1028, no. 1, p. 012119. IOP Publishing, 2018.

[5] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.

[6] S. Ergin and M. B. Gulmezoglu, "Face recognition based on face partitions using common vector approach," in *2008 3rd International Symposium on Communications, Control and Signal Processing (ISCCSP)*, 2008, pp. 624–628.

[7] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 4–13, 2005.

[8] A. Martinez, "Fisherfaces," *Scholarpedia*, vol. 6, no. 2, p. 4282, 2011.

[9] N. Kumar., P. Belhumeur and S. Nayar. "Facetracer: A search engine for large collections of images with faces", *In Computer Vision–ECCV 2008: 10th European Conference on Computer Vision*, Marseille, France, Proceedings, Part IV 10, pp. 340-353, 2008.

[10] K. Özkan and E. Seke, "Image denoising using common vector approach," *IET Image Processing*, vol. 9, no. 8, pp. 709–715, 2015.

[11] S. Sadıç and M. B. Gülmezoğlu, "Common vector approach and its combination with GMM for text-independent speaker recognition," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11394–11400, 2011.

[12] Ş. Işık, K. Özkan, and Ö. N. Gerek, "CVABS: moving object segmentation with common vector approach for videos," *IET Computer Vision*, vol. 13, no. 8, pp. 719–729, 2019.

[13] S. Günal, S. Ergin, and Ö. N. Gerek, "Spam E-mail recognition by subspace analysis," in *INISTA – International Symposium on Innovations in Intelligent Systems and Applications*, 2005, pp. 307–310.

[14] M. L. Zhang and Z. H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40, no. 7, pp. 2038–2048, 2007.

[15] P. Miller and J. Lyle, "The effect of distance measures on the recognition rates of PCA and LDA based facial recognition," in *Digital Image Processing*, 2008.

[16] M. S. Ahuja and S. Chhabra, "Effect of distance measures in PCA based face recognition," *International Journal of Enterprise Computing and Business Systems*, vol. 1, no. 2, p. 2230–8849, 2011.

[17] H. Saadatfar, S. Khosravi, J. H. Joloudari, A. Mosavi, and S. Shamshirband, "A new K-nearest neighbors classifier for big data based on efficient data pruning," *Mathematics*, vol. 8, no. 2, p. 286, 2020.

[18] A. G. Hatzimichailidis, G. A. Papakostas, and V. G. Kaburlasos, "A novel distance measure of intuitionistic fuzzy sets and its application to pattern recognition problems," *International Journal of Intelligent Systems*, vol. 27, no. 4, pp. 396–409, 2012.

[19] V. Perlibakas, "Distance measures for PCA-based face recognition," *Pattern Recognition Letters*, vol. 25, no. 6, pp. 711–724, 2004.

[20] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[21] M. Anggo and L. Arapu, "Face recognition using fisherface method," in *Journal of Physics: Conference Series*, vol. 1028, no. 1, p. 012119. IOP Publishing, 2018.

[22] K. Chomboon, P. Chujai, P. Teerarassamee, K. Kerdprasop, and N. Kerdprasop, "An empirical study of distance metrics for k-nearest neighbor algorithm," in *Proceedings of the 3rd International Conference on Industrial Application Engineering*, 2015, pp. 280–285.

[23] Y. Xie, Y. Wang, A. Nallanathan, and L. Wang, "An improved K-nearest-neighbor indoor localization method based on spearman distance," *IEEE Signal Processing Letters*, vol. 23, no. 3, pp. 351–355, 2016.

[24] M. B. Gülmezoğlu, V. Dzhafarov, R. Edizkan, and A. Barkana, "The common vector approach and its comparison with other subspace methods in case of sufficient data," *Computer Speech & Language*, vol. 21, no. 2, pp. 266–281, 2007.

[25] M. V. Valueva, N. N. Nagornov, P. A. Lyakhov, G. V. Valuev, and N. I. Chervyakov, "Application of the residue number system to reduce hardware costs of the convolutional neural network implementation," *Mathematics and Computers in Simulation*, 2020.

[26] L. Shang, Q. Yang, J. Wang, S. Li, and W. Lei, "Detection of rail surface defects based on CNN image recognition and classification," in *2018 20th International Conference on Advanced Communication Technology (ICACT)*, 2018, pp. 45–51.

[27] Y. Fan, X. Lu, D. Li, and Y. Liu, "Video-based emotion recognition using CNN-RNN and C3D hybrid networks," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 2016, pp. 445–450.

[28] M. Zhang, W. Li, and Q. Du, "Diverse region-based CNN for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2623–2634, 2018.

[29] B. Kayalibay, G. Jensen, and P. van der Smagt, "CNN-based segmentation of medical imaging data," *arXiv preprint arXiv:1701.03056*, 2017.

[30] J. Thomas, T. Maszczyk, N. Sinha, T. Kluge, and J. Dauwels, "Deep learning-based classification for brain-computer interfaces," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017, pp. 234–239.

[31] H. Ben Fredj, S. Bouguezzi, and C. Souani, "Face recognition in unconstrained environment with CNN," *The Visual Computer*, vol. 37, no. 2, pp. 217–226, 2021.

[32] S. Sharma, K. Shanmugasundaram, and S. K. Ramasamy, "FAREC—CNN based efficient face recognition technique using Dlib," in *2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, 2016, pp. 192–195.

[33] M. Arsenovic, S. Sladojevic, A. Anderla, and D. Stefanovic, "FaceTime—Deep learning based face recognition attendance system," in *2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY)*, 2017, pp. 000053–000058.

[34] S. Saxena and J. Verbeek, "Heterogeneous face recognition with CNNs," in *European Conference on Computer Vision*, 2016, pp. 483–491.

[35] K. C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, 2005.

[36] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6.

[37] S. Khare and M. Totaro, "Ensemble learning for detecting attacks and anomalies in IoT smart home," in *Proceedings of the 2020 3rd International Conference on Data Intelligence and Security (ICDIS)*, 2020, pp. 56–63.

[38] N. Butt, A. Shahid, K. N. Qureshi, S. Haider, A. O. Ibrahim, F. Binzagr, and N. Arshad, "Intelligent deep learning for anomaly-based intrusion detection in IoT smart home networks," *Mathematics*, vol. 10, p. 4598, 2022.

[39] E. Anthi, L. Williams, M. Slowinska, G. Theodorakopoulos, and P. Burnap, "A supervised intrusion detection system for smart home IoT devices," *IEEE Internet of Things Journal*, vol. 6, pp. 9042–9053, 2019.

[40] C. Stolojescu-Crisan, C. Crisan, and B. P. Butunoi, "An IoT-based smart home automation system," *Sensors*, vol. 21, p. 3784, 2021.

[41] M. R. Dhobale, R. Y. Biradar, R. R. Pawar, and S. A. Awatade, "Smart home security system using IoT, face recognition, and Raspberry Pi," *IEEE Int. J. Comput. Appl.*, vol. 176, pp. 45–47, 2020.

[42] H. Kumar and P. Padmavati, "Face recognition using SIFT by varying distance calculation matching method," *International Journal of Computer Applications*, vol. 47, no. 3, pp. 20–26, 2012.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1122-1138 YEAR: 2024
DOI: 10.17798/bitlisfen.1521143

A Study on the Examination of the 21st-Century Skills of the Personnel Working in the Ministry of Youth and Sports

Fatih HANOĞLU[1], Canan SAYIN TEMUR[2*]

[2]*Ankara Yıldırım Beyazıt University, Faculty of Sport Sciences, Ankara, Türkiye*
(ORCID: 0000-0002-5430-1875) (ORCID: 0000-0003-0583-8083)

Keywords: 21st century skills, Ministry Youth and Sports personnel, Sports.

Abstract

The objective of this study is to examine the levels of 21st-century skills among personnel employed at the Ministry of Youth and Sports, with particular consideration of the influence of socio-demographic variables. A total of 369 personnel, comprising 148 women and 221 men, participated in the research on a voluntary basis. During the data collection process, the researchers utilised a Personal Information Form and the Multidimensional 21st Century Skills Scale, which was developed by Çevik and Şentürk (2019). The scale developed by Çevik and Şentürk (2019) comprises five dimensions with a total of 41 items and is structured as a 5-point Likert-type scale. The data were analysed using the statistical software package SPSS 23.0. As the scale scores exhibited a normal distribution, an independent samples t-test was employed to compare the means of two groups, while a one-way analysis of variance (ANOVA) was used for comparing the means of more than two groups. Upon conclusion of the research, a statistically significant correlation was identified between the scores of the participants with regard to the sub-dimensions of information and technology literacy, critical thinking and problem-solving, entrepreneurship and innovation skills, social responsibility and leadership skills, and career awareness. Upon analysis according to demographic variables, a significant difference was observed in the 21st-century skills of the personnel participating in the research with regard to the variables of professional status and length of service in the institution ($p < 0.05$). Furthermore, no significant differences were identified in the 21st-century skills of the participants according to gender, marital status, duration of use of information technology tools, age group, education level, or income status ($p > 0.05$).

1. Introduction

Societies are in constant change, and the pace of this change has increased especially in the 21st century. The information produced in every field is replaced in a short time, and the skills, once practical, are inadequate to solve current problems. Unlike past times, this applies not only to fields such as engineering and medicine but to all areas of life, not only to countries in a certain geography but all over the world, not only to individuals living in cities but also to those living in rural areas, including people of all ages. In the constantly changing and developing world, the expected skills also change depending on people's changing needs. It can be seen that the concept of skill has been defined in various ways in the relevant literature. For example, its definition according to the Turkish Higher Education Qualifications Framework is "the ability to apply knowledge, solve problems, and complete tasks" [1]. According to this definition, it can be said that individuals tend to learn in different ways depending on their environment and conditions because, in a changing and developing world, an inevitable reality for individuals is adaptation. At this point, it can be said that a learner generally carries out the act of learning by obtaining information in line with his/her needs [2]. Although these needs vary depending on

*Corresponding author: canansayintemur@aybu.edu.tr
Received: 23.07.2024, Accepted: 08.12.2024

individuals' living conditions, the necessities of today have made certain skills necessary for all individuals who want to adapt to this age.

To deal with the challenges of this rapidly changing, increasingly complex, and unpredictable world, people need to rethink the way they think and organize information. This is because many jobs that took individuals much longer to do in previous centuries can be done in a very short time today, thanks to changing living conditions and technology integration. For this reason, being good at basic skills, such as reading, writing, and doing mathematics alone cannot meet employees' needs, especially in the business world. Even in low-paying jobs, individuals are expected to have the ability to produce solutions to the problems they encounter [3]. At this point, as we were entering the 21st century, the Jacques Delors Report released in 1996 shaped the discussions on how children and young people could be better prepared for the new century and ultimately led to establishing a consensus [4]. Reaching similar results, Keskin (2012) emphasized that when children reached the stage of starting out in life, they would see that many tasks were done on computers and understand that they needed to think productively, and that their environment would expect them to make a difference in whatever their field was [4], [5].

Skill, defined as the ability to accomplish a task and complete a process appropriately, depending on the individual's predisposition and education, the ability to perform a task, mastery, or talent [1], also means the abilities that are designed to be acquired, developed, and transferred to life in students during the learning process [3]. In today's world, the skills that individuals need to actively and effectively participate in the information society are called 21st-century skills [6]. Many organizations, including international ones, such as the United Nations (UN), the Organization for Economic Development and Cooperation (OECD), the European Union (EU), as well as ministries responsible for education, many public or private institutions, non-governmental organizations, and large technology companies, participated in studies for creating a framework about the 21st-century skills that individuals should acquire in the new age [4]. These efforts show that society cares about the future of children, helps them achieve their personal goals, and provides what is necessary to empower them as autonomous citizens with the self-confidence to create their future. To sum up, while the environment, society, technology, and economic structure are constantly changing in the world, it is necessary to make regulations and changes to raise individuals according to 21st-century skills to keep up with this change [7].

The 21st-century skills initiative started in 2002 with P21 and EnGauge. P21, which was first founded under the name "Partnership for 21st-Century Skills" by the US Department of Education in 2002 and later became "Partnership for 21st-Century Learning," is a national organization in which many states are actively involved [8]. Many educational institutions, including the "National Councils of Mathematics, Science and English Teachers, the National Social Sciences Council, and the Geography Education Council," worked together in this organization to develop the skills that will prepare students for the 21st century. In addition to non-governmental organizations, technology companies from the business world, such as AOL Time Warner Foundation, Apple, Dell, Microsoft, and Cisco Systems Company, also played a role in shaping P21. Aiming to serve as a catalyst to place 21st-century skills at the center of US education by building collaborative partnerships among education, business, community, and government leaders, the P21 organization was founded on the belief that there was a deep gap between the knowledge and skills most students learned at school and the knowledge and skills needed in typical 21st-century communities and workplaces. The primary goal of P21, which focused on advocating for the inclusion of 21st-century skills in education, was to transform K-12 education in America and initiate a dialogue about the importance of these skills for all students [9], [10].

When we look at the studies in this field, it can be seen that the concept was named "21st-century skills" in North America and "21st-century competencies" in Europe [11], [12]. Starting from the premise of learning to collaborate with others and connecting with technology and with an emphasis on basic skills in a knowledge-based economy, more than 250 researchers from 60 institutions came together within the scope of the ATC21S project and classified 21st-century skills under four main headings [13].

- Ways of thinking: creativity, critical thinking, problem-solving, decision-making and learning,
- Ways of working: communication and collaboration,
- Tools of working: information and communication technology (ICT) and information literacy,
- Life skills: Citizenship, life and career, and personal and social responsibility

In another classification, 21st-century skills were classified as basic, change, and humanitarian knowledge. The content of these titles is as follows [11], [14]:

- Basic knowledge: high academic standards, mathematical and scientific competencies, core subjects (mathematics, language, science, social studies, etc.), quantitative literacy, discipline of mind, sophisticated knowledge in traditional knowledge, and core/framework education programs,
- Change knowledge: original thought, creativity, innovation, creative and critical thinking, creative mind, films, games, and design
- Humanitarian knowledge: ethical reasoning, empathy, ethical thinking and reflective thinking, managing emotions, emotional intelligence, and high ethical standards

Identification and acquisition of 21st-century skills and framework studies are given importance in Turkey. A Turkish Qualifications Framework was prepared to impart key competencies designed by the EU within the scope of the "2023 Vision", and in accordance with this, 21st-century skills were included in the curriculum in 2018 [7], [15]. Based on this context, it is necessary to examine the frameworks that include 21st-century skills in the world and Turkey and to determine the importance attached to these skills by taking into account the similarities and differences between these frameworks. According to EnGauge (2003), an organization leading the way in ensuring that future generations acquire the necessary skills for the 21st century, technology changes the way the world works. As technology evolves, the skills of those who use it must also evolve [10], [16]. To remain competitive tomorrow, today's students need to develop techniques to easily adapt to changes. This organization examined the 21st-century learning skills in four groups: "digital age literacy, creative thinking skills, effective communication, and high efficiency" [17], [18]. Accordingly, in terms of the characteristics one possesses, it is thought that 21st-century skills are must-have features also in the context of sports.

As a solution to the health problems faced by modern people, sports create a dynamic environment away from the stress of daily living. With the lifestyle it provides, it helps preventive medicine and supports people's social and mental development [19]. Since individual skills in sports are behaviors that are learned later rather than spontaneously, individuals need cognitive factors that provide opportunities for self-learning. While the acquisition of these skills requires cognitive alertness due to the coordination between muscles and the brain, their acquisition by individuals is not considered only within the scope of that skill. It also requires learning certain techniques, methods, and rules [20]. In addition to critical thinking and problem-solving skills, which are some of the 21st-century skills, sportive activities, and sports management also gain meaning with phenomena, such as cognitive flexibility, effective communication, information literacy, technology literacy, financial literacy, global competencies, reflective thinking, and lateral thinking, which have begun to be discussed and researched today [9], [10]. Therefore, it can be said that 21st-century skills are also required for sports management [21].

As a result of the literature review, it was seen that there was no study on personnel working in sports organizations regarding 21st-century skills. In this context, it is thought that integration into this age and acquisition of the skills required by the age by coaches, the indispensable actors of sports, and the personnel of the Ministry of Youth and Sports, the leading institution of sports in our country, will contribute greatly to the development of sports in the country. This study aimed to determine the level of 21st-century skills that the Ministry of Youth and Sports personnel and coaches have. It is thought that the study will contribute to the relevant literature in terms of allowing the participants to see their shortcomings and improve them.

The problem statement was determined as follows according to the purpose of the research:
1. Is there a statistical difference between the 21st-century skills and demographic variables of the Ministry of Youth and Sports personnel and coaches?
The sub-problems according to the research problem statement were identified as follows:
1. Is there a significant difference between the 21st-century skill scores of the Ministry of Youth and Sports personnel and coaches according to their status in their institution?
2. Is there a significant difference between the 21st-century skill scores of the Ministry of Youth and Sports personnel and coaches in terms of gender?
3. Is there a significant difference between the 21st-century skill scores of the Ministry of Youth and Sports personnel and coaches in terms of their marital status?
4. Is there a significant difference between the 21st-century skill scores of the Ministry of Youth and Sports personnel and coaches in terms of daily use of information technology tools?
5. Is there a significant difference between the 21st-century skill scores of the Ministry of Youth and Sports personnel and coaches in terms of their total work experience?
6. Is there a significant difference between the 21st-century skill scores of the Ministry of Youth and Sports personnel and coaches in terms of their age groups?

7. Is there a significant difference between the 21st-century skill scores of the Ministry of Youth and Sports personnel and coaches in terms of their education level?
8. Is there a significant difference between the 21st-century skill scores of Ministry of Youth and Sports personnel and coaches in terms of the department they graduated from?
9. Is there a significant difference between the 21st-century skill scores of the Ministry of Youth and Sports personnel and coaches in terms of their income levels?

2. Material and Method

This study was conducted to evaluate the 21st century skills of individuals working in the Ministry of Youth and Sports in terms of their status (coach-staff-manager) and some other variables. For this purpose, the data collected using the Multidimensional 21st-Century Skills Scale developed by Çevik and Şentürk (2019) and quantitative methods were analyzed [22].

2.1. Study Design

The survey model, one of the quantitative research methods, was employed. Survey research aims to describe the opinions of a large number of individuals [23]. According to the data collected from individuals who voluntarily participated in our survey within the scope of the research, the 21st-century skills were examined in terms of some variables.

2.2. Study Group

The study group comprises 369 personnel employed at the central organization of the Ministry of Youth and Sports, including 148 women and 221 men. A convenience sampling technique was employed to ensure that the sample accurately represents the overall population. The primary rationale for selecting this technique was to include all accessible individuals, including managers and employees, in the sample. In this context, the population of the study comprises 1,450 personnel working at the Ministry of Youth and Sports in Ankara. This group was determined using convenience sampling, one of the non-random sampling methods. Convenience sampling is a method that provides convenience in terms of time and cost and allows the researcher to recruit samples from the immediate surroundings that he or she knows and can easily reach [23].

Demographic data of the participants who constituted the study group of the research is given in Table 1.

Table 1. Distribution of participants' demographic characteristics

Demographic characteristics	Group	Frequency (F)	Percentage (%)
Gender	Female	148	40.1
	Male	221	59.9
Age	18-27	74	20.1
	28-37	153	41.5
	38-47	108	29.3
	≥48	34	9.2
Marital status	Single	135	36.6
	Married	234	63.4
Status	Coach	34	9.2
	Staff	218	59.1
	Manager	117	31.7
Seniority (years)	0-5	150	40.7
	6-10	85	23.0
	11-15	56	15.2
	16-20	39	10.6
	≥21	39	10.6
Education	High school	26	7.0
	Associate degree	94	25.5
	Undergraduate degree	207	56.1
	Graduate degree	42	11.4
Total		369	100

2.3. Data Collection Tools

In this section, details about the "Personal Information Form" and the "Multidimensional 21st-Century Skills Scale" developed by Çevik and Şentürk (2019), which were determined as quantitative data collection tools, were presented.

Personal Information Form: This form consists of eight questions about demographic variables (gender, age, marital status, status, seniority, education level, income level, and the duration of using information technology tools during working hours due to work).

Multidimensional 21st-Century Skills Scale: This scale, developed by Çevik and Şentürk (2019), consists of five dimensions and a total of 41 items [23]. The author's permission to use the scale was obtained. Items on the scale are evaluated on a five-point Likert-type structure with options ranging from "1: completely agree" to "5: completely disagree." The dimensions of the scale are "information and technology literacy skills, critical-thinking and problem-solving skills, entrepreneurship and innovation skills, social responsibility and leadership skills, and career consciousness." The alpha coefficient was found as 0.86 for the total of the original scale, 0.84 for the first dimension, 0.79 for the second sub-dimension, 0.76 for the third sub-dimension, 0.73 for the fourth sub-dimension, and 0.75 for the fifth sub-dimension. The alpha coefficients found in this study for the Multidimensional 21st-Century Skills Scale are shown in Table 2.

Table 2. Reliability coefficients of the total scale and its sub-dimensions

Sub-dimensions	Count of items	Cronbach's alfa
Information and technology literacy skills	15	0.90
Critical-thinking and problem-solving skills	6	0.73
Entrepreneurship and innovation skills	10	0.86
Social responsibility and leadership skills	4	0.60
Career consciousness	6	0.81
Total	41	0.92

As seen in Table 2, Cronbach's alpha values were high in all sub-dimensions of the scale. In addition, in the general reliability analysis of the scale, Cronbach's alpha value was calculated as 0.92. Considering that measurements with a reliability coefficient of .70 and above are reliable, it can be said that the calculated reliability coefficients were adequate except for Dimension 4. Test length is a factor that increases reliability, and reliability increases as the number of items on a test increases [24]. The low Cronbach's alpha value of the social responsibility and leadership skill subscale may be due to its small number of items.

2.4. Data Collection Process

In this research, which was a master's thesis study at Ankara Yıldırım Beyazıt University Faculty of Sports Sciences, the survey technique, which is the most frequently employed data collection technique, was utilized. The approval of the Ankara Yıldırım Beyazıt University Ethics Committee was obtained (date and decision number: 16.04.2021/26 and code: 2021-174). The survey was conducted between May 2021 and August 2021 after the necessary legal permissions had been obtained. Data were collected by sending the link of the questionnaire created on Google Forms to the participants working under the Ministry of Youth and Sports (coach-staff-manager), who constituted the study group of the research.

2.5 Data Analysis

Data obtained via the "Multidimensional 21st-Century Skills Scale", which was developed using a quantitative research design, were analyzed. The IBM SPSS Statistics Version 23 software package was utilized to analyze the data according to the research problems. First, the data were examined and invalid answers, missing values, and outliers were identified and the data were cleaned. Then, considering the determined sub-problems, scores on the total scale and each dimension were examined to find out whether they differed according to gender, branch, age, school type, seniority, education level, and working hours in the institution. The adequacy of the sample size and normal distribution of scores are necessary assumptions for parametric methods; therefore, they were tested. Considering the total and sub-dimension scores, whether the scores of the variables falling into each determined category showed a normal distribution was tested using the skewness and kurtosis coefficients. When examined together in each subcategory, data were found to be normally distributed, and therefore parametric methods were employed to compare the mean scores according to the variables considered. Before performing analysis according to parametric methods, the homogeneity of variances was checked with the Levene test. According to the Levene test result, the assumption of homogeneity of variances was provided. The independent samples t-test was utilized to compare the mean values of two groups [25]. Therefore, whether the items differed between the lower and upper groups was examined using the independent samples t-test. For the independent samples t-test, Cohen's *d* is determined by calculating the mean difference between two groups for effect size. One-way analysis of variance (ANOVA) and the

Scheffe test, one of the multiple comparison tests that show whether there is a difference between groups, were used to compare the means of more than two groups. The reason for using the Scheffe test is that the variances are homogeneous.

3. Results and Discussion

This section presents and discusses the findings of the research in relation to the research objectives and the relevant literature, supported by tables for a clearer representation of data. The results are analyzed in terms of the socio-demographic variables and 21st-century skills dimensions examined in the study. The implications of these findings are explored, with particular attention paid to their potential contribution to both theory and practice. In order to provide a broader understanding of the patterns observed, comparisons are made with previous studies. The significance of the results is evaluated within the context of existing knowledge in the field.

3.1. Descriptive Findings about the Scale

Descriptive statistics about the "Multidimensional 21st-Century Skills Scale" and its sub-dimensions used in the research are given in Table 3. As seen in Table 3, the sample of the research included 369 subjects. The mean score of the participants on the total Multidimensional 21st-Century Skills Scale was 170.95. According to the evaluation of the scale scores, the participants had "high level" 21st-century skills. The mean scores for the sub-dimensions were as follows: information and technology literacy skills, 65.62; critical thinking and problem-solving skills, 24.59; entrepreneurship and innovation skills, 38.59; social responsibility and leadership skills, 15.94; and career consciousness, 26.22. The kurtosis and skewness values were in the range of -1.5 and +1.5, which showed normality [26]. In addition, the data distribution was normal according to the graphic analysis.

Table 3. Descriptive statistics regarding the scale and sub-dimensions (N=369)

Variables		Min.	Max.	Mean.	SD	Skewness	Kurtosis
Multidimensional 21st-Century Skills Scale	Information and technology literacy skills	47	75	65.62	5.92	-.113	-.610
	Critical thinking and problem-solving skills	12	30	24.59	3.48	-.638	.828
	Entrepreneurship and innovation skills	21	50	38.59	5.51	-.167	.321
	Social responsibility and leadership skills	9	20	15.94	2.24	-.321	-.033
	Career consciousness	18	30	26.22	2.99	-.352	-.644
Total score		124	205	170.95	15.15	-.074	-.049

3.2. Findings about the First Sub-Problem

The mean scores of the participants on the total and sub-dimensions of the Multidimensional 21st-Century Skills Scale were compared according to their gender. For this purpose, the independent samples t-test was employed to test whether the total and subscale scores differed statistically significantly. The results are given in Table 4.

As seen in Table 4, there was no statistically significant difference between the mean 21st-Century Skills Scale sub-dimension and total scores according to gender (p>.05).

Table 4. Independent samples t-test results for the total and sub-dimension scores of the Multidimensional 21st-Century Skills Scale according to the gender variable

Variable	Gender	N	$\overline{X}$	SD	DF	t	p
Information and technology literacy skills	Female	148	65.74	5.71	367	0.322	0.748
	Male	221	65.53	6.06			
Critical thinking and problem-solving skills	Female	148	24.87	2.97	367	1.357	0.176
	Male	221	24.39	3.77			
Entrepreneurship and innovation skills	Female	148	38.50	5.02	367	-0.243	0.808
	Male	221	38.64	5.83			
Social responsibility and leadership skills	Female	148	16.06	2.06	367	0.843	0.400
	Male	221	15.86	2.36			
Career consciousness	Female	148	26.48	2.70	367	1.359	0.175
	Male	221	26.05	3.15			
Total score	Female	148	171.65	13.17	367	0.726	0.449
	Male	221	170.48	16.35			

3.3. Findings about the Second Sub-Problem

Participants' mean scores on the total and sub-dimensions of the Multidimensional 21st-Century Skills Scale were compared according to the age variable. For this purpose, analysis of variance (ANOVA), one of the parametric methods, was performed to determine whether there was a difference between the groups. The results are presented in Table 5.

As seen in Table 5, the age variable did not yield a statistically significant difference between participants' mean scores on the total and sub-dimensions of the Multidimensional 21st-Century Skills Scale ($p>.05$). Accordingly, it can be said that the results of the 21st -century skills were not effective in age groups.

Table 5. The ANOVA results for the total and sub-dimension scores of the Multidimensional 21st-Century Skills Scale according to the age variable

Variable	Age range	N	$\bar{X}$	SD	DF	t	p
Information and technology literacy skills	18-27	74	65.76	5.94	(3, 365)	0.508	0.677
	28-37	153	65.18	6.22			
	38-47	108	66.02	5.81			
	≥48	34	66.00	4.74			
Critical thinking and problem-solving skills	18-27	74	23.92	3.10	(3, 365)	1.288	0.278
	28-37	153	24.62	3.80			
	38-47	108	24.89	3.40			
	≥48	34	24.91	2.85			
Entrepreneurship and innovation skills	18-27	74	39.20	5.05	(3, 365)	0.546	0.651
	28-37	153	38.63	5.89			
	38-47	108	38.27	5.57			
	≥48	34	38.03	5.57			
Social responsibility and leadership skills	18-27	74	15.72	2.06	(3, 365)	1.035	0.377
	28-37	153	15.84	2.35			
	38-47	108	16.07	2.28			
	≥48	34	16.44	2.00			
Career consciousness	18-27	74	26.68	2.65	(3, 365)	0.760	0.517
	28-37	153	26.17	3.09			
	38-47	108	26.06	2.99			
	≥48	34	26.00	3.17			
Total score	18-27	74	171.27	12.87	(3, 365)	0.096	0.962
	28-37	153	170.44	16.41			
	38-47	108	171.31	15.31			
	≥48	34	171.38	13.76			

3.4. Findings about the Third Sub-Problem

The mean scores on the total and sub-dimensions of the Multidimensional 21st-Century Skills Scale were compared according to participants' marital status. For this purpose, the independent samples t-test was employed to test whether the subscale and total scores differed statistically significantly. Table 6 shows the results of the analysis. According to the results in Table 6, there was no statistically significant difference between the mean total and sub-dimension scores of the 21st-Century Skills Scale according to marital status ($p>.05$). In other words, all subscale and total scores of single and married individuals were similar.

Table 6. Independent samples t-test results for the total and sub-dimension scores of the Multidimensional 21st-Century Skills Scale according to the marital status variable

Variable	Marital status	N	$\bar{X}$	SD	DF	t	p
Information and technology literacy skills	Single	135	65.94	6.19	367	0.803	0.423
	Married	234	65.43	5.76			
Critical thinking and problem-solving skills	Single	135	24.33	3.42	367	-1.089	0.277
	Married	234	24.74	3.51			
Entrepreneurship and innovation skills	Single	135	38.61	5.86	367	0.078	0.938
	Married	234	38.57	5.31			
Social responsibility and leadership skills	Single	135	15.83	2.22	367	-0.720	0.472
	Married	234	16.00	2.26			
Career consciousness	Single	135	26.27	2.98	367	0.253	0.800
	Married	234	26.19	2.99			
Total score	Single	135	170.99	15.67	367	0.035	0.972
	Married	234	170.93	14.87			

3.5. Findings about the Fourth Sub-Problem

Comparisons were made using the analysis of variance (ANOVA), one of the parametric methods, to find out whether the scores on the total and sub-dimensions of the Multidimensional 21st-Century Skills Scale differed according to participants' status variable. The results are shown in Table 7.

As seen in Table 7, the total and all sub-dimension scores of the "Multidimensional 21st-Century Skills Scale" showed a statistically significant difference according to participants' status ($p < 0.05$). Pairwise comparisons were made using the Scheffe test, one of the post-hoc tests, to determine between which status levels a significant difference existed. As a result of pairwise comparisons made between the scores of the information and technology literacy skills, critical thinking and problem-solving skills, and social responsibility and leadership skills sub-dimensions, it was determined that the significant difference was between staff and managers. When the mean scores on these sub-dimensions were examined, it was observed that managers' scores were higher than the scores of staff. As a result of the pairwise comparisons made between the total scores, it was seen that the significant difference was between coaches and staff and between staff and managers. When mean total scores were examined, it was determined that the scores of coaches and managers were higher than those of staff. Finally, the result of the pairwise comparisons made between participants' mean scores on the career consciousness and entrepreneurship and innovation skills sub-dimensions, it was seen that the significant difference was between coaches and staff. The scores of the coaches were higher than those of staff.

Table 7. The ANOVA results for the total and sub-dimension scores of the Multidimensional 21st-Century Skills Scale according to the status variable

Variable	Status	N	$\bar{X}$	SD	DF	F	P	Significant difference
Information and technology literacy skills	Coach	34	67.35	5.84	(2,366)	6.987	0.001*	2-3
	Staff	218	64.68	5.94				
	Manager	117	66.85	5.58				
Critical thinking and problem-solving skills	Coach	34	23.88	4.35	(2, 366)	4.241	0.015*	2-3
	Staff	218	24.29	3.55				
	Manager	117	25.33	2.92				
Entrepreneurship and innovation skills	Coach	34	41.50	4.47	(2, 366)	8.406	0.000*	1-2
	Staff	218	37.77	5.53				
	Manager	117	39.26	5.40				
Social responsibility and leadership skills	Coach	34	16.41	2.16	(2, 366)	8.406	0.000*	2-3
	Staff	218	15.55	2.19				
	Manager	117	16.53	2.23				
Career consciousness	Coach	34	27.44	2.35	(2, 366)	3.399	0.034*	1-2
	Staff	218	26.02	3.02				
	Manager	117	26.25	3.01				
Total score	Coach	34	176.59	13.33	(2, 366)	8.747	0.000*	1-2/2-3
	Staff	218	168.31	14.91				
	Manager	117	174.22	15.11				

3.6. Findings about the Fifth Sub-Problem

The mean scores on the total and sub-dimensions of the Multidimensional 21st-Century Skills Scale were compared according to participants' total work experience. For this purpose, analysis of variance (ANOVA), one of the parametric methods, was performed to determine whether there was a difference between the groups. Table 8 shows the results of the analysis.

When Table 8 was examined, it was seen that there was no statistically significant difference between the total and information and technology literacy skills, social responsibility and leadership skills, and career consciousness sub-dimension scores of the 21st-century skills scale according to the length of service in the institution variable ($p > .05$). The mean critical thinking and problem-solving skills sub-dimension showed a statistically significant difference according to this variable ($p < .05$). As a result of the pairwise comparisons, it was seen that the significant difference was between those with 0-5 and those with 11-15 years of service. When the mean scores were examined, it was determined that the scores of those with 0-5 years of service on the critical thinking and problem-solving skills sub-dimension were higher than the scores of those with 11-15 years of service. According to Cohen effect size, a moderate effect was found (d=0,52).

Table 8. ANOVA results about participants' total and sub-dimension scores on the Multidimensional 21st-Century Skills Scale according to their total work experience in their institution

Variabable	Year	N	$\bar{X}$	SD	DF	F	p	Significant difference
Information and technology literacy skills	0-5	150	64.87	5.99	(4,364)	1.500	0.202	-
	6-10	85	66.53	6.02				
	11-15	56	66.30	6.08				
	16-20	39	64.95	6.52				
	≥21	39	66.15	4.08				
Critical thinking and problem-solving skills	0-5	150	24.06	3.34	(4,364)	2.588	0.037*	1-3
	6-10	85	24.73	3.95				
	11-15	56	25.73	3.10				
	16-20	39	24.33	3.72				
	≥21	39	24.90	2.83				
Entrepreneurship and innovation skills	0-5	150	38.57	5.13	(4,364)	2.070	0.084	-
	6-10	85	39.92	5.59				
	11-15	56	37.96	5.91				
	16-20	39	37.69	5.85				
	≥21	39	37.51	5.74				
Social responsibility and leadership skills	0-5	150	15.73	2.20	(4,364)	0.700	0.592	-
	6-10	85	16.12	2.37				
	11-15	56	15.89	2.33				
	16-20	39	16.10	2.09				
	≥21	39	16.26	2.20				
Career consciousness	0-5	150	26.25	2.96	(4,364)	0.370	0.830	-
	6-10	85	26.47	2.91				
	11-15	56	26.21	2.86				
	16-20	39	25.85	2.97				
	≥21	39	25.97	3.48				
Total score	0-5	150	169.49	14.34	(4,364)	1.346	0.252	-
	6-10	85	173.76	16.52				
	11-15	56	172.11	15.95				
	16-20	39	168.92	14.98				
	≥21	39	170.79	13.67				

3.7. Findings about the Sixth Sub-Problem

Participants' mean total and sub-dimension scores on the Multidimensional 21st-Century Skills Scale were compared according to their education levels. For this purpose, analysis of variance (ANOVA), one of the parametric methods, was conducted to determine whether there was a difference between the groups. The results are shown in Table 9.

Table 9 showed that there was no statistically significant difference between participants' mean sub-dimension and total scores on the 21st-Century Skills Scale according to their education levels ($p>.05$). Accordingly, it can be said that the results of 21st-century skills did not affect education levels.

Table 9. ANOVA results about participants' total and sub-dimension scores on the Multidimensional 21st-Century Skills Scale according to their education levels

Variable	Education level	N	$\bar{X}$	SD	DF	F	p
Information and technology literacy skills	High school	26	66.23	6.85	(3, 365)	0.186	0.906
	Associate degree	94	65.33	5.88			
	Undergraduate degree	207	65.70	5.87			
	Postgraduate degree	42	65.48	5.76			
Critical thinking and problem-solving skills	High school	26	24.50	4.22	(3, 365)	1.149	0.329
	Associate degree	94	24.23	3.49			
	Undergraduate degree	207	24.58	3.50			
	Postgraduate degree	42	25.43	2.78			
Entrepreneurship and innovation skills	High school	26	40.23	6.20	(3, 365)	1.064	0.364
	Associate degree	94	38.16	5.61			
	Undergraduate degree	207	38.66	5.33			
	Postgraduate degree	42	38.14	5.70			
Social responsibility and leadership skills	High school	26	16.38	2.16	(3, 365)	1.101	0.348
	Associate degree	94	16.16	2.08			
	Undergraduate degree	207	15.76	2.37			
	Postgraduate degree	42	16.05	2.00			

Table 9. (Continuous) ANOVA results about participants' total and sub-dimension scores on the Multidimensional 21st-Century Skills Scale according to their education levels

Career consciousness	High school	26	27.12	2.89	(3, 365)	1.374	0.250
	Associate degree	94	26.14	2.86			
	Undergraduate degree	207	26.05	3.07			
	Postgraduate degree	42	26.69	2.83			
Total score	High school	26	174.46	16.84	(3, 365)	0.635	0.593
	Associate degree	94	170.02	14.80			
	Undergraduate degree	207	170.76	15.38			
	Postgraduate degree	42	171.79	13.79			

3.8. Findings about the Seventh Sub-Problem

The mean total and sub-dimension scores on the Multidimensional 21st-Century Skills Scale were compared according to participants' income levels. To do this, analysis of variance (ANOVA), one of the parametric methods, was performed to determine whether there was a difference between the groups. Table 10 presents the results.

The examination of Table 10 showed that no statistically significant difference existed between participants' mean sub-dimension and total scores on the Multidimensional 21st-Century Skills Scale according to their income levels ($p>.05$).

Table 10. ANOVA results about participants' total and sub-dimension scores on the Multidimensional 21st-Century Skills Scale according to their income levels

Variable	Income level	N	$\bar{X}$	SD	DF	F	p
Information and technology literacy skills	₺ 0-3000	53	65.62	6.29	(3, 365)	0.029	0.993
	₺3001-5000	92	65.73	5.15			
	₺5001-700	89	65.66	6.06			
	≥₺7001	135	65.50	6.22			
Critical thinking and problem-solving skills	₺ 0-3000	53	23.89	3.47	(3, 365)	1.173	0.320
	₺3001-5000	92	24.73	3.56			
	₺5001-700	89	24.40	3.47			
	≥₺7001	135	24.88	3.42			
Entrepreneurship and innovation skills	₺ 0-3000	53	38.87	5.23	(3, 365)	1.627	0.183
	₺3001-5000	92	39.52	4.76			
	₺5001-700	89	38.47	5.64			
	≥₺7001	135	37.91	5.95			
Social responsibility and leadership skills	₺ 0-3000	53	15.83	2.20	(3, 365)	0.640	0.590
	₺3001-5000	92	15.78	1.97			
	₺5001-700	89	16.21	2.38			
	≥₺7001	135	15.91	2.35			
Career consciousness	₺ 0-3000	53	26.94	2.79	(3, 365)	3.313	0.070
	₺3001-5000	92	25.82	2.89			
	₺5001-700	89	26.76	3.06			
	≥₺7001	135	25.86	2.99			
Total scores	₺ 0-3000	53	171.15	14.02	(3, 365)	0.249	0.862
	₺3001-5000	92	171.58	12.73			
	₺5001-700	89	171.52	16.28			
	≥₺7001	135	170.07	16.38			

4. Discussion

When the findings about participants' 21st-century skills were examined in general, it was determined that their scores on the sub-dimensions of information and technology literacy skills, critical thinking and problem-solving skills, entrepreneurship and innovation skills, social responsibility and leadership skills, and career consciousness were high. In addition, participants' mean scores on the total scale were also at a high level. According to these results, it can be said that the personnel participating in the study thought that they had most of the 21st-century skills. It can be thought that the main reason behind participants' high scores on the 21st-century skills sub-dimensions was that there were significant relationships between these skills and that the results of sub-dimensions positively affected each other. The findings of the study supported this view. Similarly, some studies in the literature indicated that there was a significant relationship between 21st-century skills (entrepreneurship, creativity, innovation, problem-solving, and leadership) and that a high level of any skill in this context also affected other skill areas [27].

A study conducted among public sector employees aimed to identify findings related to 21st century skills. A total of 261 adults participated in the study, including 110 males and 151 females. At the end of the research, it was found that there were significant relationships between the participants' self-efficacy levels, their awareness of entrepreneurship and their problem-solving skills [28]. In a study conducted by Yeni (2015), it was reported that 21st century skills are among the fundamental factors that influence the leadership style of managers [29]. In a study conducted by Elekoğlu and Demirdağ (2020), it was reported that school administrators had high levels of 21st-century skills, and furthermore, there were significant relationships between the sub dimensions of 21st-century skills (leadership, communication skills) [30].

In recent years, analogous studies conducted particularly on adult individuals have revealed that their levels of possession of 21st-century skills are notably high. In one study, the objective was to examine the demographic factors influencing entrepreneurial intentions among adults. The study involved 433 adult participants. The results indicated that the participants exhibited high levels of entrepreneurial intentions, demonstrated a comprehensive understanding of entrepreneurship, and expressed a strong desire to establish their own businesses [31]. Another study focusing on adults aimed to examine the factors influencing entrepreneurial personality traits, with 76 adult participants. The findings demonstrated that the participants exhibited pronounced entrepreneurial tendencies and that engagement in entrepreneurship-related activities was a significant contributor to the development of entrepreneurial awareness [32].

A comparable study conducted on school administrators similarly revealed that the majority of participants demonstrated possession of 21st-century skills. The study demonstrated that the administrators exhibited notably elevated levels of proficiency in managing change and in implementing and evaluating change. The same study reported significant correlations between the learning and innovation skills of school administrators and their literacy skills [33]. A further study, conducted with public employees in managerial positions, revealed that the administrators demonstrated high levels of 21st-century skills, including learning and innovation skills (communication, collaboration, critical thinking and problem-solving, creativity and innovation), literacy skills (technological literacy, media literacy, information literacy), and life and career skills (social and cultural skills, flexibility and adaptability, leadership and responsibility, productivity and accountability, self-management and initiative) [34].

When the findings of this study are evaluated in conjunction with the results of analogous studies in the existing literature, it is typically observed that individuals in the workforce demonstrate elevated levels of 21st-century skills. Several underlying factors may explain these results. In particular, it can be posited that the pervasive utilization of internet tools in the personal development process and the ease with which individuals can access desired information at any time have contributed to these outcomes.

The results indicated that the findings pertaining to 21st-century skills exhibited some discrepancies based on the professional statuses of the personnel who participated in the study. The results revealed significant differences in favour of coaches in the sub-dimension of information and technology literacy, in favour of managers in the sub-dimension of critical thinking and problem-solving, in favour of coaches in the sub-dimension of entrepreneurship and innovation, in favour of managers in the sub-dimension of social responsibility and leadership skills, and in favour of coaches in the sub-dimension of career awareness. Moreover, it was established that coaches and managers exhibited the highest scores in the overall scale sub-dimensions when compared to other personnel.

A review of the literature reveals a paucity of studies that address the factors influencing 21st-century skills among Ministry of Youth and Sports employees. Furthermore, no studies have been conducted that specifically examine the 21st-century skills of coaches, personnel, or managers. The higher prevalence of 21st-century skills among coaches and managers in this study may be attributed primarily to the higher education levels of individuals in these professions, as well as their frequent encounters with various problems in their professional lives.

A review of the literature revealed no significant differences in the sub-dimensions of information and technology literacy, critical thinking and problem-solving, entrepreneurship and innovation, social responsibility and leadership skills, and career awareness when the findings related to the 21st-century skills of personnel were examined according to gender. Moreover, no significant differences were observed in the overall scale scores according to gender. Given that gender is a factor that influences individuals' social characteristics and personal skills, numerous studies in the literature have examined 21st-century skills in relation to the gender variable [35], [36].

It is evident that some studies in the existing literature demonstrate similarities with the findings of this research project. The objective of the study, which was conducted on students in the sports departments of universities, was to examine the leadership characteristics and entrepreneurial tendencies of the students in question, with a view to establishing the influence of certain demographic factors. A total of 700 students enrolled in sports departments participated in the study on a voluntary basis. Upon evaluation of the findings related to gender, it was determined that there was no significant difference in leadership characteristics according to the gender variable. However, entrepreneurial tendencies did exhibit a significant difference based on gender [37]. Another study in this field sought to examine the findings related to the entrepreneurial awareness of adult individuals studying in sports and recreation departments based on certain demographic factors. The views of 268 individuals were gathered within the scope of the research. The study concluded that there was no difference in entrepreneurial awareness between participants based on gender [38].

A synthesis of the findings from the extant literature with the results of the present study reveals that there are no significant differences between male and female employees in terms of their levels of 21st-century skills [28], [34], [31], [39], [40], [41]. One of the primary reasons for this may be that male and female employees have similar levels of education, encounter similar challenges in their professional lives, and exhibit comparable levels of professional self-efficacy. Conversely, studies that have identified differences in 21st-century skills according to gender may attribute these differences to the fact that such research has been conducted on individuals from different professions, each with distinct professional qualifications.

A comparison of the findings regarding participants' 21st-century skills in relation to the length of service variable indicated that there was no statistically significant difference between the mean scores on the information and technology literacy, entrepreneurship and innovation skills, social responsibility and leadership skills, and career consciousness sub-dimensions. No significant differences were observed in total scale scores according to the frequency of daily IT tool usage. Conversely, the mean scores for the critical thinking and problem-solving sub-dimensions demonstrated a statistically significant correlation with the length of service in the institution. It was thus established that personnel with 0-5 years of professional seniority exhibited the most pronounced critical thinking and problem-solving abilities. It may be postulated that the principal reason for this phenomenon was that personnel in this category demonstrated a proclivity for idealism and motivation, given that they were in the nascent stages of their professional careers.

A review of the literature on this subject reveals a discrepancy in findings regarding the variable of professional seniority. In a study conducted on public personnel, it was reported that the findings related to the sub-dimensions of 21st-century skills (entrepreneurship and problem-solving) did not differ according to the professional seniority of the public personnel [28]. In a separate study conducted on public personnel, it was determined that perceptions pertaining to all sub-dimensions of 21st-century skills exhibited variation according to professional seniority. In the aforementioned study, it was observed that the perceptions of personnel with a professional experience of between one and ten years were higher in the scale's sub-dimensions (creativity and innovation, critical thinking and problem-solving, communication and collaboration, information literacy, media and technology literacy) when compared to those with a higher level of professional seniority [42].

The examination of the effect of age variable on 21st-century skills showed that the scores on information and technology literacy, critical thinking and problem-solving, entrepreneurship and innovation skills, social responsibility and leadership skills, and career consciousness sub-dimensions did not show a significant difference. The total scale score did not differ significantly across age groups, either. It can be thought that the reason for this result was that although the participants were in different age groups, their professional qualifications were similar. In analogous studies conducted on disparate sample groups in the literature, analogous results to those of this study were obtained. In the study conducted by Kara (2018) on students enrolled in the sports departments of universities, it was determined that the entrepreneurial intentions and leadership characteristics of the students exhibited significant differences according to socio-demographic variables. Conversely, it was established that the leadership characteristics and entrepreneurial intentions of the participants did not differ significantly according to the age factor [37].

Upon examination of the findings pertaining to 21st-century skills in accordance with the variable of education level, it was observed that the mean total score and the scores for information and technology literacy, critical thinking and problem-solving, entrepreneurship and innovation skills, social responsibility and leadership skills, and career

consciousness did not exhibit a statistically significant difference according to this variable. It can be postulated that this outcome may be attributed to the fact that the majority of the participants had comparable education levels. In some studies, conducted on individuals in different occupational groups and age groups in the literature, it was reported that the findings related to the sub-dimensions of 21st-century skills did not differ according to the education level variable [43]. In a further study conducted on this subject, the aim was to compare the entrepreneurship awareness of associate degree and undergraduate students according to their educational status. The findings indicated that there was no significant difference according to the education level variable [44].

The investigation into the influence of income level on 21st-century skills indicated that there were no statistically significant differences in the total scores and scores for the sub-dimensions of information and technology literacy, critical thinking and problem-solving, entrepreneurship and innovation skills, social responsibility and leadership skills, and career consciousness. A review of the literature reveals that income level is a significant predictor of certain skills, including those related to entrepreneurship. This is particularly relevant given the role of financing in economic terms in the development of entrepreneurial activities. In a study conducted on adult individuals, the objective was to examine the demographic factors affecting the entrepreneurial intention, which is among the 21st-century skills. At the conclusion of the study, it was determined that family income level was an important determinant of the entrepreneurial intentions of the participants [45], [46], [47], [48]. This situation does not support the findings of our study, which may be attributed to the relatively narrow salary differential between employees of the Ministry of Youth and Sports.

5. Conclusion and Suggestions

In conclusion, it was found that the personnel participating in this study had a high level of 21st-century skills and that these skills differed according to some socio-demographic variables. In light of the findings obtained in this study and similar research findings in the literature on this subject, the following suggestions can be made for research and practice:

• Although the investigation of the 21st-century skills of the Ministry of Youth and Sports employees is important, it is also critical for the employees to demonstrate these skills in their professional lives. Determining the level of employees' exhibition of 21st-century skills is possible through the investigation of the opinions of other personnel in business life. In this context, to determine the level of 21st-century skill use by employees, especially by those in high-rank positions, in their professional lives, studies can be conducted to examine the level of managers' possession of 21st-century skills according to the opinions of employees.

• The findings of this study and similar study results in the literature showed that some demographic variables were effective in having 21st-century skills. However, these demographic variables cover more than those discussed in this study. In this context, studies can be conducted to examine the 21st-century skills of the Ministry of Sports employees according to different demographic factors (marital status, place of employment, department graduated, etc.).

• It is expected that having 21st-century skills will positively affect the parameters of business life. It is noteworthy that studies on the effects of 21st-century skills on professional life are generally limited in the field of sports. Therefore, studies including employees from the sports sector and addressing the relationship between 21st-century skills and factors in professional life (work motivation, job satisfaction, organizational commitment, organizational cynicism, organizational communication, job performance, and leadership styles) can be conducted.

• Professional characteristics of personnel may differ depending on their status and the units where they work. At this point, status and profession variables are expected to be decisive on 21st-century skills. In this context, studies can be conducted to address 21st-century skills according to the professional status of employees and the units they work in.

• It is not possible to generalize the findings of this research to the entire business life due to the number of samples. In this context, similar studies with large sample groups involving individuals from different professions can be conducted.

• Considering that the level of employees' possession of 21st-century skills will positively affect their work, family, and social lives, in-service training courses can be organized for the personnel of the Ministry of Youth and Sports to develop these skills.

• Employees' 21st-century skills are generally shaped by their professional experience, social environment, and education. In addition, it is known that these skills can be acquired at every age. In this context, employees can be given training to develop 21st-century skills. By comparing the differences between these skills at the beginning and end of the training, the effectiveness of the training provided can be determined.

Acknowledgment

We would like to thank the reviewers and the editorial board who were involved in the publication process.

Contributions of the authors

Canan SAYIN TEMUR: Conceptualization, Research Design, Editing, Supervision, Manuscript Writing, Review, Final Approval.
Fatih HANOĞLU: Literature Review, Data Collection, Data Analysis.

Conflict of Interest Statement

There is no conflict of interest between the authors regarding this article, which is derived from a master's thesis.

Statement of Research and Publication Ethics

This research, conducted as a master's thesis at the Faculty of Sport Sciences, Ankara Yıldırım Beyazıt University, has received approval from the Ethics Committee of Ankara Yıldırım Beyazıt University and complies with research and publication ethics (date and decision number: 16.04.2021/26, code: 2021-174).

References

[1] YÖK, "Türkiye Yükseköğretim Yeterlilikler Çerçevesi Yönetmelik Taslağı," [Online]. Available: http://tyyc.yok.gov.tr/?pid=101&duyuruid=357. [Accessed: March 20, 2021].

[2] M. Yılmaz, "Öğrenme ve Bilgi İlişkisi," *Gazi University Journal of Gazi Education Faculty*, vol. 29, no. 1, pp. 173-191, 2009.

[3] R. J. Marzano and T. Heflebower, Teaching and assessing 21st century skills. [Online]. Available: https://ebookcentral.proquest.com/lib/gazi-ebooks/reader.action?docID=3404970&query=. [Accessed: March 22, 2021].

[4] S. Keskin, *İnovasyon Nasıl Yapılır? Yenilikçi Fikir ve Ürün Arayanlara Pratik Modeller*. İstanbul: Mavi, 2012.

[5] S. S. Seferoğlu and C. Akbıyık, "Eleştirel Düşünme ve Öğretimi," *Hacettepe University Journal of Education*, vol. 30, pp. 193-200, 2006.

[6] F. Güneş, "Bologna Süreci ile Yükseköğretimde Öngörülen Beceri ve Yetkinlikler," *Journal of Higher Education and Science*, vol. 2, no. 1, pp. 1-9, 2012.

[7] E. Yılmaz, *21. Yüzyıl Becerileri Kapsamında Dönüşen Okul Paradigması*, Eğitim Bilimlerinden Yansımalar, Konya: Çizgi, 2012, pp. 8-11.

[8] C. Fadel, M. Bialik and B. Trilling, *"Four-dimensional education: The competence learners need to succeed."* Boston: The Center for Curriculum Redesign, 2015.

[9] E. Wright and M. Lee, "Developing skills for youth in the 21st century: The role of elite International Baccalaureate Diploma Programme schools in China," *Springer Int*. Rev. Educ., vol. 60, pp. 199-216, 2014.

[10] S. Woods-Groves, "The Human Behavior Rating Scale–Brief: A Tool to Measure 21st Century Skills of K–12 Learners," *Psychological Reports*, vol. 116, no. 3, pp. 769-796, 2015.

[11] A. Schleicher, *Preparing Teachers and Developing School Leaders for The 21st Century: Lessons from Around the World.* Paris: OECD Publishing, 2021.

[12] National Research Council, *"Assessing 21st Century Skills: Summary of A Workshop."* National Academies Press, 2021.

[13] OECD, "Preparing teachers and developing school leaders for the 21st century: Lessons from around the world," [Online]. Available: https://www.oecd.org/site/eduistp2012/49850576.pdf. [Accessed: March 28, 2021].

[14] K. Kereluik, P. Mishra, C. Fahnoe and L. Terry, "What knowledge is of most worth: Teacher knowledge for 21st century learning," *Journal of Digital Learning in Teacher Education*, vol. 29, no. 4, pp. 127-140, 2013.

[15] Ş. D. Belet Boyacı and M. Güner Özer, "Öğrenmenin Geleceği: 21. Yüzyıl Becerileri Perspektifiyle Türkçe Dersi Öğretim Programları," *Anadolu Journal of Educational Sciences International*, vol. 9, no. 2, pp. 708-738, 2019.

[16] EnGauge, enGauge 21st Century Skills., 2003. [Online]. Available: https://www.cwasd.k12.wi.us/highschl/newsfile1062_1.pdf. [Accessed: March 19, 2021].

[17] D. R. Fisher, A. Bagiati and S. E. Sarma, "Fostering 21st century skills in engineering undergraduates through co-curricular involvement paper," 2014 ASEE Annual Conference & Exposition, Indianapolis, Indiana, 2014, [Online]. Available: https://peer.asee.org/20514. [Accessed: March 15, 2021].

[18] International Society for Technology in Education, "About ISTE," [Online]. Available: https://www.iste.org/standards/iste-standards-for-students. [Accessed: March 19, 2021].

[19] D. G. Doğan, "Sporda başarıyı etkileyen sosyal faktörler üzerine bir araştırma," M.S. thesis, Sakarya University, Institute of Social Sciences, Sakarya, Türkiye, 2001.

[20] G. Özdemir and G. Ersoy, "Engelli Sporcularda Beslenme, Sağlık ve Performans Etkileşimi," *Turkish Journal of Physical Medicine and Rehabilitation*, vol. 55, no. 3, 2009.

[21] M. Gündüz, *21. Yüzyılda Eğitim ve Öğretmen Yeterlikleri*. Ankara: Pegem Akademi, 2018.

[22] M. Cevik and C. Senturk, "Multidimensional 21st century skills scale: Validity and reliability study," *Cypriot Journal of Educational Sciences*, vol. 14, no. 1, pp. 11–28, 2019.

[23] Ş. Büyüköztürk, E. K. Çakmak, Ö. E. Akgün, Ş. Karadeniz, and F. Demirel, *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi, 2014.

[24] Y. Baykul, *Eğitimde ve psikolojide ölçme: klasik test teorisi ve uygulaması*. Ankara: Pegem Akademi Yayıncılık, 2015.

[25] M. Çakan, S. Çelikten, and T. Gündüz, "Nicel veri analizi ve yorumlanması," in *Kuramdan Uygulamaya Eğitimde Bilimsel Araştırma Yöntemleri*, B. Oral and A. Çoban, Eds. Ankara: Pegem Akademi, 2020, pp. 161–203.

[26] B. G. Tabachnick, L. S. Fidell, and J. B. Ullman, *Using multivariate statistics*, 6th ed. Boston, MA: Pearson, 2013.

[27] A. Carmeli, Y. Friedman, and A. Tishler, "Cultivating a resilient top management team: The importance of relational connections and strategic decision comprehensiveness," *Safety Science*, vol. 51, no. 1, pp. 148–159, 2013.

[28] N. Kavukcu, "The perceptions of healthcare professionals serving migrant, refugee and asylum seekers in the primary healthcare centers of Hatay, Şanlıurfa, Gaziantep and İzmir about healthcare provision, and their level of burnout," M.S. thesis, Social Sciences Institutes, Hacettepe University, Ankara, 2021.

[29] G. Yeni, "21. yüzyıl becerilerinin okul yöneticilerinin liderlik yönlerine olan etkisi," *Uluslararası Liderlik Eğitimi Dergisi (ULED)*, 2015. [Online]. Available: https://dergipark.org.tr/tr/download/article-file/1238753. [Accessed: March 19, 2021].

[30] F. Elekoğlu and S. Demirdağ, "Okul müdürlerinin 21. yüzyıl becerileri, iletişim becerileri ve liderlik stillerinin öğretmen algılarına göre incelenmesi," *Karaelmas Journal of Educational Sciences*, vol. 8, no. 1, pp. 101–117, 2020.

[31] J. Gerente, M. de Morisson Valeriano, and E. P. Moreira, "Regionalização de variáveis geomorfométricas para o mapeamento dos domínios morfoestruturais da bacia hidrográfica do rio Itajaí-Açu (SC)," *Revista Brasileira de Geomorfologia*, vol. 19, no. 3, 2018.

[32] M. Damar, "Study on the validity and reliability of the parent as secure base scale–mother version," *Milli Eğitim Dergisi*, vol. 47, no. 218, pp. 251–262, 2018.

[33] S. Yörük and D. Güler, "The relationship between psychological resilience, burnout, stress, and sociodemographic factors with depression in nurses and midwives during the COVID-19 pandemic: A cross-sectional study in Turkey," *Perspectives in Psychiatric Care*, vol. 57, no. 1, pp. 390–398, 2021.

[34] T. Ulgen, "Home care practices of cancer caregivers and the effect of perceived social support on care burden," 2021. [Online]. Available: https://www.researchsquare.com/article/rs-1055812/latest.pdf. [Accessed: Oct. 17, 2024].

[35] N. Atabey and M. S. Topcu, "The relationship between Turkish middle school students' 21st century skills and STEM career interest: Gender effect," *Journal of Education in Science Environment and Health*, vol. 7, no. 2, pp. 86–103, 2021.

[36] Á. B. Alfonso Águila, M. Calcines Castillo, R. Monteagudo de la Guardia, and Z. Nieves Achon, "Estrés académico," *Edumecentro*, vol. 7, no. 2, pp. 163–178, 2015.

[37] A. Kara, "Yükseköğretim kurumlarında spor eğitimi alan öğrencilerin girişimcilik eğilimi ve liderlik özellikleri üzerine bir araştırma," M.S. thesis, Institute of Social Sciences, İstanbul University, İstanbul, Türkiye, 2018.

[38] Ç. Özmen, "Rekreasyon ve spor yöneticiliği bölümü öğrencilerinin girişimcilik düzeylerinin değerlendirilmesine yönelik bir çalışma," M.S. thesis, Institute of Educational Sciences, Mehmet Akif Ersoy Üniversity, Burdur, Türkiye, 2015.

[39] T. Yörük, "Okul yöneticileri ve öğretmenlerin stratejik planlama süreçlerine yönelik algıları," M.S. thesis, Institute of Educational Sciences, Pamukkale University, Denizli, Türkiye, 2021.

[40] E. Kazaferoğlu, "Kişilik özellikleri ile girişimcilik eğilimi ilişkisi: Süleyman Demirel Üniversitesi öğrencileri üzerine bir araştırma," M.S. thesis, Institute of Educational Sciences, Süleyman Demirel University, Isparta, Türkiye, 2017.

[41] M. Özer, "Eğitim politikalarında sistematik uyum," M.S. thesis, Sosyal Bilimler Enstitüsü, Maltepe University, İstanbul, Türkiye, 2021.

[42] E. Kazak and S. Çiner, "Undesired behaviors of school principals and the effects of these behaviors on teachers," *Turkish Online Journal of Qualitative Inquiry*, vol. 12, no. 2, pp. 300–328, 2021.

[43] L. Özbay, "Üniversite öğrencilerinin girişimcilik eğilimlerinin belirlenmesi: Dumlupınar Üniversitesi'nde bir araştırma," M.S. thesis, Institute of Social Sciences, Dumlupınar University, Kütahya, Türkiye, 2016.

[44] S. A. Solmaz, Ö. Aksoy, S. Şengül, and M. Sarıışık, "Üniversite öğrencilerinin girişimci kişilik özelliklerinin belirlenmesi: Turizm lisans ve ön lisans öğrencileri üzerine bir alan araştırması," *Karamanoğlu Mehmetbey University Journal of Social and Economic Research***,** no. 1, pp. 41–55, 2014.

[45] F. Ö. Karataş, İ. Delen, C. Cengiz, İ. K. Nilgün, and S. Birinci, "Onuncu sınıf öğrencilerinin bilimsel süreç beceri düzeylerinin çeşitli değişkenler açısından incelenmesi," *Van Yüzüncü Yıl University Journal of Education*, vol. 15, no. 1, pp. 468–494, 2018.

[46] G. M. Abdullayeva, "Developing research paper writing for EFL/ESL undergraduate students," ELMİ İŞ, no. 117, 2007.

[47] A. Karadaş, S. Kaynak, S. Ergün, and P. P. Karaca, "Hemşirelik ve ebelik öğrencilerinin 21. yüzyıl becerilerinin bazı değişkenlere göre incelenmesi," *Ordu University Journal of Nursing Studies*, vol. 4, no. 2, pp. 232–239, 2021.

[48] D. Özdemir Özden, E. Karakuş Tayşi, Ş. Kılıç Şahin, S. Demir Kaya, and F. Ö. Bayram, "Öğretmen adaylarının 21. yüzyıl becerilerine yönelik yeterlik algıları: Kütahya örneği," *Electronic Turkish Studies*, vol. 13, no. 27, 2018.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1139-1146 YEAR: 2024
DOI: 10.17798/bitlisfen.1521246

Micromorphology of the proboscis sensilla of *Maniola jurtina* (Linnaeus, 1758) (Nymphalidae: Satyrinae)

Selma Seven ÇALIŞKAN[1*], Yağmur MENGI[1], Selami CANDAN[1]

[1]*Gazi University, Faculty of Science, Department of Biology, Ankara, Türkiye*
(ORCID: 0000-0003-4406-6768), (ORCID: 0000-0002-3794-7121), (ORCID: 0000-0002-7402-1360)

Keywords: Scanning electron microscopy, ultrastructure, sensilla, proboscis, Satyrinae.

Abstract

Lepidoptera is the most suitable insect order to comprehend the connections between nutrition and mouthparts. The proboscis, which is typically tightly coiled between the labial palps, is a flexible organ used by most adult butterflies for extracting flower nectar and other liquids. We examined the proboscis structure and sensilla of *Maniola jurtina* Linn. by scanning electron microscope (SEM) in order to contribute to the relationship between food selection and proboscis structure in Satyrinae Sensilla chaetica (s. trichodea) of various lengths is the most common species along the proboscis. Sensilla styloconica is of pluricarinate type. It is densely located only in the distal region and is in rows. It has a long quill with 6 ribs (flat protrusions) and a relatively short nail. In the study, sensilla lengths were compared with the sensilla lengths of some other lepidopteran species.

1. Introduction

Lepidopteran proboscis is one of the most remarkable body parts among insect groups in terms of nutrition. The diversity and structure of lepidopteran proboscis are intricately linked to their feeding behaviors and their capacity to ingest fluids [1-5]. While many species are known to be anthophilous, some also exhibit a varied diet, including mud puddles, sugary substances, sweat, tears, and wound exudates from mammals [6-13]. The proboscis, a flexible tube specialized for liquid food intake, typically adopts a coiled ring shape at rest. However, during feeding, it unfurls, revealing distinct regions. The proximal region is closest to the head, while the bend region emerges at approximately one-third of the proboscis's length [14]. The distal region, which is relatively lengthy, features a flexible apex containing several slits responsible for food intake [15, 16]. In this region, approximately 10–20% features dorsal legulae modifications characterized by wider interlegular spaces and reduced galeal surface area. These adaptations aid in fluid passage into the alimentary canal, contributing to an overall hydrophilic profile [3-4]. The proboscis consists of a pair of c-shaped galeae. Each galea is equipped with cuticular processes and several types of sensilla that are essential for nutritional activities [1, 17]. These sensilla vary depending on feeding habits [5, 18-20]. The form and function of the lepidopteran proboscis have been studied in various taxa encompassing numerous families [10, 15, 21-24].

There are six sensilla types in butterflies [1, 25]. Three of the most commonly seen sensilla types are: s. chaetica (=trichodea), s. basiconica and s. styloconica. Sensilla chaetica is located on the entire galea surface. Acting as a mechanoreceptor, this sensilla provides information about the width and depth of the tubular flower during flower scanning [26]. Sensilla basiconica are composed of a sensory cone, either domed or pointed, and a shallow, flexible socket [1, 27]. Present across various insect groups and distributed throughout the body, they serve as essential structures for chemosensory functions [28]. Sensilla basiconica exist in two forms: external and inner sensilla basiconica. Inner sensilla basiconica are notably fewer in number compared to those on the outer surface and are typically confined to the distal part of the proboscis [29]. The most characteristic sensilla type in lepidopterans is sensilla stylochonica.

*Corresponding author: selma@gazi.edu.tr
Received: 23.07.2024, Accepted: 04.10.2024

They are outstanding chemo-mechanic receptors limited to the apex region [27]. It is the sensilla type in which the most differentiation is observed [25]. The potential systematic importance of these senses in butterflies have been acknowledged by some researchers (example Borner, [30]) yet this issue has not been evaluated enough in systematic aspects. In their examination of the proboscis, Paulus and Krenn [16] documented 19 European satyrid species, a number of which are also present in Turkey. In a study about proboscises of butterflies conducted in North America, proboscis of 17 Satryinae species was examined [26]. In Turkey, there are only 3 publications regarding Lepidoptera proboscis [24, 31-32].

Given their biology, diversity, and distribution, Satyrinae constitutes a prominent group within butterfly communities [33-40]. The extensive diversity of Satyrinae, both in terms of species richness and morphology, has resulted in considerable uncertainty and taxonomic challenges in their classification. Comparative morphological studies on sensilla located in different places in insects may contribute to classification.

This research involves a detailed SEM examination of the proboscis morphology and sensilla types, and mouthparts in specimens of *Maniola jurtina* from Turkey. Sensilla have been compared with other lepidopter species and discussed later. The results contribute to the proboscis morphology of Satyrinae. This will be beneficial for studies that are intended to be conducted later related to the nutritional behaviors of adult lepidopters.

2. Material and Method

In this study, the proboscis of two male specimens of *M. jurtina* were examined. The specimens taken from Prof. Dr. Çalışkan's collection are kept in Gazi University Zoology Museum. The ultrastructure of the sensilla of the proboscis in adults of *M. jurtina* is elucidated using SEM imaging. For the terminology used here in proboscis and sensilla morphology, see Bänzigerl [9, 41], Büttiker *et al.* [8], Speidel *et al.* [23], Altner and Altner [27], Faucheux [42, 43], Hallberg [44] remove. For the scanning electron microscope imaging (SEM), mouthparts and proboscises of *Maniola jurtina* individuals have been cleaned with a brush and dried by air before being positioned with a double-sided tape on SEM tabs. They were later coated with a layer of gold using a Polaron SC 502 Sputter Coater and examined with a 10 kV Jeol JSM 6060 LV SEM. While stretching the butterflies, the proboscis of the softened samples were lengthened and fixed with the help of a needle. Sensilla lengths were measured with references taken from the beginning and end points of sensilla in SEM images. The example examined in the study is from the second author's master's thesis.

3. Results

The components of the mouthparts of *Maniola jurtina* are shown in Figure 1. The head of the individual shown in Figure 1a is damaged. The proboscis is four spirals (Fig. 1a). Galeae are hinged by dorsal and ventral legulae. Long bristle-shaped scales project from the pilifer and contact the nearby dorsal galeal surface. The maxillary palpus is fist-shaped next to the proboscis base (Fig. 1b). The proboscis is roughly threefold compared to the labialy palpus. The galeal surface has a rough texture, lacks spines, and has a spiny web-like structure near the dorsal ligulae. There are three types of sensilla in *Maniola jurtina*: Sensilla basiconica, sensilla styloconica, and sensilla chaetica (Figs. 2, 3, 4). Sensilla chaetica can be observed in various lengths and scattered on lateral and ventral sides in the proximal and distal regions of the galea (Figs. 2a, 2b, 3a). Sensilla basiconica can be observed in irregular rows on the external surface of the proboscis (Figs. 2a, 2b, 3a). Sensilla styloconica are in distal region of the proboscis. There has a long stylus with 6 ribs (Smooth Ridges) and a relatively short peg. Tips of the ribs bear six (Shoulder Spines) sharp spines. Length of the shoulder spines are shorter than the peg (Fig. 3b). Sensilla basiconica consist of a conical sensory structure with a flat surface surrounded by a shallow socket. Inner sensilla basiconica form a single row on the food canal (Fig. 4a). They consist of a long stylus and a short sensory nail placed in a slightly dome-shaped socket (Fig. 4b). The sensilla were not seen in the maxillary palps in the samples examined. However, pits on the palps attract attention. These pits are likely to be sensilla pits. The sensilla on the pits may have been severed because the head of the specimens was damaged.

4. Discussion

Most Satyrinae species are diurnal adults with limited dispersal capabilities, typically flying close to the ground, particularly in shaded forest areas (understory) [45-48]. Host plants of Satyrinae are mostly monocots. They are also used with certain eudicot families such as Fabaceae and Menispermaceae, and some species have been recorded feeding on Lycopodiophyta (Selaginellaceae), Bryophyta (Neckeraceae) and gymnosperms (Cycadaceae) [49-52]. Adults of most species are polyphagous and feed on the nectar of different species.

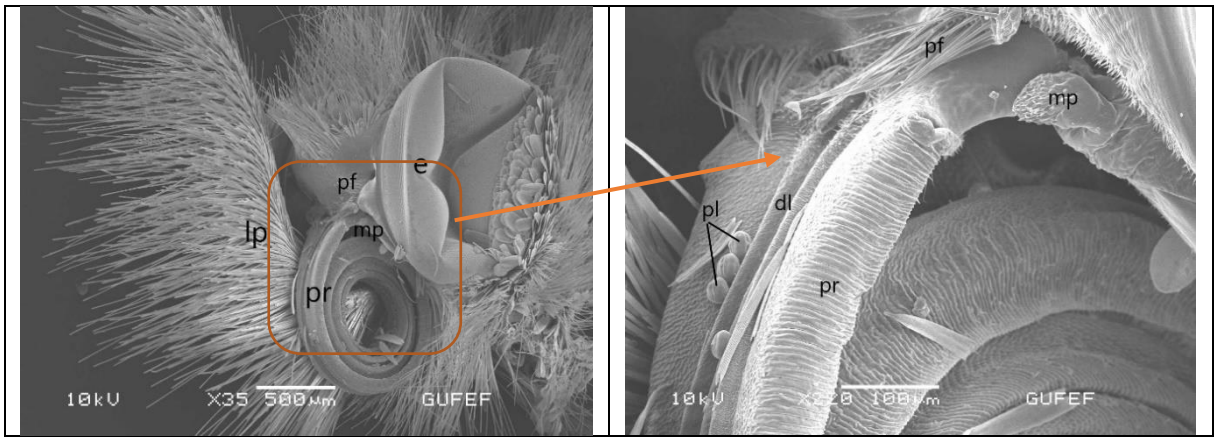


Figure 1a-b. Head and mouthparts of *Maniola jurtina* in oblique frontal view, **a.** proboscis (pr) in recoiled resting position ı: complex eye (e), pilifer (pf), maxillar palpus (mp), labial palpus (lp), **b.** Basal proboscis each proboscis (pr), dorsal legulae (dl), pollen (pl)

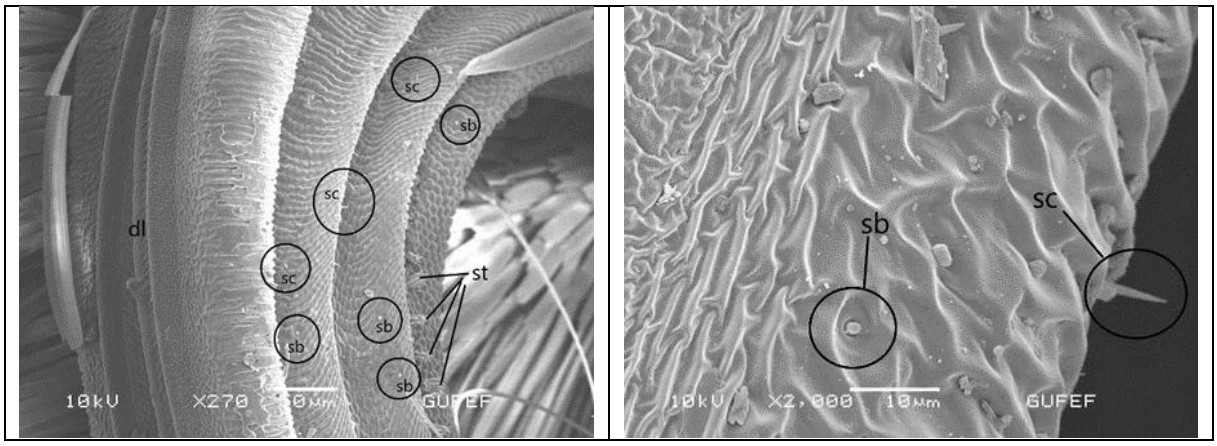


Figure 2a-b. Galea surface in *M. jurtina*, and the distribution of sensilla positioned on the galea surface, **a.** sensilla chaetica (sc), sensilla basiconica (sb), sensilla styloconica (st), **b.** Enlarged image of s. chaetica (sc), s. basiconica (sb)

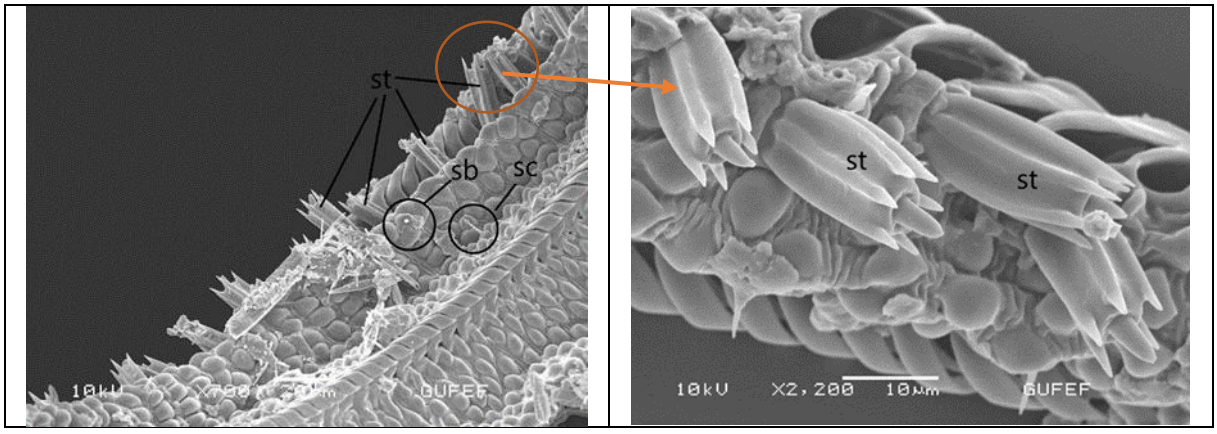


Figure 3a-b. Distal region of proboscis of *M. jurtina*, **a.** sensilla chaetica (sc), sensilla basiconica (sb), sensilla styloconica (st), **b.** Enlarged image of sensilla styloconica

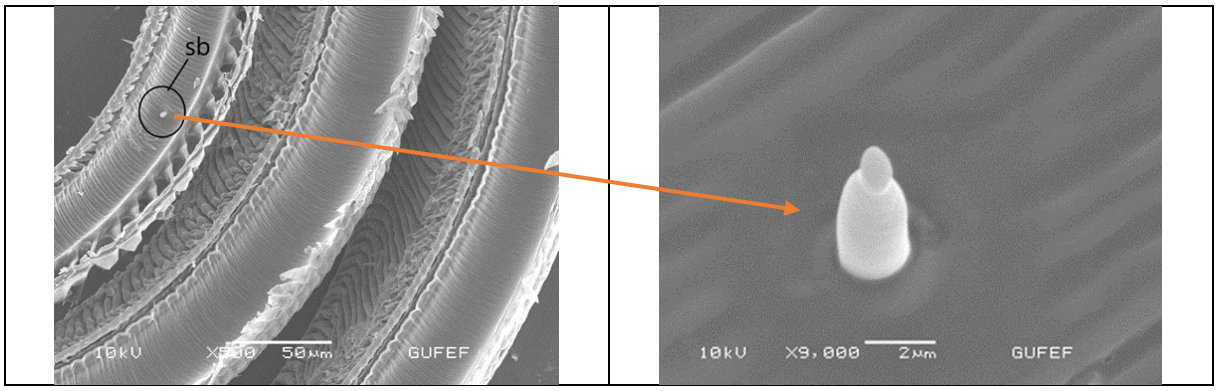


Figure 4a, b. Food canal of the proboscis of *M. jurtina*, **a.** intrasensilla basiconica (sb), **b.** Enlarged image of sensilla basiconica.

Krenn *et al.*, [53] examined the sucking proboscis of six satyrid butterfly species, including *Maniola jurtina,* in their study of the proboscis structure of nymphalids. Their study revealed variations in the sensilla equipment, particularly in the length of sensilla chaetica and the shape and quantity of sensilla styloconica. Notably lengthy sensilla trichodea were observed extending beyond the midproboscis in numerous Heliconiini (as detailed in Krenn & Penz, [54]), certain Limenitidinae, some Brassolinae, all Morphinae, and *Haetera piera* (Satyrinae). In Morphinae and *Haetera piera*, stout sensilla trichodea were predominantly situated on the ventral sides of the galeae, spanning from the base of the proboscis to just short of the tip-region. In the samples we examined, needle-shaped extensions on the ventral sides of the galea are noteworthy. Sensilla caetica are remarkably long, varying in size from 11.7 to 25.1 µm. S. chaetica length of *Aporia crataegi* ranges from 2.77 to 22.5 µm [31], ~4 µm in *Tirumala limniace* and in *Iphiclides podalirius*, ~6 µm [55] in *Vanessa cardui*, 8–48 µm [56]. The size of chaetica varies even in the same individual.

The sensilla basiconica, consisting of a broad dome-shaped base and a short sensory cone, serves as a chemoreceptive [16, 25, 26, 28, 56]. Sensilla basiconica in the alimentary canal are sensitive to mechanical stimuli and likely function to control fluid selection within the digestive tract [29]. Fluid uptake is probably faster in individuals with short inner sensilla length. Sensilla basiconica of *M. jurtina* is 6 µm long. Sensilla basiconica of *A. crataegi* is 3 µm long [31], 1.6 µm in *Homoeosoma electellum* (Hulst, 1887), 4 µm in *Monopis crocicapitella* (Clemens, 1859) and 10 µm in *Tineola bisselliella* (Hummel, 1823) [25]. According to the data, sensilla basiconica length is not related to body height [29]. The fact that the length of the inner sensilla basiconica in *M. jurtina* is twice as long as that of *A. crataegi* supports this.
Sensilla styloconica can provide the insect with important information such as the location of nectar and the length of the pollen tube [55, 54]. Sensilla styloconica increases the hydrophilicity and capillarity of the drinking area and helps fluid uptake from porous surfaces [3, 57]. Petr and Stewart [26] divided and named s. styloconica into groups according to their morphological structures. According to this study, sensilla styloconica of satyrids are divided into two groups: pluricarinate and pluridentate. Cylindriform/pluricarinate and cylindriform/pluridentate sensilla are unique identifiers for some groups of Satyrinae. Cylindrical/pluricarinate styli are specific to *Coenonympha*, *Cercyonis*, *Neominois*, and *Oeneis*; Cylindrical/pluricarinate styli are found only in *Erebia*, *Gyrocheilus* and *Oeneis*. The *M. jurtina* we examined also are the s.styloconica cylindriform/ pluricarinate type. The protrusions on the body of the cylindrical pencil are flat. They extend beyond the shoulders to the apical spines. Six apical shoulder spines form a symmetrical, rounded crown around the sensory nail.

Sensilla styloconica were observed to be longer and more abundant in non-flower-visiting nymphalids compared to flower-visiting species. Among species that visit flowers, there exists greater diversity in the shapes of sensilla styloconica. In Satyrinae, they feature longitudinally ribbed ends with apical spines, whereas in certain Nymphalinae members, they are either smooth and partially cylindrical (e.g., Ithomiinae) or smooth and flattened (e.g., Heliconiinae) [1]. The morphology of the sensilla styloconica in the examined sp. *jurtina* shows the characteristics of the satyrid group. The stylus has a cylindrical structure, unlike the either smooth and partially cylindrical sensilla styloconica seen in some members of the Nymphalinae group. The ribs on the body of the scion are smooth and have apical spines, 6 in number. These apical spines are arranged symmetrically to form a rounded crown around the sensory cone. The senslla styloconica of *M. jurtina*, studied by Krenn *et al*. [1], similarly has longitudinal projections ending in apical spines.

5. Conclusion and Suggestions

The results obtained show that Satyrinae are closer to Nymphalinae than Limenitiidae with their sensilla styloconica morphology. The usability of differences in the morphology of sensilla styloconica in the classification of higher categories in Satyrinae should be investigated.

Contributions of the Authors

S.S.Ç and Y.M. designed the study and wrote the article, S.C. contributed to Scanning Electron Microscope (SEM) images and writing the article.

Conflicts of Interest Statements

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] H. W. Krenn, "Feeding mechanisms of adult Lepidoptera: structure, function, and evolution of the mouthparts," *Annual Review of Entomology*, vol. 55, pp. 307–327, 2010, doi: 10.1146/annurev-ento-11240,085338.

[2] J. M. Zaspel, S. J. Weller, and M. A. Branham, "A comparative survey of proboscis morphology and associated structures in fruit-piercing, tear-feeding, and blood-feeding moths in Calpinae (Lepidoptera: Erebidae)," *Zoomorphology*, vol. 23, no. 3, pp. 203–225, 2011, doi: 10.1007/s00435-011-0132-1.

[3] M. S. Lehnert, D. Monaenkova, T. Andrukh, C. E. Beard, P. H. Adler, and K. G. Kornev, "Hydrophobic-hydrophilic dichotomy of the butterfly proboscis," *Journal of the Royal Society Interface*, vol. 10, pp. 330–336, 2013.

[4] M. S. Lehnert, C. E. Beard, P. D. Gerard, K. G. Kornev, and P. H. Adler, "Structure of the lepidopteran proboscis in relation to feeding guild," *Journal of Morphology*, vol. 277, pp. 167–182, 2016.

[5] D. Monaenkova, M. S. Lehnert, T. A. Alexander, C. E. Beard, B. R. Alexander, W.-K. Lin, P. H. Adler, and K. G. Kornev, "Butterfly proboscis: combining a drinking straw with a nanosponge facilitated diversification of feeding habits," *Journal of the Royal Society Interface*, vol. 9, no. 69, pp. 720–726, 2012, doi: 10.1098/rsif.2011.0392.

[6] M. J. Norris, "The feeding-habits of the adult Lepidoptera Heteroneura," *Transactions of the Royal Entomological Society of London*, vol. 85, pp. 61–90, 1936.

[7] W. Büttiker, "Biological and morphological notes on the fruit-piercing and eye-frequenting moths," *Proceedings of the XI International Congress of Entomology*, Vienna, pp. 10–12, 1962.

[8] H. Bänziger and W. Büttiker, "Records of eye-frequenting Lepidoptera from man," *Journal of Medical Entomology*, vol. 6, no. 1, pp. 53–58, 1969, doi: 10.1093/jmedent/6.1.53.

[9] H. Bänziger, "The piercing mechanism of the fruit-piercing moth Calpe [Calyptra] thalictri Bkh. (Noctuidae) with reference to the skin-piercing blood-sucking moth C. eustrigata mps," *Acta Tropica*, vol. 27, pp. 53–88, 1970.

[10] H. Bänziger, "Biologie der lacriphagen Lepidopteren in Thailand und Malaya," *Revue Suisse de Zoologie*, vol. 79, pp. 1381–1469, 1973.

[11] H. Bänziger, "Fruit-piercing moths (Lep., Noctuidae) in Thailand: a general survey and some new perspectives," *Mitteilungen der Schweizerischen Entomologischen Gesellschaft*, vol. 55, pp. 213–240, 1982.

[12] J. A. Downes, "Lepidoptera feeding at puddle-margins, dung, and carrion," *Journal of the Lepidopterists' Society*, vol. 27, no. 2, pp. 89–99, 1973.

[13] P. H. Adler and D. L. Pearson, "Why do male butterflies visit mud puddles?," *Canadian Journal of Zoology*, vol. 60, pp. 322–325, 1982.

[14] L. E. S. Eastham and Y. E. E. Eassa, "The feeding mechanism of the butterfly Pieris brassicae L.," *Philosophical Transactions of the Royal Society B*, vol. 239, pp. 1–43, 1955.

[15] H. W. Krenn, "Functional morphology and movements of the proboscis of Lepidoptera (Insecta)," *Zoomorphology*, vol. 110, pp. 105–114, 1990, doi: 10.1007/BF01632816.

[16] H. F. Paulus and H. W. Krenn, "Vergleichende Morphologie des Schmetterlingsrüssels und seiner Sensillen-ein Beitrag zur phylogenetischen Systematik der Papilionoidea (Insecta, Lepidoptera)," *Journal of Zoological Systematics and Evolutionary Research*, vol. 34, no. 4, pp. 203–216, 1996, doi: 10.1111/j.1439-0469.1996.tb00826.x.

[17] H. W. Krenn, J. D. Plant, and N. U. Szucsich, "Mouthparts of flower-visiting insects," *Arthropod Structure and Development*, vol. 34, pp. 1–40, 2005, doi: 10.1016/j.asd.2004.10.002.

[18] J. M. Zaspel, S. Coy, K. Habanek, and S. J. Weller, "Presence and distribution of sensory structures on the mouthparts of self-medicating moths," *Zoologischer Anzeiger*, vol. 253, no. 1, pp. 6–10, 2013, doi: 10.1016/j.jcz.2013.06.001.

[19] L. E. Gilbert, "Pollen feeding and reproductive biology of Heliconius butterflies," *Proceedings of the National Academy of Sciences*, vol. 69, no. 6, pp. 1403–1407, 1972.

[20] J. K. Yuvaraj, M. N. Andersson, O. Anderbrant, and C. Löfstedt, "Diversity of olfactory structures: A comparative study of antennal sensilla in Trichoptera and Lepidoptera," *Micron*, vol. 111, pp. 9–18, 2018, doi: 10.1016/j.micron.2018.05.006.

[21] W. Büttiker, H. W. Krenn, and J. F. Putterill, "The proboscis of eye-frequenting and piercing Lepidoptera (Insecta)," *Zoomorphology*, vol. 116, no. 2, pp. 77–83, 1996, doi: 10.1007/BF02526872.

[22] H. W. Krenn, "Proboscis assembly in Lepidoptera—a once in a lifetime sequence of events," *European Journal of Entomology*, vol. 94, pp. 495–501, 1997.

[23] W. Speidel, H. Fanger, and C. M. Naumann, "The surface microstructure of the noctuid proboscis (Lepidoptera: Noctuidae)," *Zoologischer Anzeiger*, vol. 234, pp. 307–315, 1996.

[24] Y. Mengi and S. S. Çalışkan, "A morphological comparative study of proboscis sensilla between Pierinae and Coliadinae (Lepidoptera: Pieridae)," *SHILAP Revista de Lepidopterología*, vol. 52, no. 206, pp. 261–267, 2024, doi: 10.57065/shilap.

[25] M. J. Faucheux, "Sensillum types on the proboscis of the Lepidoptera: a review," *Annales de la Société Entomologique de France*, vol. 49, no. 1, pp. 73–90, 2013, doi: 10.1080/00379271.2013.767514.

[26] D. Petr and W. K. Stewart, "Comparative morphology of sensilla styloconica on the proboscis of North American Nymphalidae and other selected taxa (Lepidoptera): Systematic and ecological considerations," *Transactions of the American Entomological Society*, vol. 130, no. 4, pp. 293–409, 2004.

[27] H. Altner and I. Altner, "Sensilla with both, terminal pore and wall pores on the proboscis of the moth, Rhodogastria bubo Walker (Lepidoptera: Arctiidae)," *Zoologischer Anzeiger*, vol. 216, nos. 3–4, pp. 129–150, 1986.

[28] B. D. Walters, P. J. Albert, and R. Y. Zacharuk, "Morphology and ultrastructure of sensilla on the proboscis of the adult spruce budworm, Choristoneura fumiferana (Clem.) (Lepidoptera: Tortricidae)," *Canadian Journal of Zoology*, vol. 76, pp. 466–479, 1998, doi: 10.1139/z97-218.

[29] E. Städler, M. Städler-Steinbruchel, and W. D. Seabrook, "Chemoreceptors on the proboscis of the female eastern spruce budworm: morphological and histological study," *Mitteilungen der Schweizerischen Entomologischen Gesellschaft*, vol. 47, pp. 63–68, 1974.

[30] C. Borner, "22. Ordn. Lepidoptera, Schmetterlinge. In Brohmer, P. (Ed.), Fauna von Deutschland," 4th ed., Leipzig, Germany: X + 561 pp., 1932.

[31] S. Candan and N. K. Özyurt, "The morphology of the sensilla on the proboscis of Aporia crataegi (Linnaeus, 1758) (Lepidoptera: Pieridae)," *Comm. Journal of Biology*, vol. 6, no. 2, pp. 184–189, 2022, doi: 10.31594/commagene.1187494.

[32] S. S. Çalışkan, S. Candan, and N. K. Özyurt, "Ultrastructure of the sensilla on the proboscis of adult Tretopteryx pertusalis (Lepidoptera: Pyralidae)," *Transactions of the American Entomological Society*, vol. 149, no. 2, pp. 179–184, 2023.

[33] P. J. DeVries, "Patterns of butterfly diversity and promising topics in natural history and ecology," in *La Selva: Ecology and Natural History of a Neotropical Rainforest*, L. A. MacDade et al., Eds., Chicago, IL: University of Chicago Press, pp. 187–194, 1994.

[34] P. J. DeVries, D. Murray, and R. Lande, "Species diversity in vertical, horizontal, and temporal dimensions of a fruit-feeding butterfly community in an Ecuadorian rainforest," *Biological Journal of the Linnean Society*, vol. 62, no. 3, pp. 343–364, 1997.

[35] K. S. Brown Jr. and A. V. L. Freitas, "Atlantic Forest butterflies: indicators for landscape conservation," *Biotropica*, vol. 32, pp. 934–956, 2000, doi: 10.1111/j.1744-7429.2000.tb00631.x.

[36] K. S. Brown Jr. and A. V. L. Freitas, "Butterflies communities of urban forests in Campinas, São Paulo, Brazil: Structure, instability, environmental correlates, and conservation," *Journal of Insect Conservation*, vol. 6, pp. 217–231, 2002, doi: 10.1023/A:1024462523826.

[37] T. W. Pyrcz and J. Wojtusiak, "The vertical distribution of Pronophiline butterflies (Nymphalidae, Satyrinae) along an elevational transect in Monte Zerpa (Cordillera de Mérida, Venezuela) with remarks on their diversity and parapatric distribution," *Global Ecology and Biogeography*, vol. 11, no. 3, pp. 211–221, 2002, doi: 10.1046/j.1466-822x.2002.00285.x.

[38] D. Tobar, J. O. Rangel, and M. G. Andrade, "Diversidad de mariposas (Lepidoptera: Rhopalocera) en la parte alta de la cuenca del río el Roble (Quindío-Colombia)," *Caldasia*, vol. 24, pp. 393–409, 2002.

[39] D. B. Ribeiro, P. I. Prado, K. S. Brown Jr., and A. V. L. Freitas, "Additive partitioning of butterfly diversity in a fragmented landscape: Importance of scale and implications for conservation," *Diversity and Distributions*, vol. 14, pp. 961–968, 2008, doi: 10.1111/j.1472-4642.2008.00505.x.

[40] L. V. Vu, "Diversity and similarity of butterfly communities in five different habitat types at Tam Dao National Park, Vietnam," *Journal of Zoology*, vol. 277, pp. 15–22, 2009, doi: 10.1111/j.1469-7998.2008.00498.x.

[41] H. Bänziger, "Skin-piercing blood-sucking moths III: Feeding act and piercing mechanism of *Calyptra eustrigata* (Hmps) (Lep., Noctuidae)," *Mitteilungen der Schweizerischen Entomologischen Gesellschaft*, vol. 53, pp. 127–142, 1980.

[42] M. J. Faucheux, "Morphology and distribution of sensilla on the cephalic appendages, tarsi and ovipositor of the European sunflower moth, *Homoeosoma nebulella* Den. & Schiff. (Lepidoptera: Pyralidae)," *International Journal of Insect Morphology and Embryology*, vol. 20, pp. 291–307, 1991, doi: 10.1016/0020-7322(91)90018-5.

[43] M. J. Faucheux, "Sensilla on the antennae, mouthparts, tarsi and ovipositor of the sunflower moth, *Homoeosoma electellum* (Hulster) (Lepidoptera: Pyralidae): A scanning electron microscopic study," *Annales des Sciences Naturelles, Zoologie et Biologie Animale*, vol. 16, pp. 121–136, 1995.

[44] E. Hallberg, "Fine-structural characteristics of the antennal sensilla of *Agrotis segetum* (Insecta: Lepidoptera)," *Cell and Tissue Research*, vol. 218, pp. 209–218, 1981, doi: 10.1007/BF00210105.

[45] P. J. DeVries, *The Butterflies of Costa Rica and Their Natural History: Papilionidae, Pieridae, and Nymphalidae*, Princeton, NJ: Princeton University Press, 327 pp., 1987.

[46] C. Kremen, "Biological inventory using target taxa: A case study of the butterflies of Madagascar," *Ecological Applications*, vol. 4, pp. 407–422, 1994.

[47] A. L. Viloria, *Studies on the Systematics and Biogeography of Some Montane Satyrid Butterflies (Lepidoptera)*, Ph.D. dissertation, University of London, London, UK, 493 p., 1998.

[48] D. L. Murray, *Systematics of Neotropical Satyrinae Butterflies (Nymphalidae: Satyrinae: Euptychiina) Based on Larval Morphology and DNA Sequence Data and the Evolution of Life History Traits*, Ph.D. dissertation, Louisiana State University, New Orleans, LA, 380 p., 2001.

[49] M. C. Singer, P. R. Ehrlich, and L. E. Gilbert, "Butterfly feeding on lycopsid," *Science*, vol. 172, pp. 1341–1342, 1971.

[50] M. C. Singer and J. Mallet, "Moss feeding by satyrine butterfly," *Journal of Research on the Lepidoptera*, vol. 24, p. 392, 1986.

[51] P. R. Ackery, "Hostplants and classification: A review of nymphalid butterflies," *Biological Journal of the Linnean Society*, vol. 33, pp. 95–203, 1988.

[52] G. W. Beccaloni, Á. L. Viloria, S. K. Hall, and G. S. Robinson, *Catalogue of the Hostplants of the Neotropical Butterflies*, Zaragoza, Spain: Sociedad Entomológica Aragonesa, 536 pp., 2008.

[53] H. W. Krenn, K. P. Zulka, and T. Gatschnegg, "Proboscis morphology and food preferences in Nymphalidae (Lepidoptera, Papilionoidea)," *Journal of Zoology (London)*, vol. 253, pp. 17–26, 1998.

[54] H. W. Krenn and C. M. Penz, "Mouthparts of *Heliconius* butterflies (Lepidoptera: Nymphalidae): A search for anatomical adaptations to pollen-feeding behavior," *International Journal of Insect Morphology and Embryology*, vol. 27, no. 4, pp. 301–309, 1998, doi: 10.1016/S0020-7322(98)00022-1.

[55] L. Ma, H. K., P. Li, J. Liu, and X. Yuan, "Ultrastructure of the proboscis sensilla of ten species of butterflies (Insecta: Lepidoptera)," *PLOS ONE*, vol. 14, no. 3, p. e0214658, 2019, doi: 10.1371/journal.pone.0214658.

[56] H. W. Krenn, "Proboscis sensilla in *Vanessa cardui* (Nymphalidae, Lepidoptera): Functional morphology and significance in flower-probing," *Zoomorphology*, vol. 118, pp. 23–30, 1998, doi: 10.1007/s004350050053.

[57] F. Molleman, H. W. Krenn, M. E. Van Alphen, P. M. Brakefield, P. J. DeVries, and B. J. Zwaan, "Food intake of fruit-feeding butterflies: Evidence for adaptive variation in proboscis morphology," *Biological Journal of the Linnean Society (London)*, vol. 86, pp. 333–343, 2005, doi: 10.1111/j.1095-8312.2005.00533.x.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1147-1157 YEAR: 2024
DOI: 10.17798/bitlisfen.1521704

Performance Evaluation of the Time-Frequency Transformation Methods on Electrical Machinery Fault Detection

Mehmet Emin KILIC[1], Yunus Emre ACAR[2*]

[1]*Hey Teknoloji, Kahramanmaras, Türkiye*
[2]*Selcuk University, Faculty of Technology, Department of Electrical and Electronics Engineering, Konya, Türkiye*
(ORCID: 0009-0003-2381-7873) (ORCID: 0000-0002-6809-9006)

Keywords: Fault Detection, Machine Learning, Predictive Maintenance, Time-Frequency Transform, Transfer Learning, Vibration.

Abstract

Detecting faults in electrical machine systems is crucial for developing maintenance strategies. Modern technology enables personalized maintenance planning for system components by continuously or periodically monitoring systems with sensors. The first step in condition-based maintenance planning is predicting faults from sensor data. Monitoring vibration signals is one of the most preferred approaches for fault diagnosis in electrical machine systems. We have used a dataset containing vibration data recorded to detect intentionally created faults in an electrical machine system. The paper spots three popular methods to convert the time domain data into the frequency domain: power spectral density signal, spectrogram images, and scalogram images. Furthermore, we have analyzed the performance of the popular machine learning and deep learning methods with frequency-domain inputs. We have reported the results with accepted performance metrics such as accuracy, precision, recall, and F1 score. Our findings indicate that spectrogram images with the InceptionV3 model achieve maximum accuracy of over 98% accuracy among. The findings also highlight the necessity of carefully selecting model parameters based on the data type.

1. Introduction

Electric machine systems are among the most crucial components of the industry and hold a pivotal position in manufacturing facilities [1]. Faults or unplanned downtime in electric machine systems can lead to extensive financial consequences and damage to the reputation of facilities [2]. Consequently, developing maintenance strategies for electric machine systems becomes beneficial and imperative. While academic studies categorize maintenance strategies in various ways, many researchers unanimously agree on dividing maintenance activities into three fundamental groups: corrective, preventive, and predictive maintenance [3-5].

Corrective maintenance, which emphasizes post-failure recovery, and preventive maintenance, which aims to prevent potential failures before they occur, represent traditional approaches. Since corrective maintenance requires intervention after a failure, it is preferred in areas where unplanned downtime is not critical and is tolerable. On the other hand, preventive maintenance, in facilities where continuity is crucial, deems postponing the failure to a more suitable time by taking overly precautious measures sufficient. This approach involves creating a periodic maintenance and repair plan considering the average lifespan of electric machine system components. However, inherent unpredictability remains a challenge for electric machine systems. Determining a precise timeframe for potential failure is difficult due to the numerous factors influencing the expected lifespan of system components [6]. This uncertainty often leads facilities to adopt cautious measures and rely on maintenance programs based on generalized service life. Selçuk [4] has elucidated that

*Corresponding author: yacar@selcuk.edu.tr
Received: 23.07.2024, Accepted: 04.10.2024

such overly precautionary measures, especially concerning personnel and spare parts inventories, often lead to economic inefficiencies.

In recent years, supported by technological advancements and the proliferation of sensor technology, predictive maintenance offers personalized maintenance programs for electric machine systems [7]. In this approach, sensors continuously monitor the health of the motor and promptly flag any abnormalities. Beyond mere detection, the aim is to evaluate the progression and criticality of abnormalities and then make informed decisions regarding appropriate maintenance or repair interventions.

Two primary techniques dominate predictive maintenance in electric machine systems: current-based and vibration-based methods. The former approach, as detailed by Niu et al. [8], relies on the analysis of the motor's current signal. On the other hand, the vibration-based approach utilizes strategically positioned sensors to identify types of faults that create a vibration signature [9]. While time-domain analyses have occasionally been employed in both techniques, many modern studies [10] support frequency spectrum analysis. Tools such as Fast Fourier Transform (FFT), Short-Time Fourier Transform (STFT), Wavelet Transform (WT), and Power Spectral Density (PSD) commonly feature prominently in these analyses.

Features derived from these analytical tools are then interpreted. Historically, expert judgments formed the basis of such interpretations. However, the emergence of complex computational techniques like fuzzy logic, artificial neural networks, machine learning, and deep learning has ushered in a new era of analysis [11]. Garcia et al. [12], outlining the main stages of the standard predictive maintenance process, have enumerated various stages ranging from continuous monitoring of motor health to the strategic preparation of maintenance and repair plans based on multiple determinants. The first step in predictive maintenance is the disaggregation of faults using current or vibration signatures. Predicting how each disaggregated fault will deteriorate over time constitutes the next step. Proposing specific maintenance and repair plans for the predicted types of faults represents the final step of the approach. This study sheds light on the classification stage of predictive maintenance, a critical intersection in a comprehensive process. Inspired by recent advancements, the applicability of machine learning and deep learning methodologies to detect mechanical abnormalities via vibration signatures of electric machines is discussed. To provide a holistic perspective, results from machine learning and deep learning models are presented using 1-D power spectral density and 2-D spectrogram and scalogram images.

2. Material and Method

The most encountered electrical faults are winding and rotor bar faults, while bearing faults, misalignment, and imbalance are the most prevalent mechanical issues. This article focuses on classifying the mechanical faults in electric machine systems using vibration signals.

2.1. Dataset

The Machine Fault Dataset, MaFaulDA [13], was utilized in the article. The experimental setup comprises a circular metal disk combined with a DC motor. By placing various weights onto the disc, the imbalance failure is simulated. The overall system contains two bearings: one placed between the motor and the disk and the other positioned at the end of the system after the disk. Throughout the remainder of the article, the locations of the bearings will be referred to as Location 1 and Location 2, respectively. The dataset includes radial, tangential, and axial vibration signals recorded from these locations. After analyzing the data, we decided to use only the vibration signals on the radial axis. The data for each experiment has a length of 250,000, with an observation period of 5 seconds and a sampling frequency of 50 kHz. The dataset contains six distinct classes of data, consisting of five fault conditions and one healthy condition. Descriptions of the dataset are tabulated in Table 1.

2.2. PSD

PSD is a powerful tool for analyzing the frequency spectrum of signals. It serves as a mathematical measure of the power distribution across frequencies and finds widespread use in various applications such as communication [14, 15], power systems [16], optics [17], and more. Since anomalies in electric motors manifest differently in the frequency spectrum, PSD is also beneficial for fault classification studies.

There are various PSD estimation methods, broadly classified as parametric and non-parametric. Parametric estimation techniques require modeling the system with a limited number of parameters using autoregressive, moving average, and autoregressive moving average models. The accuracy of estimation is directly proportional to the accuracy of the model created.

Table 1. The data amount for each class and dataset explanation.

Class	Number of Data	Explanation
Normal	49	The data set has been created by 49 distinct experiments with different rotational speeds varying between 737 rotations per minute (rpm) and 3686 rpm.
Imbalance	333	The imbalance cases are created by placing various weights from 6g to 35g on the disc. For 44 different rpm values, 333 experiments have been conducted.
Horizontal misalignment	197	Misalignments are created by horizontally shifting the motor shaft from 0.5mm to 2 mm over 197 experiments.
Vertical misalignment	301	Misalignments are created by vertically shifting the motor shaft 0.51 mm to 1.9 mm over 301 experiments.
Overhang	513	Deliberately damaged bearings (outer and inner track faults, rolling element faults) have been placed at Location 1 in the experiments.
Underhang	558	Deliberately damaged bearings (outer and inner track faults, rolling element faults) have been placed at Location 2 in the experiments.
Total	1951	

Moreover, as the data size increases, the complexity of the method also increases. On the other hand, non-parametric approaches like the periodogram are more flexible and do not require a model or prior information about the system. Hence, they are suitable for fault detection applications in electric machine systems with nonlinear characteristics and difficult-to-model behavior.

This study focuses on the non-parametric PSD estimation technique known as the Welch approach. The estimated PSD of the recorded vibration signals is utilized as input for deep learning methods.

Within the Welch approach, the signal is divided into K segments, each of length M, with the inclusion of a designated overlap ratio. Following this division, these segments undergo multiplication by a window function represented as $w(n)$. Utilizing a window jump size of R, the calculated weighted segments, designated as $x_m(n)$, are determined as outlined in (1).

$$x_m(n) \cong w(n)x(n+mR)$$
$$n = 0,1,2,\dots,M-1, m = 0,1,2,\dots,K-1 \quad (1)$$

Subsequent to this, the periodogram is derived through the computation of the square of the absolute value of the N-length Fourier transform applied to the divided segments. The periodograms for each segment are calculated as in (2).

$$P_{x_m,M(w_k)} \cong \frac{1}{M}\left|\sum_{n=0}^{N-1} x_m(n)e^{\frac{-j2\pi nk}{N}}\right|^2 \quad (2)$$

Finally, the periodograms of all subsegments are averaged to obtain the overall PSD to decrease the variance. The overall PSD, $\hat{S}_x^W(w_k)$, is evaluated as given in (3).

$$\hat{S}_x^W(w_k) \cong P_{x_m,M(w_k)} = \sum_{m=0}^{K-1} P_{x_m,M(w_k)} \quad (3)$$

Figure 1(a) and Figure 1 (b) depict sample PSD estimates obtained from vibration signals taken respectively from Location 1 and Location 2 for each class provided in Table 1. Figure 1(a) illustrates that different peak points occur at distinct frequencies for each class, reinforcing the idea that the PSD of vibration signals obtained from this point is beneficial to distinguish between different types of faults. However, as shown in Figure 1(b), the sensor at Location 2 exhibits a level of DC noise that complicates the direct use of the data.

2.3. Spectrogram

A spectrogram is a powerful tool that visually represents the dynamic frequency changes over time, commonly used in analyzing sound and vibration signals. Vibration is a crucial precursor to faults and failures in electrical machines, with frequency spectra in vibration signals providing vital information about the nature and severity of faults. Hence, representing both time and frequency makes spectrograms a logical choice in fault diagnosis for electrical machine systems.

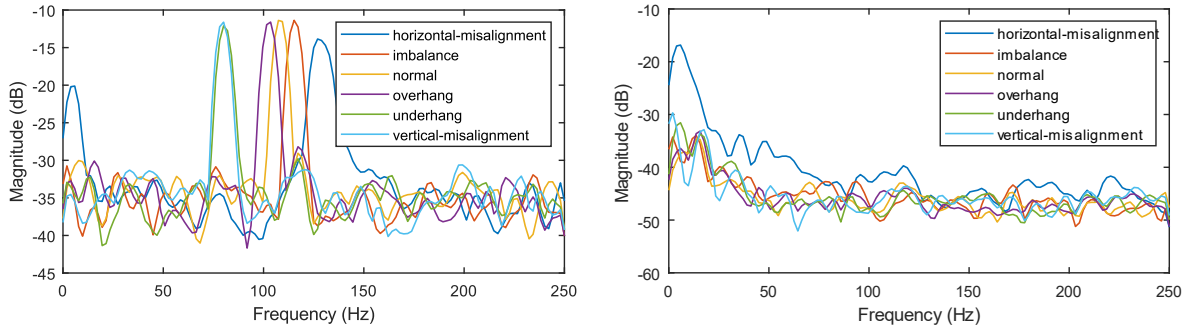


Figure 1. (a) The PSD estimation of the vibration signal on Location 1 and (b) Location 2 for each class

The Fourier Transform is a fundamental approach to discerning a signal's frequency content, deriving the frequency spectrum of a continuous signal using (4).

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \qquad (4)$$

where $x(t)$ is the time-dependent continuous signal, and $X(\omega)$ is the counterpart of $x(t)$ in the frequency domain. While Fourier Transform provides insights into the signal's frequencies, it lacks details about the occurrence time of each frequency component. When the frequency content by time is required to be analyzed, the STFT is one of the most preferred methods. Segmenting the signal into smaller time windows and applying the Fourier Transform after multiplying the signal by a window function, a frequency spectrum emerges. The signal then undergoes multiplication by the time-shifted version of the window function, yielding another frequency spectrum. This process spans the defined time duration of the signal. When the resulting frequency spectra, all of equal length, are arrayed as rows in a matrix, a comprehensive depiction of time-dependent frequency changes materializes. The STFT is as expressed in (5)

$$STFT(\omega, \tau) = \int_{-\infty}^{\infty} x(t) W(t-\tau) e^{-j\omega t} dt \qquad (5)$$

where τ is the time delay to slide the window function $W(t)$. For the discrete input signals, discrete-time FT is replaced by the FT and the rest of the procedures remain the same. Figure 2 visualizes the STFT processes [18].

The processes visualized in Figure 2 can be summarized as follows.

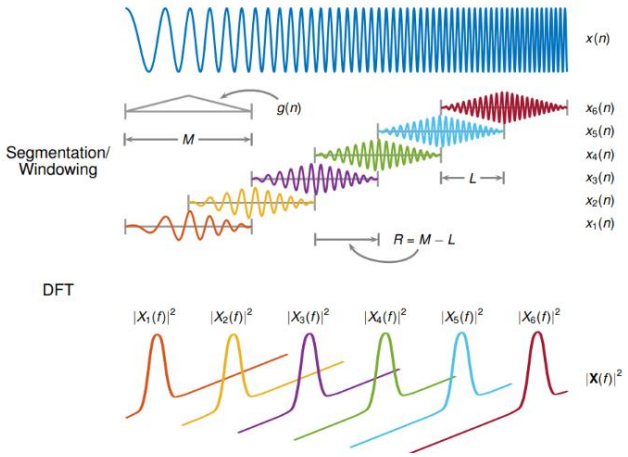


Figure 2. Visualization of STFT: segmentation, windowing, and DFT [18]

- The input signal is segmented, and each segment is multiplied by a proper window function of length.
- The window function slides over the input signal by L with $R = M - L$ samples overlap of consecutive segments.
- The frequency content of each segment is extracted by DFT and placed in the rows of a matrix. Thus, a 2D representation of the frequency by time change is obtained.

2.4. Scalogram

Much like a spectrogram, a scalogram visually illustrates how a signal's frequency changes over time. Unlike spectrograms, which rely on the Fourier Transform, scalograms based on the WT introduce a visually engaging exploration of time and frequency relationships.

In contrast to spectrograms, which offer a fixed time resolution through STFT, scalograms present a flexible narrative by balancing time and frequency resolutions. Unlike Fourier's sinusoidal signals, WT employs a mother function with distinct features. The signal to be analyzed is expressed

through this mother function, known for its irregularity, zero average, limited duration, and general asymmetry. Scalogram is obtained as a 2D representation of the input signal with the shifted and scaled variations of the mother function. While one of the dimensions is time, scale is the other one. Scale represents the mother function's different compressions or dilations, inversely related to frequency. While frequency increases, scale decreases, bringing about sharper time resolution. Conversely, a drop in frequency results in an increase in scale, widening the time window.

This tradeoff provides a zoom for detailed observation, enabling precise time resolution during sudden frequency changes and a high-frequency resolution within the domain of low-frequency elements. For a time-dependent continuous signal $x(t)$, Continuous Wavelet Transform (CWT) is expressed as in (6)

$$CWT(\tau,s) = \frac{1}{\sqrt{|s|}}\int_t x(t)\psi^*\left(\frac{t-\tau}{s}\right)dt \qquad (6)$$

where s is the scale referring to the compressions or dilations, τ is the time shift, $\psi(t)$ is a properly selected mother function. A mother function is required to be continuous, has null moments, and quickly decreases to zero with time (or band-limited). The admissibility condition to check any square-integrable candidate function is given in (7)

$$\int \frac{|\Psi(\omega)|^2}{|\omega|}d\omega < \infty \qquad (7)$$

where $\Psi(\omega)$ is the FT of the $\psi(t)$. When using CWT to create a scalogram, the procedures are similar to the ones to create spectrograms. The input signal is multiplied with $\psi^*(\tau,s)$ for different values of τ and s. While the time axis of the 2D scalogram is created by varying τ values, the scale axis corresponds to the different values of s. As previously mentioned, increasing the scale increases the time resolution and decreases the frequency resolution. The difference between CWT and STFT in terms of time and frequency resolution is visualized in Figure 3.

2.5. Classifiers

This study focuses on popular machine learning and deep learning methods as classifiers. While Decision Tree (DT), k-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Linear Discriminant (LD) are preferred among the machine methods, Xception, InceptionV3, and MobileNet deep learning models have been meticulously chosen among Keras models, considering both dimension and performance metrics. Various fine-tuning strategies have been experimented upon the selected models to achieve the highest success.

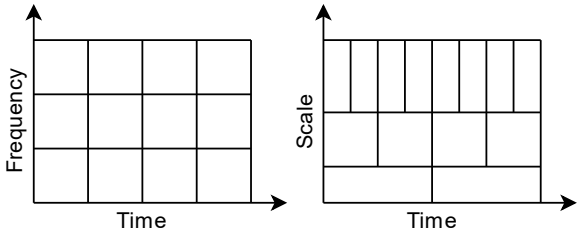


Figure 3 Decomposition of the time-frequency plane for (a) STFT and (b) CWT

3. Results and Discussion

The conducted study leverages vibration signals from an electrical machine system to classify mechanical faults within the system. 1-D power spectral density signals and 2-D scalogram and spectrogram images are created from the vibration signals. Machine learning techniques accept the power spectral density signals, while deep learning models use the scalogram and spectrogram images as inputs. Each machine learning algorithm was meticulously selected based on its proven effectiveness in handling similar datasets, ensuring a comprehensive evaluation of their performance in classifying mechanical faults. The deep learning models were chosen for their advanced capabilities in processing image data, leveraging transfer learning to enhance model accuracy and efficiency. The performances of the classifiers are evaluated with the well-accepted performance metrics explained in Table 2. The results reported in this study are obtained from the testing phase, utilizing a 5-fold cross-validation approach. This method involves dividing the dataset into five subsets, with each subset serving as a test set in turn, while the others are used for training. The performance metrics presented are the averages of the five folds.

In Table 2, True Positive (TP) refers to the number of examples correctly predicted as positive by the model, while True Negative (TN) is the number of examples correctly predicted as negative. False Positive (FP) represents the number of examples incorrectly predicted as positive but are negative, and False Negative (FN) is the number of examples incorrectly predicted as negative but are positive.

Table 2. The performance metrics

Metric	Description	Mathematical Expression
Accuracy (%)	The ratio of correctly classified instances to the total number of instances.	$\frac{TP}{TP+TN+FP+FN}$
Precision (%)	The ratio of true positive predictions to the total predicted positives.	$\frac{TP}{TP+FP}$
Recall (%)	The ratio of true positive predictions to the actual positives.	$\frac{TP}{TP+FN}$
F1 Score (%)	The harmonic mean of precision and recall.	$2 \times \frac{Precision \times Recall}{Precision + Recall}$

The performance of machine learning approaches is comparatively presented in Table 3. Table 3 provides a comparative overview of the performance of various classification methods applied to the dataset using accepted performance metrics. Moreover, the parameters yielding the best results are also provided. The learning parameters for the machine learning models were optimized using a random search strategy. For the Decision Tree classifier, the maximum number of splits was randomly selected from values between 4 and 100, with the optimal value chosen based on the highest accuracy. Various split rules, including the Gini's diversity index, twoing rule, and maximum deviance reduction, were evaluated. For the SVM classifier, linear, quadratic, and cubic kernels were tested, along with both one-vs-one and one-vs-all approaches for multiclass classification. Box constraint levels were randomly sampled from a range of 1 to 10. In the case of the KNN classifier, distance metrics such as Euclidean, cosine, Chebyshev, and Hamming were explored, and the number of neighbors was determined through random search from a range of 1 to 100. For Discriminant Analysis, both linear and quadratic kernels were considered, with covariance structures evaluated as full and diagonal. Among the examined methods, the KNN algorithm exhibited the highest performance with 96.56% accuracy, followed by SVM (95.85%) and DT (94.3%), while the LD method showed the lowest performance at 75.71%. These results highlight the non-linear nature of fault classes in the dataset and the effectiveness of algorithms like KNN in capturing complex patterns. The combination of PSD signal processing and KNN proves to be a robust method for fault detection in electrical machines. These findings underscore the importance of selecting an appropriate classification method tailored to the dataset for optimal performance. For a more detailed analysis of the classification performance of the methods, further information can be obtained through the analysis of confusion matrices provided in Figure 4 (a)-(d).

Table 3. Performance of Machine Learning approaches with PSD signals

Method	Parameters	Acc. (%)	Prec. (%)	Rec. (%)	F1 Score (%)
DT	max. # of splits: 100 split criterion: Gini diversity index	94.3	92.31	90.81	91.55
SVM	kernel function: linear box constraint level: 1 multiclass meth.: one vs one	95.85	94.18	95.98	95.08
KNN	# of neighbors: 1 distance metric: Euclidean distance weight: equal	96.56	96.75	96.72	96.74
LD	preset: Linear discriminant covariance structure: full	75.71	68.22	67.45	67.83

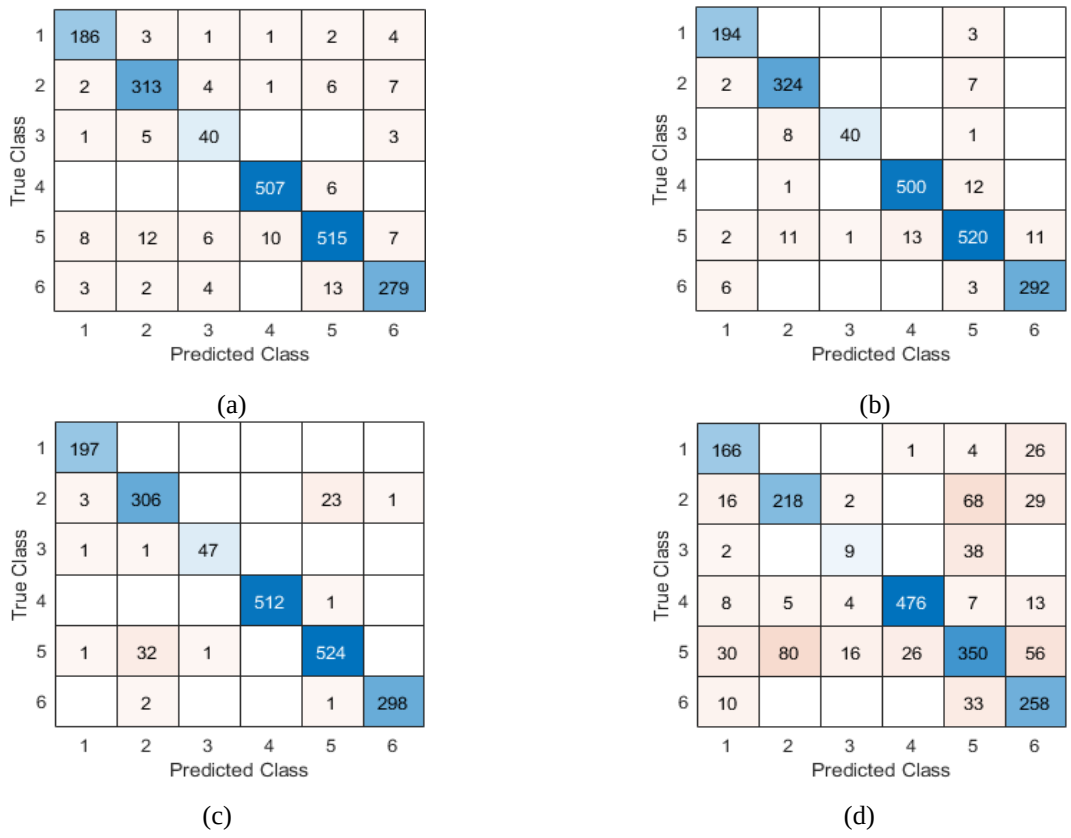


Figure 4. Confusion matrices for (a) DT, (b) SVM, (c) KNN, and (d) LD

Figure 4 shows that the KNN model demonstrates the most balanced and superior performance, particularly excelling in classifying 'normal', 'imbalance', and 'horizontal misalignment' categories, though it shows some confusion between 'overhang' and 'underhang' classes. SVM exhibits comparable performance to KNN but with slightly more errors in the 'vertical misalignment' class. While the DT model performs well overall, it struggles more with distinguishing between 'imbalance' and 'horizontal misalignment'. The LD model underperforms compared to the others, particularly in differentiating 'overhang' and 'underhang' classes, suggesting these fault types are not linearly separable. These results indicate that non-linear models, especially KNN and SVM, are more effective for fault detection.

As the second part of the study, deep learning models with the inputs of scalograms and spectrograms are analyzed using a transfer learning approach across three different models. The performances of Xception, InceptionV3, and MobileNet deep learning models are comparatively evaluated through heatmaps. The study presents how the performance of the methods changes with the fine-tuning and learning rate parameters. The learning rate has been optimized using a random search within the range of 5e-6 to 0.0002, and the fine-tuning steps have been selected from values between 10 and 200. The performance of each model for varying fine-tuning and learning rate parameters can be observed for the scalogram images in Figure 5 through the heatmaps provided.

Figure 5 illustrates the performance of three deep learning models (InceptionV3, Xception, and MobileNet) using scalogram inputs, represented as heatmaps across various fine-tuning steps and learning rates. The MobileNet architecture demonstrates superior performance, achieving a peak accuracy of 96.94% with 50 fine-tuning steps and a learning rate of 0.00002. InceptionV3 also exhibits robust performance, reaching 94.73% accuracy, while Xception shows comparatively lower performance.

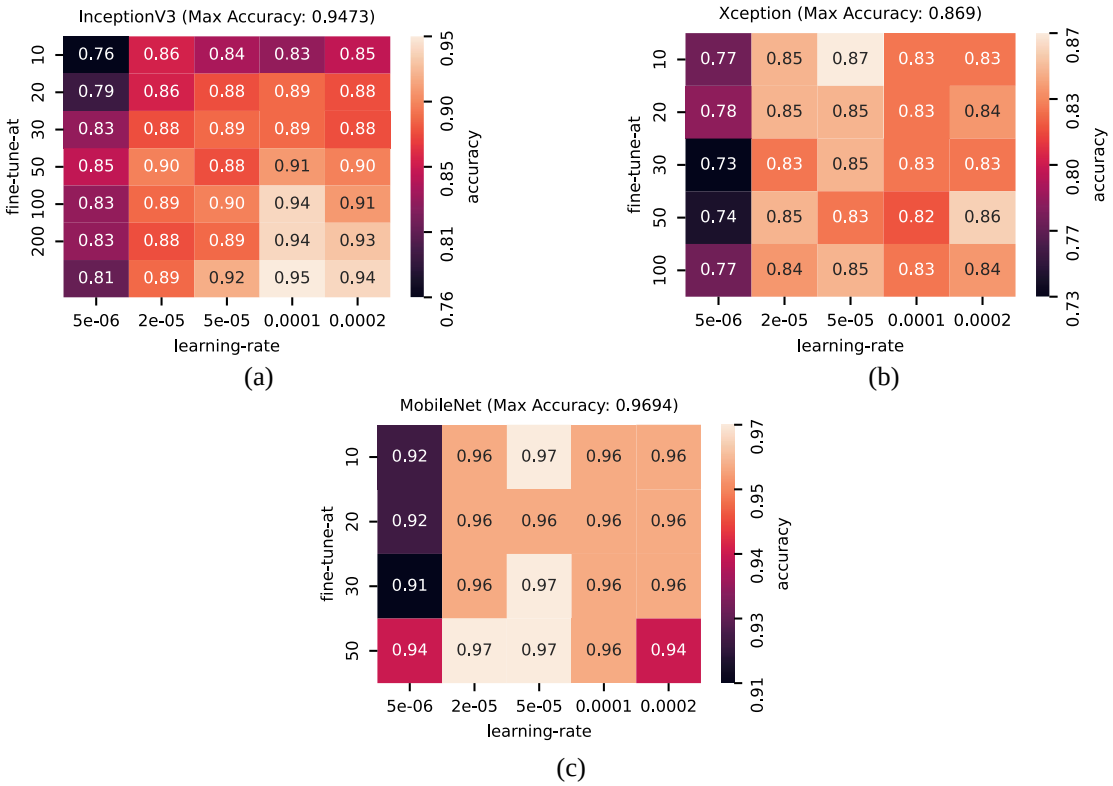


Figure 5. Heatmaps for the models (a) InceptionV3, (b) Xception, and (c) MobileNet with the scalogram images

These results suggest that the MobileNet architecture is particularly adept at extracting relevant features from scalogram representations of vibration signals for fault classification. Figure 6 displays the heat maps to indicate the effects of fine-tune-at and learning-rate parameters for the spectrogram images.

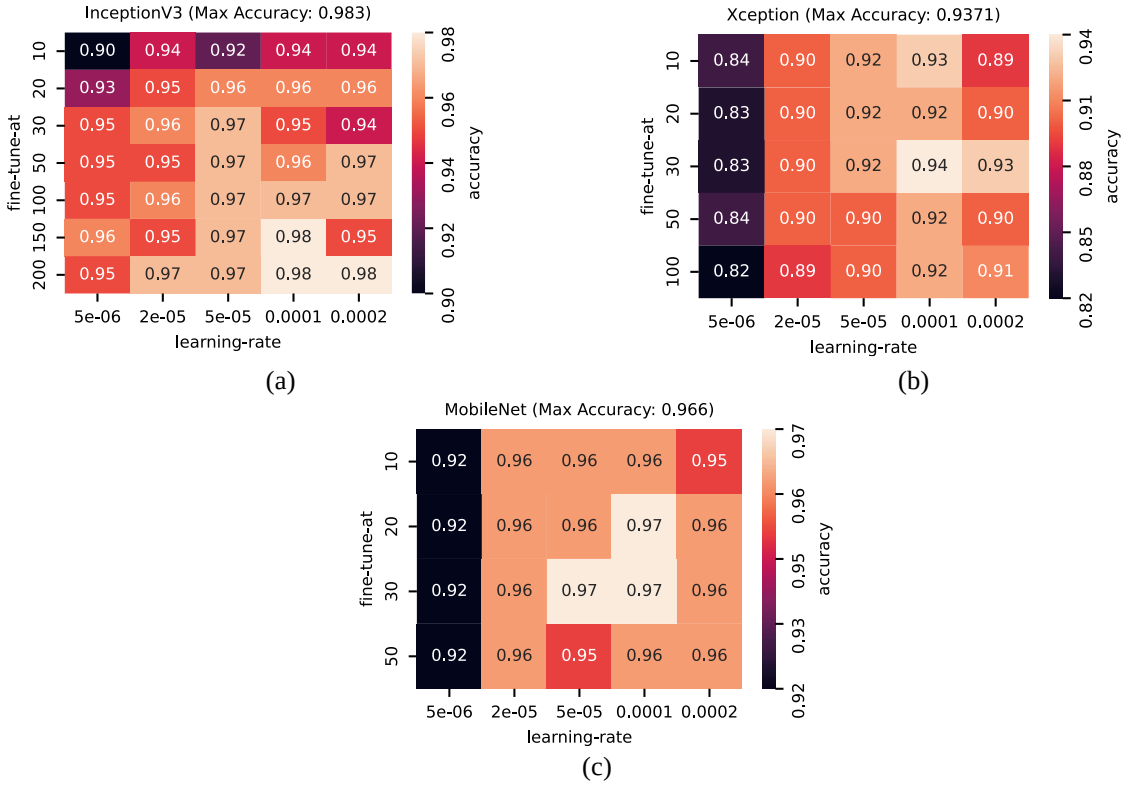


Figure 6. Heatmaps for the models (a) InceptionV3, (b) Xception, and (c) MobileNet with the spectrogram images

Figure 6 presents heatmaps depicting the performance of InceptionV3, Xception, and MobileNet models using spectrogram inputs across various fine-tuning steps and learning rates. Notably, the InceptionV3 model achieves the highest accuracy of 98.30% with 150 fine-tuning steps and a learning rate of 0.0001. Xception's performance significantly improves with spectrogram inputs compared to scalograms, while MobileNet maintains consistent performance. The superior results obtained with spectrogram inputs, particularly for InceptionV3, indicate that this time-frequency representation may capture fault-related features more effectively for the given dataset. The distinct optimal hyperparameter combinations for each model emphasize the necessity of model-specific tuning. These findings highlight the potential of deep learning approaches, especially InceptionV3 with spectrogram inputs, for highly accurate fault detection in electrical machinery.

Furthermore, our study underscores the necessity of selecting the most suitable model for different data representations, such as spectrograms and scalograms. For example, the InceptionV3 model outperforms MobileNet when processing spectrogram data, emphasizing the importance of evaluating the data type and model architecture features together.

Moreover, it's evident that the choice of learning rate significantly impacts the effectiveness of fine-tuning strategies. Carefully selecting the optimal learning rate tailored to each model and data type can substantially enhance accuracy rates. Table 4 summarizes the fine-tune-at and learning rate values where deep learning models achieve maximum accuracy.

In conducting a general assessment of our findings, we have compared the performance of traditional machine learning methods using PSD inputs with deep learning methods using spectrogram and scalogram inputs. While machine learning approaches have demonstrated strong performance, with KNN achieving the highest accuracy of 96.56% among PSD-based methods, deep learning models, particularly when using spectrogram inputs, have shown even higher accuracy. The InceptionV3 model with spectrogram inputs outperformed all other methods, achieving an impressive 98.30% accuracy. This suggests that hierarchical feature learning in deep neural networks is particularly effective for capturing complex patterns in vibration data for fault detection.

However, it's worth noting that the performance gap between the best machine learning model (KNN with PSD) and the best deep learning model (InceptionV3 with spectrograms) is relatively small (1.74%). This indicates that well-tuned traditional machine learning methods can still be competitive for fault detection tasks, especially when interpretability or computational resources are a concern. The choice between these approaches may depend on factors such as dataset size, need for interpretability, and available computational resources. Table 5 summarizes the best-performing models from each approach, highlighting their respective strengths in fault detection for electrical machinery.

Table 4. Performance of Transfer Learning models for scalogram and spectrogram images

Model	Scalogram images			Spectrogram images		
	Acc. (%)	fine-tuning step	learning rate	Acc. (%)	fine-tuning step	learning rate
InceptionV3	94.73	200	0.0001	98.30	150	0.0001
Xception	86.90	10	0.00005	93.71	30	0.0001
MobileNet	96.94	50	0.00002	96.60	30	0.00005

Table 5. Comparison of best-performing models for each approach

Approach	Model	Input	Acc. (%)	Prec.(%)	Rec. (%)	F1 score (%)
ML	KNN	PSD	96.56	96.75	96.72	96.74
ML	SVM	PSD	95.85	94.18	95.98	95.08
DL	InceptionV3	Spectrogram	98.30	98.10	98.44	98.56
DL	MobileNet	Scalogram	96.94	97.08	97.32	97.52

4. Conclusion and Suggestions

This study explores the performance of machine learning and deep learning models to classify the vibration data collected from electrical machines for fault detection. We first examine traditional machine learning methods like DT, KNN, SVM, and LD alongside three different deep learning models using transfer learning. The findings shed light on the effects of various data types and model architectures on classification performance.

Among the machine learning methods, which are fed by the 1-D PSD signal, the KNN model achieved the highest accuracy rate of 96.56%. On the other hand, the InceptionV3 deep learning model surpassed the highest accuracy with spectrogram images. The results highlight the importance of fine-tuning strategies, particularly emphasizing the significance of learning rate and fine-tune-at parameters. The InceptionV3 model has surpassed the others with an accuracy of 98.30% by selecting the fine-tune-at and learning rate parameters as 150 and 0.0001. The outcomes of the study can guide data analysis and fault diagnosis processes in industrial systems.

Acknowledgment

This work was supported by TUBITAK 1507 under grant number 7220463.

Contributions of the authors

Yunus Emre Acar: Conceptualization, Methodology, Software, Investigation, Writing, Visualization, Supervision.

Mehmet Emin KILIC: Project administration, Funding acquisition, Methodology, Software, Investigation, Writing, Visualization.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] R. Hu, J. Wang, A. R. Mills, E. Chong, and Z. Sun, "Current-residual-based stator interturn fault detection in permanent magnet machines," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 1, pp. 59–69, 2020.

[2] P. Nunes, J. Santos, and E. Rocha, "Challenges in predictive maintenance–A review," *CIRP Journal of Manufacturing Science and Technology*, vol. 40, pp. 53–67, 2023.

[3] G. A. Susto, A. Schirru, S. Pampuri, S. McLoone, and A. Beghi, "Machine learning for predictive maintenance: A multiple classifier approach," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 3, pp. 812–820, 2014.

[4] S. Selcuk, "Predictive maintenance, its implementation and latest trends," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 231, no. 9, pp. 1670–1679, 2017.

[5] Y. Ran, X. Zhou, P. Lin, Y. Wen, and R. Deng, "A survey of predictive maintenance: Systems, purposes and approaches," *arXiv preprint arXiv:1912.07383*, 2019.

[6] T. D. Popescu, D. Aiordachioaie, and A. Culea-Florescu, "Basic tools for vibration analysis with applications to predictive maintenance of rotating machines: an overview," *The International Journal of Advanced Manufacturing Technology*, pp. 1–17, 2022.

[7] N. Vafaei, R. A. Ribeiro, and L. M. Camarinha-Matos, "Fuzzy early warning systems for condition based maintenance," *Computers & Industrial Engineering*, vol. 128, pp. 736–746, 2019.

[8] G. Niu, X. Dong, and Y. Chen, "Motor Fault Diagnostics Based on Current Signatures: A Review," *IEEE Transactions on Instrumentation and Measurement*, 2023.

[9] F. Al-Badour, M. Sunar, and L. Cheded, "Vibration analysis of rotating machinery using time–frequency analysis and wavelet techniques," *Mechanical Systems and Signal Processing*, vol. 25, no. 6, pp. 2083–2101, 2011.

[10] J. R. Rivera-Guillen, J. De Santiago-Perez, J. P. Amezquita-Sanchez, M. Valtierra-Rodriguez, and R. J. Romero-Troncoso, "Enhanced FFT-based method for incipient broken rotor bar detection in induction motors during the startup transient," *Measurement*, vol. 124, pp. 277–285, 2018.

[11] J. Dalzochio et al., "Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges," *Computers in Industry*, vol. 123, p. 103298, 2020.

[12] M. C. Garcia, M. A. Sanz-Bobi, and J. Del Pico, "SIMAP: Intelligent System for Predictive Maintenance: Application to the health condition monitoring of a wind turbine gearbox," *Computers in Industry*, vol. 57, no. 6, pp. 552–568, 2006.

[13] F. Ribeiro, "MaFaulDa-machinery fault database," *Signals, Multimedia, and Telecommunications Laboratory*, 2016.

[14] R. López-Valcarce, "General form of the power spectral density of multicarrier signals," *IEEE Communications Letters*, vol. 26, no. 8, pp. 1755–1759, 2022.

[15] L. Niu and F. Li, "Cooperative Spectrum Sensing for Internet of Things Using Modeling of Power-Spectral-Density Estimation Errors," *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7802–7814, 2021.

[16] Z. Zhao, K. Peng, R. Xian, and X. Zhang, "Localization of Oscillation Source in DC Distribution Network Based on Power Spectral Density," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 1, pp. 156–167, 2023.

[17] S. Benkner, A. Herzog, S. Klir, W. D. Van Driel, and T. Q. Khanh, "Advancements in Spectral Power Distribution Modeling of Light-Emitting Diodes," *IEEE Access*, vol. 10, pp. 83612–83619, 2022.

[18] "Signal Processing Toolbox," The MathWorks Inc., Available: https://www.mathworks.com/help/signal/ug/spectrogram-computation-with-signal-processing-toolbox.html (accessed February 1, 2024).

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1158-1168 YEAR: 2024
DOI: 10.17798/bitlisfen.1523524

Classification of ECG Signals Using GAN, SMOTE, and VAE Data Augmentation Methods: Synthetic vs. Real

Turgut ÖZSEVEN[1*]

[1]*Tokat Gaziosmanpaşa University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Tokat, Türkiye*
(ORCID: 0000-0002-6325-461X)

Keywords: Synthetic data, Data augmentation, GAN, SMOTE, VAE, Heartbeat classification, Deep learning.

Abstract
Classification is separating data into predefined categories by obtaining descriptive features. In the classification process, machine and deep learning algorithms assume that the class samples are evenly distributed. In particular, the dataset size used in deep learning is significant for classification success. However, obtaining balanced data distribution in real-life problems is very difficult. This negatively affects class-based accuracy. Various methods are used in the literature to overcome the unbalanced data problem. This study investigated the effects of GAN, SMOTE, and VAE methods on ECG data. For this purpose, the heartbeat signals in the MIT-BIH dataset were used. To test the performance of the methods, a performance comparison was made using real and synthetic data, and finally, the model trained with synthetic data was tested with real data. According to the results, 96.5% accuracy was obtained with the real data. The highest classification accuracy of 100.0% was obtained in VAE when using only synthetic data. In training with synthetic data and test results with real data, the highest classification success was 86.4% with SMOTE. When synthetic and real data sets are used together, the highest success rate is 98.6% with VAE. In addition, the accuracy of all classes is evenly distributed after data augmentation.

1. Introduction

An electrocardiogram (ECG) is the recording and graphing of electrical activities in the heart to study the heart's working. It is a physiological signal that contains important pathological information for detecting heart diseases. Therefore, the amount of ECG data is often very large. Manually reviewing this data by experts is both time-consuming and subjective. Computer-aided diagnosis methods are used to overcome this problem [1]. Traditional computer-aided diagnosis methods consist of two stages: feature extraction (discrete wavelet transform, time-domain features, frequency-domain features, etc.) and classification [2]. Recently, feature extraction has not been used as a separate process with deep learning models. Deep learning models perform the feature extraction process within the model. Convolutional neural network (CNN), one of the deep learning models, has shown high performance in many studies, such as biosignal classification, image recognition, and arrhythmia classification [2].

Traditional and deep learning-based computer-aided diagnosis systems assume a balanced distribution of class samples. In particular, the dataset size used in deep learning is extremely important for classification success. However, obtaining balanced data distribution in real-life problems is very difficult. This negatively affects class-based accuracy. Unbalanced data increases classification accuracy for classes with more samples [3]. Because some arrhythmias in ECG data are rare, unbalanced data is much more prominent [1].

Undersampling and oversampling methods are used to overcome the unbalanced data problem. In

*Corresponding author: turgut.ozseven@gop.edu.tr
Received: 28.07.2024, Accepted: 25.09.2024

the undersampling method, data is discarded from the class with more examples, and the number of examples is equalized to the class with a smaller number of examples. Oversampling, on the other hand, is to balance the number of samples for the class with a large number of samples by reusing the samples in the class with a low number of samples [4]. The most basic method used for oversampling in image-based studies is to change the dimensional and morphological structure of the image. However, both methods corrupt the original structure of the data. Undersampling leads to data waste. In oversampling, it causes an overfitting problem because the same data is reused. To overcome these problems, deep learning-based synthetic data production methods are mostly used. The most widely used method is the Generative Adversarial Network (GAN). Also, a few studies used the Synthetic Minority Oversampling Technique (SMOTE) for oversampling [5] - [9]. Another deep learning-based method, Variational Autoencoders (VAE), has been used in limited studies [10].

GAN contains two neural networks, a generator and a discriminator. The generator generates the candidate image, and the discriminator evaluates whether the image produced is real or fake. Both networks are competitively trained in parallel. The discriminator compares the real image with the image produced by the generator. The comparison result is sent to the generator, which will update itself according to the result. This cycle will continue until the generator produces images that are more like the image held by the discriminator. This process can be followed with several iterations, and the loss curve [3]. Studies on synthetic data generation and data balancing in recent years mostly use GAN [11] - [15]. Studies show that the data produced with GAN can be used for training purposes [11-13]. Tran et al. (2024) produced synthetic data with GAN by converting ECG signals into a two-dimensional format. High success was achieved with the proposed method, and a new benchmark was set for ECG synthesis [14]. In the study using Bi-LSTM and CBAM based GAN for data augmentation, 99.46% classification success was achieved on the MIT-BIH-AR dataset [15]. Hyland et al. (2017) used recurrent conditional GANs (RCGAN) for the synthetic generation of time series data and concluded that they could be used with a small drop in performance when tested with real data [16]. In the study to overcome the data imbalance problem in the MIT-BIH database, the GAN method increased the classification accuracy by about 9% [3]. Synthetic data generation is not only used with the physiological structure of the signal. Wulan et al. (2020) used GAN with signal structure, spectrogram, and wavelet transform [17]. In another study, a SynSigGAN model was proposed for producing ECG, EEG, EMG, and PPG synthetic data [18]. While the GAN can be designed separately for each class in the synthetic ECG data production process, there are also class-based models [19]. Zhou et al. (2021) proposed the ACE-GAN model, which provides class-based data augmentation, and achieved 99.0% success in the MIT-BIH dataset [2]. The GAN method was also used in hybrid form with other methods. Zhu et al. (2019) used BiLSTM for the generator and CNN for the discriminator in the GAN model [20]. Salazar et al. (2021) suggested that GAN and vector Markov Random Field based GANSO model be used in datasets with a very small number of samples [6]. The proposed model was compared with SMOTE, and higher accuracy was obtained in GANSO. In the study in which the data augmentation method called ProEGAN-MS was proposed, the results were compared with the GAN, and it was determined that the proposed model showed higher success [1]. In another study, VAE and GAN were used together, and 98.5% accuracy was obtained in the MIT-BIH dataset [10]. In the study where the transformer and GAN were used together, the TCGAN was proposed, and a success of 94.7% was achieved in the 1D-CNN and MIT-BIH datasets [21]. While creating GAN models, 1D-CNN, 2D-CNN, or BiLSTM models are used. In the study in which GAN-1D-CNN and GAN-BiLSTM models were used, data augmentation was performed with GAN and classification with 1D-CNN and BiLSTM [22]. According to the results of the study, GAN-BILSTM achieved higher success. GAN increased the models' classification success by approximately 11% [22]. In another study, the ResNet CNN model was used with BiLSTM, and the classification success was 99.4% in the MIT-BIH dataset [23]. GAN is used not only for data augmentation/production but also for noise reduction in real data. Global discriminator, local discriminator, and GAN were used in the study to generate noisy ECG signals due to scanning and digitizing ECG recordings [24].

SMOTE was first proposed in 2002, and the interpolation of the original samples with their neighbors was used to generate synthetic samples [25]. For the classification of ECG signals, features were obtained with wavelet packet decomposition and 1D-CNN, and selective ensemble learning framework and SMOTE were used. Classification success was 96.3% in the MIT-BIH dataset [9]. In another study using MIT-BIH and SMOTE, the classification success was 98.3% [8]. In another study called ArrhyNet, classification success was obtained at 92.7% in the MIT-BIH dataset using wavelet

transform, SMOTE, and CNN [7]. For ECG classification, the hybrid structure is used not only for data augmentation but also for classification. The results of the study, in which CNN and LSTM were used together for classification and SMOTE and Tomek were used for data augmentation, showed that data augmentation increased the classification success by 20% [5].

As explained above, GAN is the most common method for synthetic data generation in recent years. The SMOTE method is an older method than the GAN method. However, both SMOTE and VAE have been used in limited studies [5-9]. Some studies did not detail the results before data augmentation while giving the classification accuracy [1, 2, 6, 21-23]. This poses a problem when comparing the performance of synthetic data. In addition, data production methods are not often compared in many places. This poses a challenge for researchers to formulate a study plan. In this study, heartbeat data in the MIT-BIH database were obtained, and synthetic data were produced using these data with GAN, SMOTE, and VAE methods. Then, the performance of the real and synthetic data in ECG classification was compared with ten different analyses. Therefore, the contributions of this research article are as follows: 1) The efficiency of synthetic data in ECG classification was analyzed. 2) The classification success performances of three different data generation methods were compared. 3) The performance of the systems trained only with synthetic data in classifying real data was examined. 4) ECG classification success was evaluated when the hybrid data set was used by increasing the data. 5) A guide was created for researchers working in ECG classification on synthetic data generation.

The remainder of this article is organized as follows: Section II contains the dataset and methods used in the study. Section III contains the results of the analyses performed. Section IV contains the conclusion and discussion.

2. Material and Method

2.1. Data Descriptions

The MIT-BIH [26] database, created in cooperation with MIT and Beth Israel Hospital, was used. This database contains 48 ECG recordings of 47 participants, 30 minutes long. Annotation of the database was done by two or more cardiologists. Each recording contains two channels, MLII and V1, V2, V4, and V5, and we used the data of the MLII channel. Since 90% of the MIT-BIH database contains a normal heartbeat, it has a highly unbalanced structure [27]. Forty-two ECG recordings containing beats belonging to four classes (F: ventricular fusion beat, N: normal beat, S: Supraventricular premature beat, V: ventricular premature beat) were used in the analyses. The annotations in the MIT-BIH database and the heartbeat distribution used according to the AAMI standards are given in Table 1. Since heartbeat creates different graphs for the same diagnosis in different patients, the records used in the dataset were chosen as participant-dependent.

Table 1. Heartbeat distribution used

AAMI	MIT-BIH	Beat Count	Record Numbers
F	F	749	205, 208, 213, 219, 223
N	N, L, R, e, j	10440	100, 101, 103, 105, 106, 108, 109, 111, 112, 113, 115, 116, 117, 118, 119, 121, 122, 123, 124, 200, 201, 202, 205, 208, 209, 210, 212, 213, 215, 220, 221, 222, 223, 230, 234
S	A, a, J, S	2612	100, 101, 103, 108, 112, 113, 116, 117, 118, 121, 124, 200, 201, 202, 203, 205, 208, 209, 210, 213, 215, 220, 222, 223, 228, 232, 233, 234
V	V, E	4428	105, 108, 109, 111, 114, 116, 118, 119, 121, 123, 201, 202, 207, 208, 209, 213, 214, 215, 219, 221, 223, 228, 230, 233, 234

2.2 Convolutional Neural Network (CNN)

CNN is a multi-layered perceptron that carries out the feature extraction and classification process with the help of more than one layer. Therefore, the hardware and data it needs is more than that of traditional methods. CNN obtains features from the data with a certain number of filters and filter sizes in each convolutional layer, giving a linear output. The activation function is used after the convolutional layer to use the model on nonlinear problems. Another layer used in CNN architecture is the dropout layer. This layer prevents overfitting by removing some neurons from the model. The layer used for classification in CNN models is the dense layer [28]. Within the scope of the study, a simple CNN model

was created to perform the analysis. The developed model is given in Figure 1.

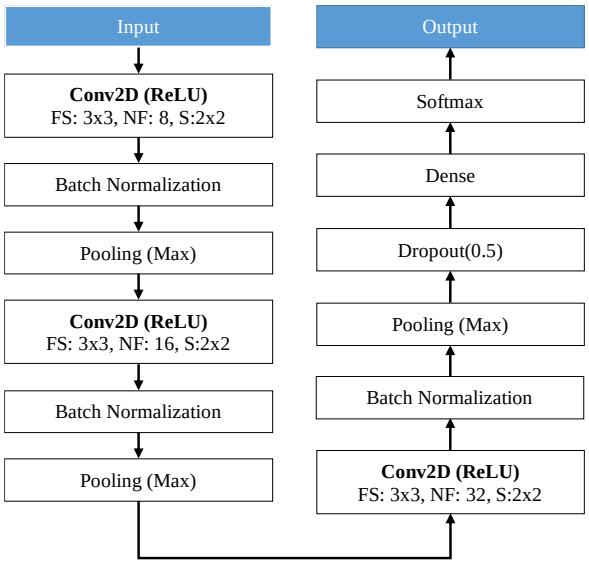


Figure 1. The structure of the 2D-CNN model used in the study (FS: FilterSize, NF: NumFilters, S:Stride)

The parameters used in the training phase of the created CNN model are given in Table 2.

Table 2. Parameters used for the CNN model

Parameters	Value
Data selection for Training, Validation, and Testing	Random permutation
The part of the data reserved for training, validation, and testing	70%, 15%, %15
Optimizer	Adam
Learning Rate	0.001
Epochs	20
Mini Batch Size	64

2.3. Generative Adversarial Network (GAN)

The GAN approach was first proposed by Goodfellow et al. in 2014 [29]. It is used for data generation in many fields, such as image, video, natural language processing, audio analysis, and time series. GAN includes the generator (G) and discriminator (D) model. The generator is used to generate data. It generates synthetic data *G(z)* using random noise *z*. On the other hand, the discriminator tries to determine whether the sample is synthetic by taking samples from real and synthetic data. The GAN training process is defined as a game between two competing networks. The discriminator learns to distinguish between real and fake samples. The Generator learns to deceive the discriminator. The input parameter is *x* representing the data, and the output *D(x)* is the probability that *x* is real data. The formula for GAN is given in Equation 1 [4]:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim P_{data}(x)}[logD(x)] + \mathbb{E}_{z\sim P_z(z)}[\log(1 - D(G(z)))] \quad (1)$$

where *D(G(z))* is the probability that network *D* decides that the data produced by *G* is real. The purpose of *G* is to maximize *D(G(z))*. The stronger *D*'s ability, the bigger *D(x)* should be and the smaller *D(G(x))* should be. Meanwhile, *V(D, G)* will get bigger. P_{data} represents the distribution of actual data. P_z represents the distribution of noise, usually a Gaussian distribution, from which we can form a synthetic image. $\mathbb{E}_x$ and $\mathbb{E}_z$ represent the expected log probability from different outputs of both real and generated images.

The model used for synthetic data generation with GAN is given in Figure 2.

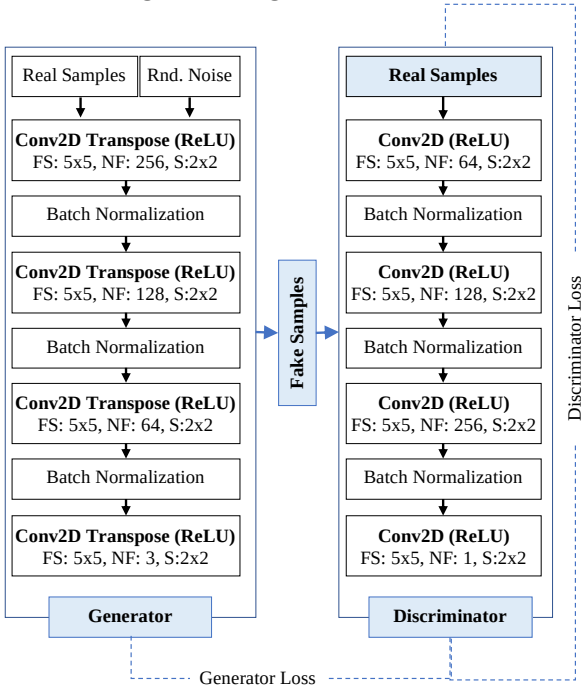


Figure 2. The structure of the GAN model used (epochs: 500, minibatchsize: 128, learningrate: 0.0002)

Samples of synthetic images obtained with the GAN model are given in Figure 3.

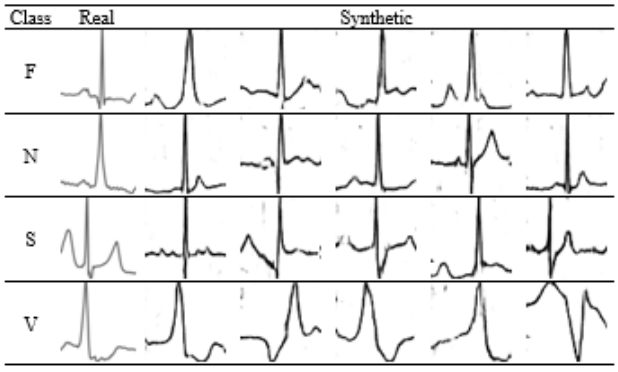


Figure 3. Synthetic images obtained with GAN.

2.4. Synthetic minority Over-sampling Technique (SMOTE)

The SMOTE method, first introduced in 2002, is used for oversampling for classes with a small number of samples and undersampling for classes with a large number of samples [25]. Unlike random sampling methods, it produces synthetic samples based on the *k* nearest neighbors of the samples examined. The details of the algorithm are given below [30]:

Suppose we have a dataset $D \in R^{nxm}$ with *m* number of features and *n* number of samples and *K* categories. Calculate the Euclidean distance between each sample in the minority class $T(T \in K)$ and other classes. A sampling rate is then adjusted based on the sample imbalance rate to determine the sample multiplier *N*. For each minority sample $X_i (i = 1, \dots, num(T))$, several samples are randomly selected among *k* neighbors $X_i^j (j = 1, \dots, k)$. For each randomly selected nearest neighbor X_i^j, the new sample is synthesized with sample X_i according to the formula:

$$X_{synthetic} = X_i + \left(X_i - X_i^j\right).rand() \quad (2)$$

where *rand()* is a uniformly distributed non-random number. Samples of synthetic images obtained with SMOTE are given in Figure 4.

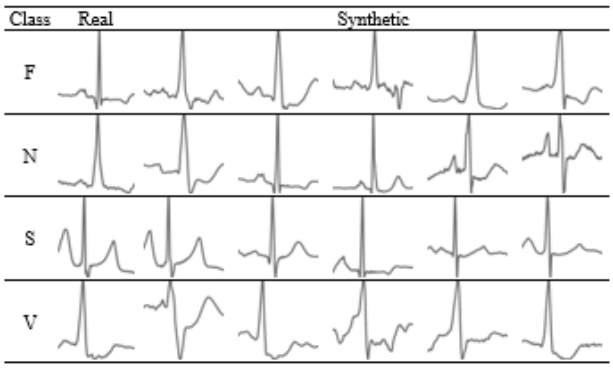


Figure 4. Synthetic images obtained with SMOTE.

2.5. Variational Autoencoder (VAE)

The main purpose of the autoencoder is to compress data with very low losses. It consists of two parts: an encoder and a decoder. The encoder compresses the data, and the decoder decompresses the compressed data. Compression and low-level data representations with the encoder are expressed as latent vectors. In standard autoencoders, each input (x) is encoded by passing through the encoder and converted into a latent vector. This vector is called code. Then, the code is passed through the decoder and restored to its original state (x'). In VAE, conversely, by adding the variation term to the autoencoder, the encoder output is obtained as mean (z_μ) and variance (z_σ) values. The code (z) is obtained by sampling from these values. The decoder part is the same in the autoencoder and VAE [10], [31].

$$z_\mu, z_\sigma = E(x; Q_e), z = z_\mu + \epsilon z_\sigma = q(z|x) \quad (3)$$

As given in Equation 3, the z code is derived from the real values of the data, $q(z|x)$. The reconstructed data x' is obtained from the distribution expressed by $p(x|z)$ (Equation 4).

$$x' = G(z; Q_d) = p(x|z) \quad (4)$$

For synthetic data generation with VAE over ECG signals, a 2-layer CNN architecture is used in the encoder and decoder. For the encoder, the filter size in the layers is 3x3, and the number of filters is 8 and 16. For the decoder, the filter size is 3x3 and the filter numbers are 16 and 8. Learning rate 0.0001, epoch 100, and mini batch size 32 were used in the model's training process. Samples of synthetic images obtained with VAE are given in Figure 5.

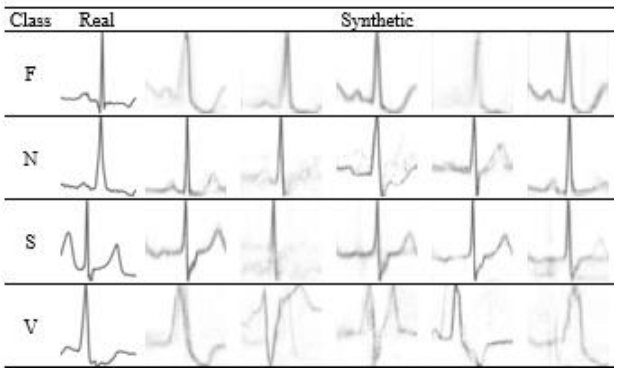


Figure 5. Synthetic images obtained with VAE.

3. Experimental Results

Our study performed ten different analyses for ECG classification with real and synthetic data. The first of these analyses was made with real data, and the other was performed with synthetic and real data. Three analyses were performed for each of the three data generation methods. While performing the analyses, 70%, 15%, and 15% of the dataset were used for training, validation, and testing, respectively. Specificity, accuracy, recall, and f1-score were used to evaluate its performance.

3.1. Classification Results with Real Data

Heartbeat images obtained from the MIT-BIH dataset were used to analyze real data. The dataset has an extremely unbalanced structure in its current form. No action was taken to solve the unbalanced problem in the data set. The dataset contains 749 samples for class F, 10440 for class N, 2612 for class S, and 4428 for class V. Training and validation curves obtained after training with the real dataset are given in Figure 6.

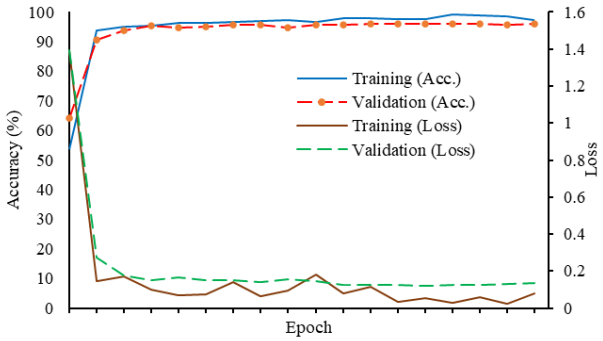


Figure 6. Training and validation curves with the real dataset.

When the graph obtained with the real data set is examined, the model's training has been carried out successfully, even though the data set is unbalanced. In addition, since there is no significant difference between training and validation, the overfitting problem has not been encountered. To analyze the results obtained with the real data in more detail, classification was made with the test dataset, and the details are given in Table 3.

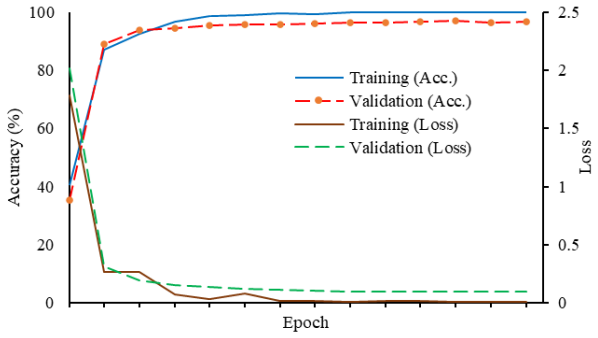
Table 3. ECG classification metrics with real dataset

Class	Acc. (%)	Specificity	Recall	f1-score	Overall
F	74.11	0.9850	0.9088	0.9283	96.53 %
N	98.40				
S	93.11				
V	97.89				

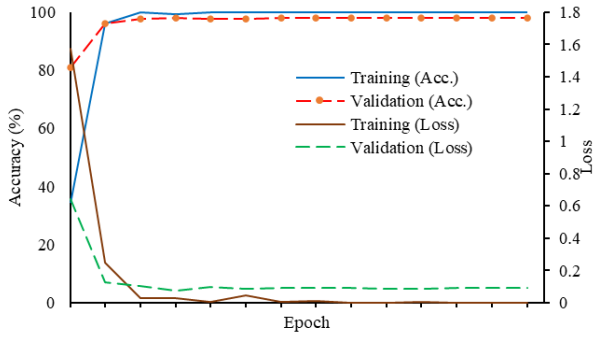
When the results in Table 3 are examined, the overall classification accuracy with the test data set is 96.53%. When class-based classification accuracies were examined, the lowest success was obtained in the F class with the lowest number of samples. In addition, class-based classification accuracies vary in proportion to the number of class samples.

3.2. Classification Results with Synthetic Data

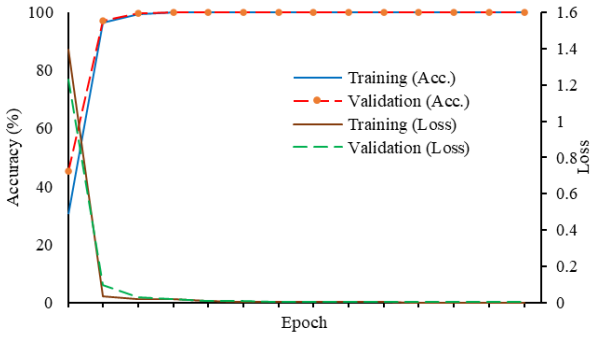
GAN, SMOTE, and VAE methods were used to generate synthetic data. With these methods, 1000 synthetic samples were produced for each class. The training and validation curves obtained after training the CNN model with the synthetic dataset are given in Figure 7.



a)



b)



c)

Figure 7. Training and validation curves with synthetic data a) GAN, b) SMOTE, c) VAE

There is no significant difference between training and validation values in all three data generation methods. After a certain iteration, it continued horizontally in three graphs. However, the VAE method has higher training and lower loss values than the other methods. To test the validity of these results obtained after the training process, classification was made with the test dataset, and the details are given in Table 4.

Table 4. ECG classification metrics with a synthetic dataset

Method	Accuracy (%)				Specificity	Recall	f1-score	Overall Acc. (%)
	F	N	S	V				
GAN	100.00	95.33	94.67	96.67	0.9889	0.9667	0.9666	96.67
SMOTE	98.00	98.67	96.67	98.00	0.9928	0.9783	0.9784	97.83
VAE	100.00	100.00	100.00	100.00	1	1	1	100.00

When the results given in Table 4 are examined, it is seen that the synthetic data are classified with high accuracy. Also, class-based accuracy rates are evenly distributed. However, the fact that all metrics of the VAE model are 100% raises the question of whether the model has an overfitting problem. Although the training, validation, and test data show that there is no overfitting problem, the use of real data and synthetic data in the next analysis will further clarify the result.

3.3. Results of Training with Synthetic Data and Testing with Real Data

In this analysis, GAN, SMOTE, and VAE models trained in the previous analysis were tested with real ECG data to examine the effectiveness of synthetic data. Random 500 samples from each class of real ECG data were used for the test. The results obtained are given in Table 5.

Table 5. Results of the test were obtained using synthetic data and training real data

Method	Accuracy (%)				Specificity	Recall	f1-score	Overall Acc. (%)
	F	N	S	V				
GAN	85.25	82.02	75.15	96.97	0.9495	0.8485	0.8476	84.85
SMOTE	67.88	89.72	90.30	97.58	0.9546	0.8637	0.8616	86.37
VAE	97.37	94.34	68.69	82.63	0.9525	0.8576	0.8550	85.76

According to the results given in Table 5, the three methods provided similar overall accuracy rates. However, when class-based accuracies were examined, there were differences according to the methods. The highest accuracy in the F and N classes was obtained with the VAE. In S and V classes, the SMOTE method provided higher accuracy than other methods.

3.4. Classification Results with Hybrid Data

This analysis created a dataset of 2000 samples for each class using 1000 randomly selected samples from the real and synthetic datasets. However, since the real dataset in the F class is 749, synthetic data is used more in this class. This data set was randomly divided into 70% training, 15% validation, and 15% testing. Then, the training process was carried out using the CNN model in other analyses. The training and validation curves obtained after training the CNN model are given in Figure 8.

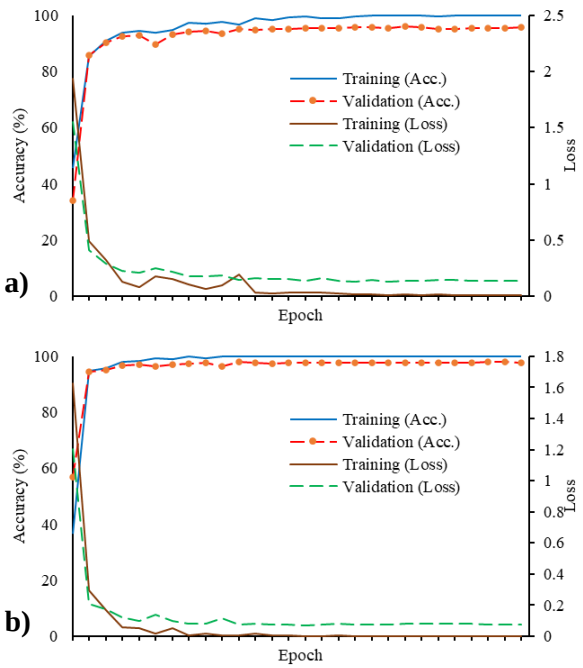


Figure 8. Training and validation curves with synthetic and real data **a)** GAN+Real, **b)** SMOTE+Real

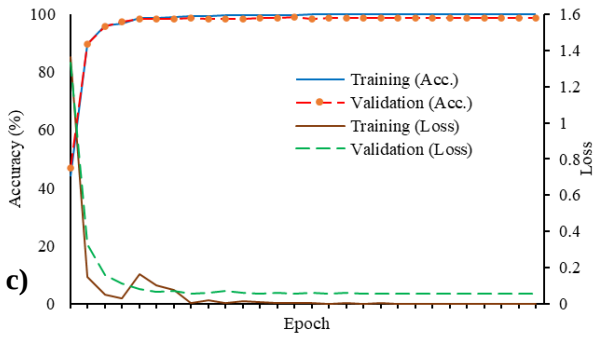


Figure 8. (Continuous) Training and validation curves with synthetic and real data **c)** VAE+Real

There is no significant difference between training and validation values in all three trainings. After a certain iteration, it continued horizontally in three graphs. However, the VAE method has higher training and lower loss values than the other methods. To test the validity of these results obtained after the training process, classification was made with the test dataset, and the details are given in Table 6.

Table 6. ECG classification metrics with synthetic and real datasets

Method	Accuracy (%)				Specificity	Recall	f1-score	Overall Acc. (%)
	F	N	S	V				
GAN+Real	96.00	95.00	96.00	97.33	0.9869	0.9608	0.9609	96.08
SMOTE+Real	98.33	99.67	95.67	97.33	0.9925	0.9775	0.9775	97.75
VAE+Real	98.33	99.67	97.67	98.67	0.9953	0.9858	0.9858	98.58

When the results in Table 6 are examined, the combined use of synthetic and real datasets increased the overall classification accuracy and contributed to the balanced and high class-based accuracies. The highest accuracy was obtained with the VAE method if analyzed on a method basis.

4. Discussion and Conclusion

ECG contains a visual representation of electrical signals related to heart rhythm. The amount of data is very large, as these signals can be one day or longer depending on the duration of the symptoms. Although the amount of data is very large, some arrhythmias are rare so that the data can have an extremely unbalanced structure. A similar situation exists in the MIT-BIH dataset used in this study. When the sample distribution of the 4-classes used is examined, there are 749 examples in the F class and 10440 in the N class.

This study analyzed the effectiveness of GAN, SMOTE, and VAE methods on ECG data to solve the unbalanced data problem and generate a limited number of data synthetically. For this purpose, real datasets, only synthetic datasets, synthetic data, and training and testing with real data and analyses were made using both datasets. The accuracy of these analyses on the test data set is summarized in Table 7.

Table 7. Summary of test accuracies obtained as a result of this study

Data	Accuracy (%)				Overall (%)
	F	N	S	V	
Real	74.11	98.40	93.11	97.89	96.53
GAN	100.00	95.33	94.67	96.67	96.67
SMOTE	98.00	98.67	96.67	98.00	97.83
VAE	100.00	100.00	100.00	100.00	100.00
Training GAN Test Real	85.25	82.02	75.15	96.97	84.85
Training SMOTE Test Real	67.88	89.72	90.30	97.58	86.37
Training VAE Test Real	97.37	94.34	68.69	82.63	85.76
GAN+Real	96.00	95.00	96.00	97.33	96.08
SMOTE+Real	98.33	99.67	95.67	97.33	97.75
VAE+Real	98.33	99.67	97.67	98.67	98.58

With the real dataset, the overall accuracy is 96.53%. However, in class-based accuracies, the accuracy of the F class is low. Also, class-based accuracies are proportional to the number of samples it contains. Class-based accuracies were balanced in the analysis results made with only synthetic data, and VAE achieved 100% classification success. The fact that all classes have 100% success in the VAE method

creates the question of whether there may be an overfitting problem. However, the highest classification success was also obtained with VAE in the hybrid dataset. Although training ECG signals with only synthetic data and testing with real data is lower than other methods, it has achieved a classification success of over 85%. The use of data augmentation increased the overall classification success by 2.05%.

The comparison of our results with the results of studies using the MIT-BIH database and data augmentation in the literature is given in Table 8.

Table 8. Comparison of our analysis results with the literature

Model	Data Aug.	Acc. (%)	References
CNN	SMOTE	92.7	[7]
CNN	TCGAN	94.7	[21]
CNN	**GAN**	**96.08**	**This study**
CNN	**SMOTE**	**97.75**	**This study**
CNN	SMOTE	98.0	[3]
CNN	SMOTE	98.3	[8]
CNN	GAN	98.3	[3]
CNN	VAE-ACGAN	98.5	[10]
CNN	**VAE**	**98.6**	**This study**
CNN	GAN	99.0	[2]
CNN + (ensemble) CNN-LSTM	SMOTE	99.0	[5]
ResNet-BiLSTM	GAN	99.4	[23]

The accuracy rates obtained in this study are similar to those of the studies in the literature. The most important difference from the literature studies is that data augmentation was used, and the classification success of synthetic data was also investigated. Synthetic data generation is usable for representing ECG signals if the results are generally evaluated. Models can only be trained with synthetic data and used with real data. In addition, how much the real data used in the production of synthetic data represents, the problem directly affects the results. Although the general success of the data generation methods used is close, there are differences in class-based success. Therefore, using hybrid data augmentation methods will provide higher classification accuracies. Another process that affects the production of synthetic data is the architecture of the methods used. The layers used for GAN and VAE and the parameters of these layers directly affect the result. Neighborhood parameters for SMOTE are similarly effective in achieving the result. As a result, GAN, SMOTE, and VAE methods can be used successfully for synthetic data generation and positively affect classification success.

The main purpose of this study is not to achieve high classification accuracy. For this reason, the CNN model was used at the basic level. In future studies, higher success can be achieved by developing the CNN model, which is used for synthetic data production. In addition, the variability of class-based achievements according to the model indicates that hybrid data generation methods will show higher success.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] H. Yang, J. Liu, L. Zhang, Y. Li, and H. Zhang, 'Proegan-ms: A progressive growing generative adversarial networks for electrocardiogram generation', *IEEE Access*, vol. 9, pp. 52089–52100, 2021.

[2] Z. Zhou, X. Zhai, and C. Tin, 'Fully automatic electrocardiogram classification system based on generative adversarial network with auxiliary classifier', *Expert Systems with Applications*, vol. 174, p. 114809, 2021.

[3] A. M. Shaker, M. Tantawi, H. A. Shedeed, and M. F. Tolba, 'Generalization of convolutional neural networks for ECG classification using generative adversarial networks', *IEEE Access*, vol. 8, pp. 35592–35605, 2020.

[4] W. Li, Y. M. Tang, K. M. Yu, and S. To, 'SLC-GAN: An automated myocardial infarction detection model based on generative adversarial networks and convolutional neural networks with single-lead electrocardiogram synthesis', *Information Sciences*, vol. 589, pp. 738–750, 2022.

[5] H. M. Rai, K. Chatterjee, and S. Dashkevych, 'The prediction of cardiac abnormality and enhancement in minority class accuracy from imbalanced ECG signals using modified deep neural network models', *Computers in Biology and Medicine*, vol. 150, p. 106142, 2022.

[6] A. Salazar, L. Vergara, and G. Safont, 'Generative Adversarial Networks and Markov Random Fields for oversampling very small training sets', *Expert Systems with Applications*, vol. 163, p. 113819, 2021.

[7] S. S. Aphale, E. John, and T. Banerjee, 'ArrhyNet: a high accuracy arrhythmia classification convolutional neural network', in *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, IEEE, 2021, pp. 453–457.

[8] S. K. Pandey and R. R. Janghel, 'Automatic detection of arrhythmia from imbalanced ECG database using CNN model with SMOTE', *Australasian physical & engineering sciences in medicine*, vol. 42, no. 4, pp. 1129–1139, 2019.

[9] H. Ge, K. Sun, L. Sun, M. Zhao, and C. Wu, 'A selective ensemble learning framework for ECG-based heartbeat classification with imbalanced data', in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2018, pp. 2753–2755.

[10] C. Du, P. X. Liu, and M. Zheng, 'Classification of imbalanced electrocardiosignal data using convolutional neural network', *Computer Methods and Programs in Biomedicine*, vol. 214, p. 106483, 2022.

[11] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, 'Synthetic data augmentation using GAN for improved liver lesion classification', in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, IEEE, 2018, pp. 289–293.

[12] Z. Qin, Z. Liu, P. Zhu, and Y. Xue, 'A GAN-based image synthesis method for skin lesion classification', *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105568, 2020.

[13] T. Golany, G. Lavee, S. T. Yarden, and K. Radinsky, 'Improving ECG classification using generative adversarial networks', in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, pp. 13280–13285.

[14] T. D. Tran, T. T. K. Dang, and N. Q. Tran, 'An Innovative Approach for Long ECG Synthesis with Wasserstein GAN Model', in *Computational Science and Its Applications – ICCSA 2024*, O. Gervasi, B. Murgante, C. Garau, D. Taniar, A. M. A. C. Rocha, and M. N. Faginas Lago, Eds., Cham: Springer Nature Switzerland, 2024, pp. 339–351. doi: 10.1007/978-3-031-64608-9_22.

[15] F. Zhou and J. Li, 'ECG data enhancement method using generate adversarial networks based on Bi-LSTM and CBAM', *Physiol. Meas.*, vol. 45, no. 2, p. 025003, Feb. 2024, doi: 10.1088/1361-6579/ad2218.

[16] C. Esteban, S. L. Hyland, and G. Rätsch, 'Real-valued (medical) time series generation with recurrent conditional gans', *arXiv preprint arXiv:1706.02633*, 2017.

[17] N. Wulan, W. Wang, P. Sun, K. Wang, Y. Xia, and H. Zhang, 'Generating electrocardiogram signals by deep learning', *Neurocomputing*, vol. 404, pp. 122–136, 2020.

[18] D. Hazra and Y.-C. Byun, 'SynSigGAN: Generative adversarial networks for synthetic biomedical signal generation', *Biology*, vol. 9, no. 12, p. 441, 2020.

[19] E. Piacentino, A. Guarner, and C. Angulo, 'Generating synthetic ecgs using gans for anonymizing healthcare data', *Electronics*, vol. 10, no. 4, p. 389, 2021.

[20] F. Zhu, F. Ye, Y. Fu, Q. Liu, and B. Shen, 'Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network', *Scientific reports*, vol. 9, no. 1, pp. 1–11, 2019.

[21] Y. Xia, Y. Xu, P. Chen, J. Zhang, and Y. Zhang, 'Generative adversarial network with transformer generator for boosting ECG classification', *Biomedical Signal Processing and Control*, vol. 80, p. 104276, 2023.

[22] R. R. Sarra, A. M. Dinar, M. A. Mohammed, M. K. A. Ghani, and M. A. Albahar, 'A Robust Framework for Data Generative and Heart Disease Prediction Based on Efficient Deep Learning Models', *Diagnostics*, vol. 12, no. 12, p. 2899, 2022.

[23] S. Ma, J. Cui, W. Xiao, and L. Liu, 'Deep Learning-Based Data Augmentation and Model Fusion for Automatic Arrhythmia Identification and Classification Algorithms', *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[24] S. Huang, P. Wang, and R. Li, 'Noise ECG generation method based on generative adversarial network', *Biomedical Signal Processing and Control*, vol. 81, p. 104444, 2023.

[25] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, 'SMOTE: synthetic minority over-sampling technique', *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[26] G. B. Moody and R. G. Mark, 'The impact of the MIT-BIH Arrhythmia Database', *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, May 2001, doi: 10.1109/51.932724.

[27] Ö. Yakut, S. Solak, and E. Bolat, 'IIR Based Digital Filter Design for Denoising the ECG Signal', *Journal Of Polytechnic*, vol. 21, no. 1, 2018, doi: 10.2339/politeknik.386970.

[28] M. A. Kızrak and B. Bolat, 'Derin öğrenme ile kalabalık analizi üzerine detaylı bir araştırma', *Bilişim Teknolojileri Dergisi*, vol. 11, no. 3, pp. 263–286, 2018.

[29] A. Courville and Y. Bengio, 'Generative adversarial nets', *Advanc in Neural*, 2014.

[30] Y. Xiong, L. Wang, Q. Wang, S. Liu, and B. Kou, 'Improved convolutional neural network with feature selection for imbalanced ECG Multi-Factor classification', *Measurement*, vol. 189, p. 110471, 2022.

[31] C. Güzel Turhan and H. Bilge, 'Scalable image generation and super resolution using generative adversarial networks', *Journal of the Faculty of Engineering and Architecture of Gazi University*, vol. 35, no. 2, 2020.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1169-1180 YEAR: 2024
DOI: 10.17798/bitlisfen.1539189

Detection of Stroke (Cerebrovascular Accident) Using Machine Learning Methods

Hadice ATEŞ[1], Abidin ÇALIŞKAN[2*]

[1]*Batman University, Department of Electrical and Electronics Engineering, Batman, Türkiye*
[2]*Batman University, Department of Computer Engineering, Batman, Türkiye*
(ORCID: 0000-0001-5039-6400) (ORCID: 0000-0002-0462-6997)

Keywords: Stroke, Artificial intelligence, Machine Learning, Classification, Accuracy.

Abstract

Stroke occurs when the blood flow to the brain is suddenly interrupted. This interruption can lead to the loss of function in the affected area of the brain and cause permanent damage to the corresponding part of the body. Stroke can develop due to various factors such as age, occupation, chronic diseases, and a family history of stroke. Assessing these factors and predicting stroke risk is often a costly and time-consuming process, which can increase the risk of permanent damage for the individual. However, with today's technology, Artificial Intelligence (AI) and Machine Learning (ML) models can process millions of data points to determine stroke risk within seconds. In this study, the risk of stroke in individuals is predicted most reliably using ML methods such as Logistic Regression (LR), Decision Tree (DT), Support Vector Machines (SVM), and k-Nearest Neighbors (KNN), with the aim of saving time, protecting human health, and enabling early diagnosis of the disease. As a result of the study, the highest accuracy rate was achieved by the DT model with 91%. The accuracy rates of the other models were found to be 89% for SVM, 81% for KNN, and 75% for LR.

1. Introduction

Stroke, medically referred to as “stroke” or “paralysis,” is a condition that leads to the loss of muscle movement in a specific area of the body or the entire body [1]. It typically occurs due to an interruption or reduction in blood flow to the brain. This interruption causes brain cells to be damaged due to a lack of oxygen and nutrients [1], [2]. Depending on the region of the brain affected by the stroke, it can lead to various sensory and motor function losses [2]. Stroke is categorized into three main types:

Ischemic stroke occurs as a result of a blockage in an artery that supplies blood to the brain and is the most common type of stroke. It usually occurs due to the blockage of blood vessels by causes such as blood clots or atherosclerosis.

Hemorrhagic stroke happens when a blood vessel in the brain ruptures and blood leaks into the brain tissue. This type of stroke is associated with conditions such as high blood pressure, aneurysms, or head trauma.

A transient ischemic attack is a temporary condition caused by a short-term reduction in blood flow to the brain. It typically lasts for a few minutes and does not cause permanent damage but can be a warning sign of a more severe stroke [3], [4].

The symptoms of a stroke appear suddenly and usually include sudden numbness and weakness in the face, arm, or leg, sudden confusion, difficulty speaking or understanding speech, sudden loss of vision or blurred vision, sudden difficulty walking, dizziness, loss of balance or coordination, and sudden severe headache. The presence of any of these symptoms requires urgent medical intervention [5].

Several factors increase the risk of stroke, including high blood pressure, diabetes, high cholesterol, smoking, obesity, physical inactivity, a family history of stroke, and heart disease.

*Corresponding author: abidin.caliskan@batman.edu.tr
Received: 27.08.2024, Accepted: 08.09.2024

Understanding these risk factors is crucial for stroke prevention and early intervention [6].

Stroke treatment varies depending on the type and severity of the stroke. In the case of an ischemic stroke, medications or surgical interventions may be used to dissolve or remove the clot. For hemorrhagic stroke, the goal is to stop the bleeding and reduce the pressure on the brain. The rehabilitation process includes methods such as physical therapy, speech therapy, and occupational therapy. Early diagnosis and treatment can significantly reduce the effects of stroke and increase the chances of recovery [6], [7].

The diagnosis of stroke is made through physical examination and clinical evaluation. The physician evaluates the patient's symptoms, medical history, and when the stroke symptoms began. This includes checking consciousness, speech ability, mobility, balance, and coordination. Additionally, the "FAST" test is used for rapid diagnosis; this test focuses on factors such as the face, arms, speech, and time [8].

AI and ML are two significant concepts that have a profound impact on the modern technology world. AI is a field of science aimed at enabling computers and machines to exhibit human-like intelligence behaviors. These behaviors include abilities such as problem-solving, learning, reasoning, and decision-making. ML, on the other hand, is a subfield of AI that allows computers to automatically perform specific tasks by learning from data. ML enables systems to improve their performance through experiences and examples without being explicitly programmed [9].

AI and ML are used in many fields. These technologies have created a significant revolution in the healthcare sector. They are particularly used in areas such as early diagnosis of diseases, treatment planning, drug discovery, and patient monitoring. AI-based systems can analyze large datasets to detect disease symptoms more quickly and accurately, contributing to the development of more effective treatment processes. ML models, trained on millions of medical data points, can successfully perform complex tasks such as predicting stroke risk or forecasting disease progression [10].

The use of AI and ML in the healthcare sector also makes it possible to develop clinical decision support systems. These systems assist doctors in making more accurate diagnoses and treatment decisions by providing better insights into patients' conditions. Additionally, with these technologies, patient data can be analyzed in real-time, enabling the early identification and prevention of potential complications. By making healthcare services more efficient, accessible, and personalized, AI and ML offer significant conveniences in the field of healthcare [11], [12].

2. Literature Review

Several studies have been conducted on stroke prediction datasets in the literature, utilizing various AI models.

Emon et al. (2020) evaluated 10 different ML methods for stroke prediction and found that the Weighted Voting Algorithm provided the highest performance with an accuracy rate of 97% [13].

Singh and Choudhary predicted stroke risk using AI techniques, achieving an accuracy rate of 97.7% by using the CART algorithm for feature selection, PCA for dimensionality reduction, and Backpropagation Neural Network (BP) for classification [14].

Sevli identified stroke risk using the RFC technique and achieved an accuracy rate of 98.84% by applying resampling to solve data problems; the study found that age, body mass index, and glucose level were significant in risk prediction [15].

Revanth et al. (2020) compared SVM, CART, RFC, and MLP models for stroke risk prediction and reported that SVM demonstrated the best performance with an accuracy rate of 98.99% [16].

Cheon et al. (2019) applied DO techniques using a large patient dataset and achieved an AUC value of 83.48% by performing dimensionality reduction with PCA, indicating relatively high model accuracy [17].

Shoily et al. (2019) compared the Naive Bayes, J48, k-Nearest Neighbor (KNN), and Random Forest algorithms for stroke detection and found that Naive Bayes performed worse with an accuracy rate of 85.6%, while J48, KNN, and Random Forest algorithms produced better results with an accuracy rate of 99.8% [18].

Pradeepa et al. (2020) applied spectral clustering and various ML algorithms to identify stroke symptoms using social media data, showing that PNN provided the highest performance with an accuracy rate of 89.90% [19].

Li et al. (2019) used LR, Naive Bayes, Bayesian Network, DT, Neural Network, Random Forest, and other algorithms to classify stroke risk levels, reporting that the boosting model created with DT achieved the best performance with a recall rate of 99.94%, while Random Forest stood out with a precision rate of 97.33% [20].

In Table 1, the performance values of the study conducted and literature studies have been compared.

Table 1. Performance Comparisons in the Literature

Authors	Year of Study	Dataset Used	Methods	Accuracy Rate
Proposed study	**2024**	**Stroke Prediction Dataset**	**LR** **KNN** **DVM** **DT**	**%75** **%89** **%81** **%91**
Emon and Choudhary	2020	Stroke Prediction Dataset	Weighted Voting Algorithm	%97
Singh et al.	2017	Cardiovascular Health Study (CHS) dataset	CART algorithm, Dimensionality reduction with PCA, Backpropagation Neural Network (BP)	%97
Sevli	2021	Stroke Prediction Dataset	RFC technique, resampling	%98
Revanth et al.	2020	Stroke Prediction Dataset	DVM, CART, RFC, MLP models	%98
Cheon et al.	2019	China National Stroke Screening Data	PCA	%83
Shoily et al.	2019	Stroke Prediction Dataset	Naive Bayes, J48, KNN, Random Forest	%85
Pradeepa et al.	2020	Social Media Content	Various ML algorithms	%89
Li et al.	2019	China National Stroke Screening Data	LR, DT, Naive Bayes	%97

3. Materials and Methods

3.1. Lojistic Regression

LR is a statistical analysis method primarily used to establish relationships between two or more categorical outcomes [21]. This model is widely preferred in cases where the dependent variable is binary (e.g., “yes” or “no”) or multi-class. Unlike linear regression, LR predicts the probabilities of outcomes and uses a logistic function that limits these probabilities between 0 and 1 [22].

In the LR model, the effect of independent variables (input variables) on the dependent variable (output variable) is examined. This effect is used to predict the likelihood of an event occurring (e.g., the probability of a patient contracting a particular disease) [23]. The model's fundamental assumption is that the probability distribution of the dependent variable can be well modeled by a logistic function. This function expresses the probability of the dependent variable as a linear combination of the independent variables [24].

LR is applied in many fields, including medical diagnosis, marketing analytics, social sciences, and credit risk assessment. The model allows for both understanding the impact of independent variables and producing useful results in classification problems [25].

This model is particularly powerful in classification problems and can yield highly accurate results when the structure of the data and the nature of the problem are suitable [26].

3.2. K-Nearest Neighbors

The working principle of the KNN algorithm is quite simple: when a new data point needs to be classified, the distance between this point and the other points in the dataset is calculated [27]. These distances are usually measured using metrics such as Euclidean distance or Manhattan distance. As a result of these calculated distances, the nearest K neighbors are selected, and the majority class of these neighbors is assigned as the predicted class for the new data point. The value of K is a parameter that determines the number of neighbors and directly affects the model's performance [28].

The KNN algorithm does not perform any complex calculations during model training, which classifies it as a “lazy learning” algorithm [29]. Although this feature means that KNN can be slow for large datasets, it can produce highly effective results for small and medium-sized datasets [30].

KNN is widely used in fields such as medicine, bioinformatics, marketing analytics, and recommendation systems. The simplicity and interpretability of the algorithm make it an attractive choice for data scientists and analysts.

However, the performance of the KNN algorithm depends on the chosen value of K, the distance metric used, and the characteristics of the dataset [31]. Since KNN relies on the proximity relationships of data points, it is important to properly scale the data and carefully select the features [32].

3.3. Support Vector Machines

SVM is a supervised learning method and a powerful algorithm used to solve classification and regression problems. SVM is particularly known as a flexible and robust classification technique that yields effective results in high-dimensional datasets [33].

The fundamental principle of SVM is to find a hyperplane (or decision boundary) that best separates the data. This hyperplane aims to separate data points from different classes with the maximum margin. The margin refers to the distance between the hyperplane and the nearest data points. By finding the hyperplane that maximizes this margin, SVM ensures accurate classification of the data [33], [34].

If the data is linearly separable, SVM directly finds this hyperplane. However, if the data is not linearly separable, SVM uses kernel functions to transform the data into a higher-dimensional space. This transformation allows the data to become linearly separable, enabling SVM to perform effective classification even on complex datasets. Commonly used kernel functions include linear, polynomial, radial basis function (RBF), and sigmoid functions [35].

One of the advantages of SVM is its ability to generalize well on large datasets. Additionally, its resistance to overfitting makes SVM a reliable model. However, the computational cost of SVM can be high, especially for large and complex datasets. Moreover, selecting optimal hyperparameters (e.g., the kernel function and regularization parameter) directly affects the model's performance and requires careful tuning [36], [37].

SVM is widely used in various fields such as bioinformatics, image processing, text classification, and genetic analysis. The high accuracy rates it demonstrates in classification tasks make SVM a significant tool in data science and ML projects [38].

3.4. Decision Trees

DT is a popular ML algorithm used to solve both classification and regression problems. DTs are employed to make decisions based on certain features of the data and to visualize these decisions in a simple and understandable way. This algorithm creates a model that resembles a tree structure; this model starts from the root node and branches out into branches and leaves [39].

DTs classify or predict data points based on the independent variables (features) in the dataset. Each node represents a feature, and the best splitting point for this feature is selected. This split is defined as a condition that divides the dataset into subgroups. The branching process continues until the data is fully separated or further splitting is no longer meaningful. Ultimately, the leaf nodes represent a specific class or value [40].

DTs work by creating clear decision rules on the data. For example, in a medical diagnosis system, DTs can be used to predict which disease a patient may have based on their age, symptoms, and test results. The tree evaluates the probabilities for each disease based on these factors and makes its final decision [41].

One of the biggest advantages of DTs is that the models they create are explainable and can be visualized. This allows for an easy understanding and interpretation of the model's decision-making process. Additionally, DTs can work with both numerical and categorical data and have the ability to handle missing data [42].

However, DTs also have some disadvantages. For example, an overgrown decision tree can negatively impact the model's performance. Pruning techniques can be applied to prevent this situation. Moreover, DTs may not be able to model the relationship between variables in a complex and non-linear way [43].

DTs have a wide range of applications in finance, marketing, medicine, biology, and many other fields. Furthermore, they form the basis for more complex algorithms such as Random Forests and Gradient Boosting. These algorithms combine multiple decision trees to create more robust and generalized models [44], [45].

3.5. Model Performance Metrics

Various measurement criteria are used to evaluate the performance of models used in ML methods. Among these criteria, accuracy, precision, sensitivity, and F1 score are the most common. It is important to calculate these metrics to

objectively assess the success of a model. Each of these metrics reflects different performance characteristics.

The accuracy metric, commonly used to evaluate algorithm performance, represents the ratio of correctly classified examples to the total predictions made by the model. This metric encompasses both true positive and true negative predictions and is calculated as shown in formula (1) [46], [47].

$$Accuracy = \frac{TP+TN}{TP+TN+TN+FN} \quad \ldots\ldots\ldots(1)$$

Precision refers to the proportion of truly positive examples among the instances that the model has classified as positive. This metric indicates the accuracy of the predictions for the positive class and is explained by formula (2) [48], [49].

$$Precision = \frac{TP}{TP+TF} \quad \ldots\ldots\ldots(2)$$

Sensitivity (or recall) shows how accurately the model identifies true positive examples. This metric determines how many of the positive class examples are correctly classified as positive and is calculated as shown in formula (3) [50], [51].

$$Recall = \frac{TP}{TP+FN} \quad \ldots\ldots\ldots(3)$$

The F1 score provides a balanced assessment of precision and sensitivity by taking their harmonic mean. It is used to evaluate the overall performance of the model, particularly in cases of class imbalance, and the calculation method is shown in formula (4) [52], [53].

$$F1\ Score = 2 * \frac{precision*recall}{precision+recall} \quad \ldots\ldots\ldots(4)$$

3.6. Dataset

In this study, the "Stroke Prediction Dataset" obtained from the Kaggle repository was used. This dataset includes features such as age, gender, and smoking status of individuals [54]. The dataset contains a total of 11 features, consisting of 10 input features and 1 output feature, aimed at classifying and predicting stroke risk. The features and their descriptions are listed below:

Age: The age of the individual, represented as numerical data.

Gender: Specifies the gender of the individual, represented as categorical data.

Hypertension: Indicates whether the individual has hypertension; represented as numerical data (1 – Hypertension present, 0 – No hypertension).

Work Type: Specifies the occupation of the individual, represented as categorical data (e.g., Government, Not working, Self-employed).

Heart Disease: Indicates whether the individual has heart disease; represented as numerical data (0 – No heart disease, 1 – heart disease present).

Marital Status: Shows the marital status of the individual, represented as categorical data (Yes, No).

Residence Area: Specifies the area where the individual lives, represented as categorical data (Urban, Rural).

Body Mass Index (BMI): Represents the body mass index of the individual, represented as numerical data.

Average Glucose Level: Indicates the average glucose level in the individual's blood, represented as numerical data.

Smoking Status: Specifies the smoking status of the individual, represented as categorical data (Previously smoked, currently smoking, never smoked).

Stroke Status: Indicates whether the individual has previously had a stroke; represented as numerical data (1 – Had a stroke, 0 – Did not have a stroke).

This dataset has been used in various academic research and theses, providing valuable information for assessing stroke risk [54].

4. Findings and Discussion

In the study, ML methods such as LR, KNN, SVM, and DT were used to detect the presence of stroke risk. To facilitate the readability of the results, performance metrics were visualized using the "Model Evaluation Matrix" and the "Confusion Matrix".

In the LR model, a moderate regularization with C = 0.6 has been applied, which helps prevent overfitting. L1 penalty (Lasso) has been chosen using penalty = '11', which simplifies the model by reducing some features to zero. Additionally, solver = 'liblinear' has been selected, which is a suitable and effective solver, especially for small datasets.

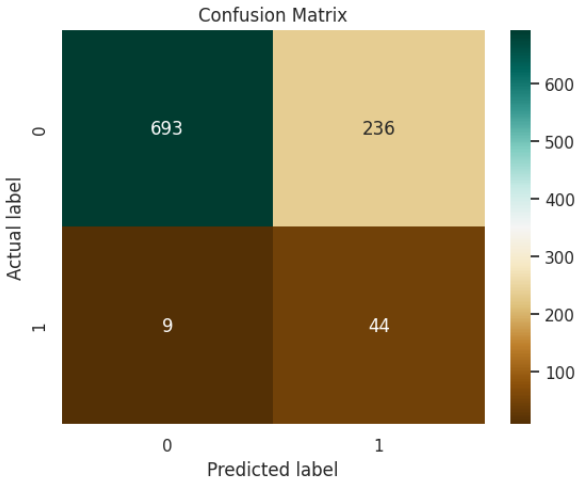


Figure 1. Confusion Matrix of the LR Model

Examining Figure 1, it can be seen that True Negatives (TN) (693) are correctly predicted. This high number of true negatives indicates that the model performs well in predicting negative classes. False Positives (FP) (236) are instances where true negatives were incorrectly predicted as positive. This indicates that the model misclassified some negative examples as positive. False Negatives (FN) (9) are instances where true positives were incorrectly predicted as negative. This suggests that the model missed some positive classes by predicting them as negative. True Positives (TP) (44) are instances where positive classes were correctly predicted. This shows that the model is successful in identifying positive classes, though there is room for improvement.

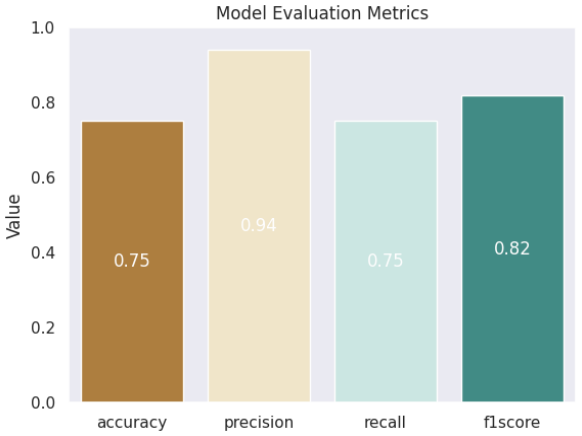


Figure 2. Model Evaluation Matrix of the LR Model

Examining Figure 2, it is observed that your model achieves high precision in classifying positive classes but may miss some positive examples, resulting in a moderate overall accuracy.

In the KNN model, the 'manhattan' metric has been identified as the most suitable option for measuring the distance between neighbors. With n_neighbors = 3, the best results were achieved by considering 3 neighbors. Furthermore, weights = 'distance' has been used to weight the neighbors based on their distances, meaning that closer neighbors have a greater impact on the prediction. With these hyperparameters, the model has achieved optimal performance based on the distance of neighbors.

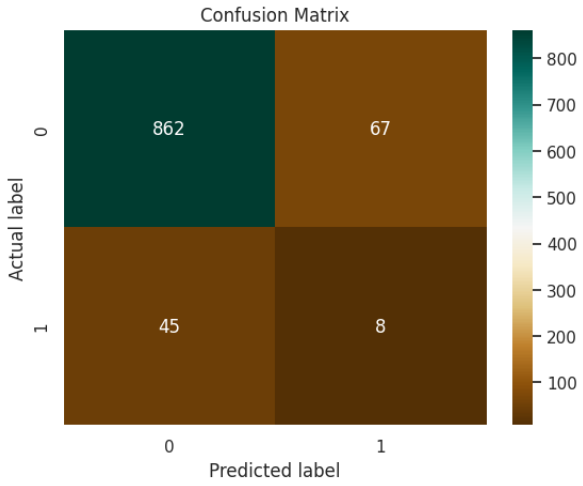


Figure 3. Confusion Matrix of the KNN Model

Figure 3 shows that TN (862) correctly predicts the true negative classes. This high true negative value indicates that the model predicts negative classes well. FP (67) shows that some true negatives are predicted as positive. This indicates that the model makes errors in predicting negative classes as positive. FN (45) shows that some true positives are predicted as negative. This means the model mistakenly predicts positive classes as negative, thus missing positive classes. TP (8) shows that the model correctly predicts true positives. This low true positive value indicates that the model is weak in identifying positive classes.

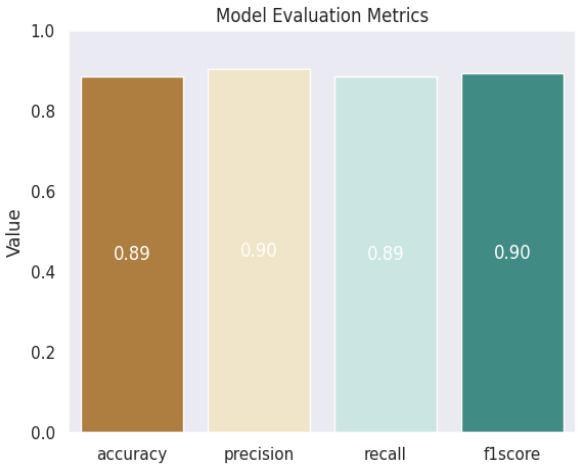


Figure 4. Model Evaluation Matrix of the KNN Model

Figure 4 shows that the model performs well overall, providing high precision and recall for positive classes and maintaining a good balance between precision and recall.

In the SVM model, the hyperparameters include a penalty parameter C = 10, which allows the model to learn a more complex boundary with a higher penalty. The 'rbf' kernel function has been chosen as it is effective on non-linear data. Additionally, the class_weight = 'balanced' setting attempts to balance the classes, optimizing the model's performance according to the proportion of each class in the dataset.

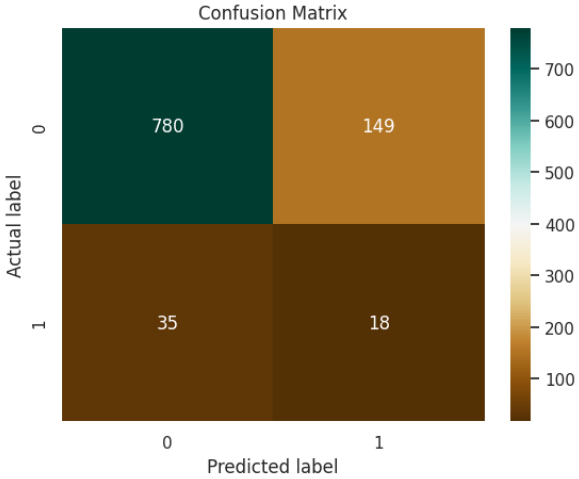


Figure 5. Confusion Matrix of the SVM Model

Figure 5 shows that TN (780) correctly predicts the true negative classes. This high true negative value indicates that the model predicts negative classes well. FP (149) shows that some true negatives are predicted as positive. This indicates that the model makes errors in predicting negative examples as positive. FN (35) shows that some true positives are predicted as negative. This means the model mistakenly predicts some positive examples as negative, thus missing positive classes. TP (18) shows that the model correctly predicts true positives. This low true positive value indicates that the model is weak in identifying positive classes.

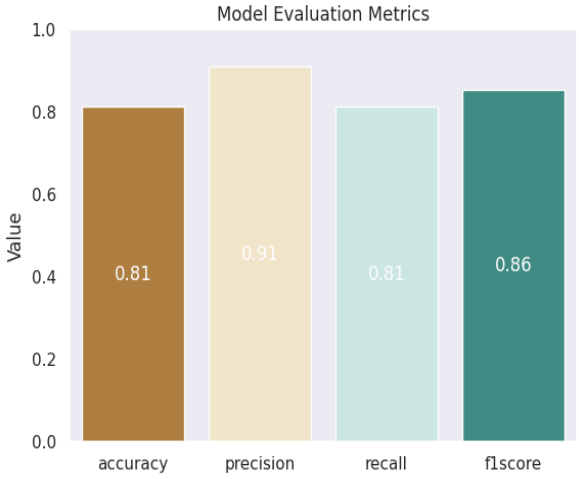


Figure 6. Model Evaluation Matrix of the SVM Model

Figure 6 shows that the model generally performs well, with high precision and strength in identifying positive classes. However, with a sensitivity of 81%, it suggests that the model might miss some positive examples.

For the KA model, the 'gini' impurity is used as the splitting criterion, which aims to increase the homogeneity of class distributions during splitting. The tree can grow up to a maximum depth of 20 levels (max_depth = 20), meaning it can have up to 20 branching layers. Additionally, at least 1 sample can be present in a leaf node (min_samples_leaf = 1).

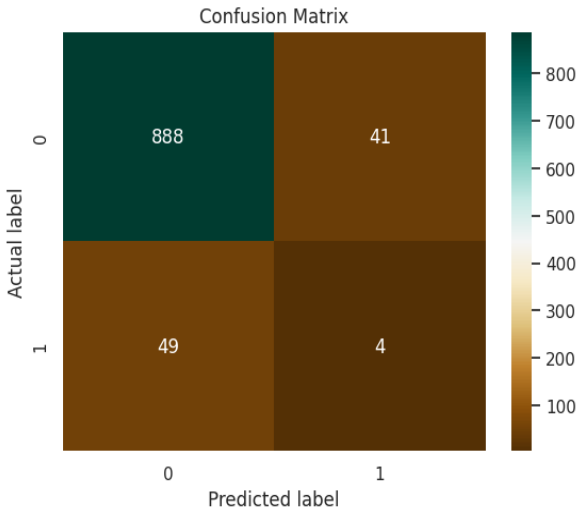


Figure 7. Confusion Matrix of the DT Model

Figure 7 shows that TN (888) correctly predicts the true negative classes. This high true negative value indicates that the model predicts negative classes very well. FP (41) shows that some true negatives are predicted as positive. This indicates that the model mistakenly predicts some negative examples as positive. FN (49) shows that some true positives are predicted as negative. This means the model mistakenly predicts some positive examples as negative, thus missing positive classes. TP (4) shows that the model correctly predicts true positives. This low true positive value indicates that the model is quite weak in identifying positive classes.

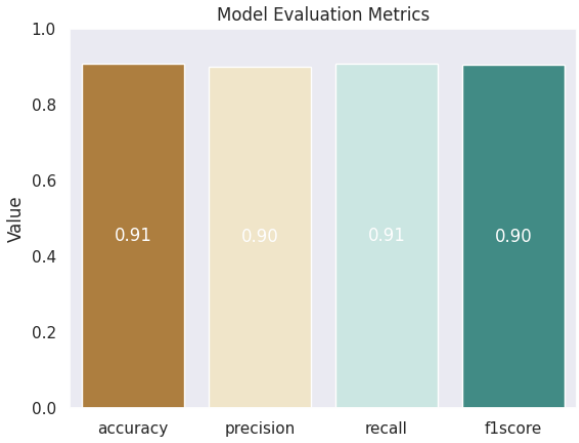


Figure 8. Model Evaluation Matrix of the DT Model

Table 2. Comparison of Performance Metrics for Used ML Models

Model	Class	Precision	Recall	F1-Score	Accuracy (%)
LR	No Risk	0.98	0.74	0.84	75,23
	Risk Present	0.15	0.83	0.26	
KNN	No Risk	0.95	0.92	0.92	88,83
	Risk Present	0.10	0.15	0.15	
SVM	No Risk	0.95	0.83	0.89	81,02
	Risk Present	0.10	0.33	0.16	
DT	No Risk	0.94	0.95	0.95	91,02
	Risk Present	0.08	0.07	0.08	

Figure 8 shows that the model has both high overall performance and high success in identifying positive classes. The model's results are quite satisfactory and can generally be considered reliable.

Table 2 compares the performance of the ML models used in the study for the “Risk Present” and “No Risk” classes. The LR model shows high precision in recognizing the “No Risk” class. However, recall is somewhat lower, indicating that some “Risk Present” cases may be missed. The F1-score and overall accuracy reflect this situation. The KNN algorithm demonstrates high performance for the “No Risk” class. Both precision and recall are quite high, resulting in good overall accuracy for the model. However, it shows poor performance in the “Risk Present” class, with both precision and recall being low, indicating difficulty in distinguishing this class. The SVM model exhibits very good performance for the “No Risk” class, with high F1-score and accuracy, though recall is somewhat lower. It is noted for having the highest performance in the “No Risk” class, with both precision and recall being quite high and the F1-score being close to perfection.

5. Conclusion and Suggestions

This study investigated the risk factors associated with stroke and assessed their impact on the development of the disease. The findings indicate that hypertension, diabetes, smoking, and high cholesterol levels increase stroke risk, with hypertension being the most significant determinant. Additionally, it was found that ischemic stroke is more prevalent, whereas hemorrhagic stroke is associated with higher mortality rates. The critical role of early diagnosis and intervention, particularly thrombolytic therapy and rehabilitation, in the recovery process of patients was emphasized.

The study developed a method for stroke detection and diagnosis using ML models and demonstrated that these models contribute to medical decision support systems. The model's accuracy was reported to be 91%. Future studies aim to enhance

success by employing different ML models and performance metrics, with more effective feature selection and data preprocessing steps.

It is recommended to explore additional ML models and advanced techniques to improve the accuracy and reliability of stroke risk prediction. Experimenting with ensemble methods and deep learning algorithms could provide better insights and enhance model performance.

Future research should focus on identifying and incorporating additional relevant features that could influence stroke risk. This may involve integrating new biomarkers, genetic factors, or lifestyle variables to create more comprehensive models.

Emphasizing rigorous data preprocessing and cleaning is crucial to ensure the quality and consistency of the data used in model training. Addressing issues such as missing values, outliers, and data imbalance will help improve model robustness.

It is important to validate the models on diverse and independent datasets to ensure their generalizability and applicability in different populations. Cross-validation techniques and external validation studies should be employed to confirm the model's effectiveness.

Efforts should be made to integrate these predictive models into clinical decision support systems to assist healthcare professionals in early stroke detection and management. Providing user-friendly interfaces and decision aids will enhance the practical utility of these models.

Regularly updating the models with new data and research findings is essential to maintain their relevance and accuracy. Continuous monitoring and iterative improvements based on real-world performance will contribute to better patient outcomes.

By addressing these recommendations, the accuracy and impact of stroke risk prediction models can be further improved, leading to more effective prevention and management strategies.

Contributions of the authors

A.Ç. and H.A. designed the study, performed the experiments and wrote the article. H.A. performed the calculations, A.Ç checked the language and contributed to the writing of the manuscript.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] S. Emon, M. M. Talukder, M. M. Rahman, and M. S. Islam, "Stroke prediction using machine learning algorithms," in *2020 IEEE Region 10 Symposium (TENSYMP)*, Dhaka, Bangladesh, 2020, pp. 101–104.

[2] M. Singh, V. S. Chouhan, and S. Gupta, "Prediction of stroke using artificial intelligence," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, Greater Noida, India, 2017, pp. 1467–1471.

[3] E. Sevli, "Prediction of stroke using Random Forest Classifier," in *2020 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, Sakheer, Bahrain, 2020, pp. 1–6.

[4] K. Revanth, G. Deepak, and K. N. V. Satya Prasad, "A comparative study on stroke prediction using various machine learning models," in *2020 International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2020, pp. 1385–1390.

[5] S. Cheon, D. A. Lee, and J. K. Kim, "Detection of stroke using deep learning algorithms," in *2019 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, 2019, pp. 1–4.

[6] S. Shoily, A. B. K. Singh, and M. Rahman, "Analysis of machine learning algorithms for stroke detection," in *2019 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, Rajshahi, Bangladesh, 2019, pp. 1–4.

[7] N. Pradeepa, R. Gupta, and R. Singh, "Detection of stroke symptoms using machine learning and social media analytics," in *2020 IEEE Symposium on Signal Processing and Information Technology (ISSPIT)*, Bilbao, Spain, 2020, pp. 1–6.

[8] H. Li, J. Liu, and Y. Wang, "Stroke risk prediction with various machine learning algorithms," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, San Diego, CA, USA, 2019, pp. 2148–2151.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[10] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2012, pp. 1106–1114.

[11] L. Zhang, "Machine Learning Techniques for Big Data Analysis," Ph.D. dissertation, Dept. of Computer Science, University of California, Berkeley, CA, USA, 2018.

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Cambridge, MA: MIT Press, 2016.

[13] M. U. Emon, M. S. Keya, T. I. Meghla, M. M. Rahman, S. Al Mamun, and M. S. Kaiser, "Performance analysis of machine learning," in *Fourth International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2020, pp. 1464–1469.

[14] M. S. Singh and P. Choudhary, "Stroke prediction using artificial intelligence," in *8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, Bangkok, Thailand, 2017, pp. 158–161.

[15] O. Sevli, "İnme (felç) riskinin MÖ kullanılarak tespiti," in *7. Uluslararası Mühendislik Mimarlık ve Tasarım Kongresi (7th International Congress on Engineering, Architecture and Design)*, İstanbul, Türkiye, 2021, pp. 661–667.

[16] S. Revanth, S. Sanjay, and V. Vijagayaganth, "Stroke prediction using machine learning algorithms," *International Journal of Disaster Recovery and Business Continuity*, vol. 11, no. 1, pp. 3081–3086, 2020.

[17] S. Cheon, J. Kim, and J. Lim, "The use of deep learning to predict stroke patient mortality," *International Journal of Environmental Research and Public Health*, vol. 16, no. 11, pp. 1–12, 2019.

[18] T. I. Shoily, T. Islam, S. Jannat, S. A. Tanna, T. M. Alif, and R. R. Ema, "Detection of stroke disease using machine learning algorithms," in *Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, 2019, pp. 1–6.

[19] S. Pradeepa, K. Manjula, S. Vimal, M. S. Khan, N. Chilamkurti, and A. K. Luhach, "DRFS: Detecting risk factor of stroke disease from social media using machine learning techniques," *Neural Processing Letters*, 2020, pp. 1–19.

[20] X. Li, D. Bian, J. Yu, M. Li, and D. Zhao, "Using machine learning models to improve stroke risk level classification methods of China national stroke screening," *BMC Medical Informatics and Decision Making*, vol. 19, pp. 1–7, 2019.

[21] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–220, 1972.

[22] M. E. T. M. MacKinnon, "Logistic regression analysis for binary data," *Journal of Applied Statistics*, vol. 28, no. 6, pp. 759–769, Jun. 2001.

[23] A. Agresti, "Categorical data analysis," *Journal of the American Statistical Association*, vol. 95, no. 450, pp. 1121–1130, Dec. 2000.

[24] S. M. Bhandari, J. P. Lee, and K. J. Kim, "Application of logistic regression in predicting heart disease," in *Proc. IEEE International Conference on Data Mining (ICDM)*, Dec. 2019, pp. 1124–1130.

[25] S. Wang and X. Zhang, "Improving logistic regression with feature selection for high-dimensional data," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2018, pp. 1552–1560.

[26] L. Huang and X. Liu, "Logistic regression for multiclass classification problems," in *Proc. IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2017, pp. 987–994.

[27] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[28] M. A. Hall and L. A. Smith, "Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper," *IEEE Transactions on Neural Networks*, vol. 15, no. 5, pp. 103–110, Sep. 2003.

[29] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognition*, vol. 45, no. 1, pp. 346–362, Jan. 2012.

[30] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, Dec. 2006, pp. 1473–1480.

[31] A. Beygelzimer, S. Kakade, and J. Langford, "Cover trees for nearest neighbor," in *Proc. 23rd International Conference on Machine Learning (ICML)*, Jun. 2006, pp. 97–104.

[32] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *Proc. IEEE International Conference on Computer Vision Theory and Applications (VISAPP)*, Feb. 2009, pp. 331–340.

[33] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[34] J. Weston and C. Watkins, "Support vector machines for multi-class pattern recognition," in *European Symposium on Artificial Neural Networks (ESANN)*, vol. 4, pp. 219–224, Apr. 1999.

[35] B. Scholkopf, A. Smola, R. Williamson, and P. Bartlett, "New support vector algorithms," *Neural Computation*, vol. 12, no. 5, pp. 1207–1245, May 2000.

[36] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge: Cambridge University Press, 2000.

[37] B. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annual ACM Workshop on Computational Learning Theory (COLT)*, Jul. 1992, pp. 144–152.

[38] A. J. Smola and B. Scholkopf, "A tutorial on support vector regression," in *Proc. International Conference on Artificial Neural Networks (ICANN)*, Aug. 1998, pp. 52–78.

[39] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar. 1986.

[40] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[41] S. Murthy, "Automatic construction of decision trees from data: A multi-disciplinary survey," *Data Mining and Knowledge Discovery*, vol. 2, no. 4, pp. 345–389, Dec. 1998.

[42] R. Kohavi, "Scaling up the accuracy of Naive-Bayes classifiers: A decision-tree hybrid," in *Proc. 2nd International Conference on Knowledge Discovery and Data Mining (KDD)*, Aug. 1996, pp. 202–207.

[43] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," in *Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, Sep. 2006, pp. 3–42.

[44] H. Blockeel and L. De Raedt, "Top-down induction of first-order logical decision trees," in *Proc. 9th International Conference on Machine Learning (ICML)*, Jul. 1998, pp. 55–63.

[45] V. Tümen and A. S. Sunar, "Predicting the work-life balance of employees based on the ensemble learning method," *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, vol. 12, no. 2, pp. 344–353, 2023.

[46] A. Çalışkan, "Finding complement of inefficient feature clusters obtained by metaheuristic optimization algorithms to detect rock mineral types," *Transactions of the Institute of Measurement and Control*, vol. 45, no. 10, pp. 1815–1828, 2023.

[47] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.

[48] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, Jul. 2009.

[49] S. B. Çelebi and Ö. A. Karaman, "Multilayer LSTM model for wind power estimation in the SCADA system," *European Journal of Technique (EJT)*, vol. 13, no. 2, pp. 116–122, 2023.

[50] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd International Conference on Machine Learning (ICML)*, Pittsburgh, PA, USA, Jun. 2006, pp. 233–240.

[51] V. Tümen, "SpiCoNET: A hybrid deep learning model to diagnose COVID-19 and pneumonia using chest X-ray images," *Traitement du Signal*, vol. 39, no. 4, pp. 1169, 2022.

[52] N. Chinchor, "MUC-4 evaluation metrics," in *Proc. 4th Conference on Message Understanding*, McLean, VA, USA, Jun. 1992, pp. 22–29.

[53] A. Çalışkan, "Classification of tympanic membrane images based on VGG16 model," *Kocaeli Journal of Science and Engineering*, vol. 5, no. 1, pp. 105–111, 2022.

[54] Kaggle, "Stroke Prediction Dataset," [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset. [Accessed: May 8, 2024].

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1181-1193 YEAR: 2024
DOI: 10.17798/bitlisfen.1539250

Multi-Region Detection of eye Conjunctiva Images Using DNCNN and YOLOv8 Algorithms

Emine CENGİL[1*]

[1] *Bitlis Eren University, Department of Computer Engineering, Bitlis, Türkiye*
(ORCID: 0000-0003-4313-8694)

Keywords: Convolutional Neural Network (CNN), Artificial Intelligence, YOLOv8, Image Enhancement, Object Detection.

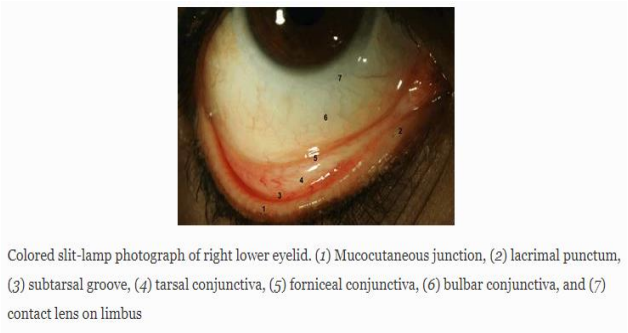
Abstract
Artificial intelligence is encountered in many areas today. It makes our lives easier with its use in our daily lives. With the advancement of medical big data and artificial intelligence, eye images have begun to be used in the detection of endocrine, cardiovascular, neurological, renal, hematological and many other diseases. It is possible to find more connections between systemic disorders and eye disorders and apply them to increase the effectiveness of artificial intelligence. The eye is an anatomically complex organ. Detection of the conjunctiva regions of the eye generally plays an important role in the diagnosis of eye diseases and applications related to eye health. The conjunctiva is a thin membrane tissue that covers the inner surface of the eyelids and the white part of the eye. Detection and analysis of this region is used in the examination of inflammation, redness, dryness and other disorders in the eye. The relevant regions were found using conjunctiva images in the study. Conjunctiva region detection Images were taken from a public database and enhanced with the image enhancement method DNCNN. The YOLO algorithm is applied to raw images and DNCNN enhanced images separately using the same parameters. As a result, the effect of the deep learning based method on finding the truth in images is presented with F1-confidence curve, precision-confidence curve, recall-confidence curve, precision-recall curve and confusion matrix metrics. In the proposed method, the mAP value is given as 0.984 in all classes.

1. Introduction

The eyes are among the majority of vital systems in the body. Eyes are windows that connect us to the world. All our lives; genetics, age, existing diseases and environmental factors can threaten our eye health. There are multiple types of vision issues. Since the visual system is an important organ for humans, external eye abnormalities need to be detected early. The transparent membrane that encompasses the interior area of the eyelids and the white part of the eye is called the conjunctiva [1]. Human optics are safe from dust particles thanks to the conjunctiva. It acts as a lubricant and prevents friction when opening and closing the eye.

Three anatomical sections of the mucosa that make up the conjunctiva are (1) palpebral, (2) fornix, and (3) bulbar. Although it is more tightly fastened to the higher tarsus, the palpebral conjunctiva is bonded to both the lower and upper tarsi. Each eyelid's posterior third of the meibomian gland apertures is where the palpebral conjunctiva starts. Cells change at this mucocutaneous junction from the skin's keratinized stratified squamous epithelium to the marginal conjunctiva's nonkeratinized stratified squamous epithelium. The conjunctiva and lacrimal puncta join at this mucocutaneous junction as it comes to the nose [2]. The conjunctiva is highly vascularized and many microvessels are easily accessible for imaging studies. Figure 1 shows the anatomical structure of the eye conjunctiva.

*Corresponding author: ecengil@beu.edu.tr
Received: 27.08.2024, Accepted: 29.12.2024



Colored slit-lamp photograph of right lower eyelid. (*1*) Mucocutaneous junction, (*2*) lacrimal punctum, (*3*) subtarsal groove, (*4*) tarsal conjunctiva, (*5*) forniceal conjunctiva, (*6*) bulbar conjunctiva, and (*7*) contact lens on limbus

Figure 1. Eye conjunctiva anatomy [2]

Ideal blood hemoglobin levels are associated with health issues and can be used as indicators of several disorders. A blood sample is typically used in an invasive manner utilizing various equipment to assess hemoglobin levels. Certain indications have historically been interpreted physically. Pallor of the face, nail beds, conjunctiva, and palm wrinkles are among these signs. Conjunctival pallor may be a more sensitive indicator of anemia than palm or nail bed pallor, according to studies [3]. In addition, diabetes can be detected using conjunctival images.

Numerous research is available in the literature using artificial intelligence techniques in the fields of classification, segmentation and detection of eye disorders. There are also studies conducted to detect diabetes and anemia using eye conjunctiva images.

In their study focusing on the problem of automatic classification of eye disorders, A. K. Bitto and I. Mahmud [4] distinguished between normal eyes, eyes with conjunctivitis, and eyes with cataracts using the convolutional neural network architectures of VGG-16, ResNet-50, and Inception-v3. In the dataset they used, Inception-v3 had the highest accuracy rate in detecting eye diseases with 97.08% verification accuracy, while ResNet-50 achieved the second highest accuracy with 95.68% and finally VGG-16 achieved 95.48%.

S. Dhalla et al. [5] used eye conjunctivitis images to detect anemia through computerized analysis. The paper examined the simultaneous picture segmentation five deep learning-based models' performance: UNet, UNet++, FCN, PSPNet, and LinkNet. A specially created dataset of 2592 palpebral pictures of pediatric nuclei is used for the experiments. LinkNet delivered the best results. For intersection-union (IoU) and dice score performance measurements, accuracy was 94.17%, 90.14%, and 93.78% in the pertinent dataset, respectively.

The goal of E. Purwanti et al. [6] is to use a deep learning method to classify images of the palpebral conjunctiva in order to diagnose anemia. There are three CNN designs in use: ResNet-50, MobileNetV2, and AlexNet. The accuracy of the AlexNet, ResNet-50, and MobileNetV2 designs was 97.19%, 97.94%, and 89.93%, respectively, according to the results.

X. Li et al. [7] aimed to automatically detect diabetes using conjunctival images. For this purpose, researchers designed a learning model. Both healthy people and individuals with type 2 diabetes provided images for the collection. Using conjunctival pictures, a hierarchical multitask network model (HMT-Net) was created. The model underwent thorough evaluation and comparison with alternative approaches. The proposed model achieved 75.17% accuracy.

A non-invasive optical system is suggested as the mechanism for automatic diabetes detection by E. R. Ghugare et al. [8]. The anterior conjunctiva and eye pallor are examined as part of the research in order to identify diabetes. The eyes of both non-diabetic and diabetic participants were analyzed using medical imaging. For the pertinent job, a CNN-based prediction model was developed. Ultimately, the method provided 96% accuracy.

Pallavi et al. [9] created a bot that is powered by artificial intelligence. Two models serve as the foundation for the bot service: one for segmenting the Region of Interest and the other for classifying anemic cases from normal ones. For the purpose of training the model, data were gathered from 160 anemics and 140 non-anemics people. For the segmentation method, we were able to obtain an Intercection Over Union (IOU) score of 0.922; for the classification model, we were able to obtain a validation accuracy of 0.9699 and a validation recall of 0.95.

The process and technology for obtaining significant information from a digital conjunctival image were detailed by G. Dimauro et al. [10]. Here, efficient characteristics are employed to guarantee that every picture is part of a diagnostic probability class for anemia. Conjunctiva photos were taken with a novel, user-friendly, low-cost apparatus intended to maximize ambient light independence. The palpebral conjunctiva was manually extracted from photos to assess the system's performance, or it could be done semi-automatically using the SLIC Superpixel technique. Tests were run on pictures that came from 102 individuals. A few classification methods for determining anemic status were tried, and SMOTE and ROSE procedures were assessed to balance the data set.

S. Wei et al. [11] aims to evaluate the applicability of the deep learning method in determining the rating of bulbar conjunctival injection. 1401 color anterior segment pictures that displayed the bulbar conjunctiva and cornea were

gathered. The bulbar conjunctival injection scores, as recorded by human ophthalmologists, served as the ground truth. Two models based on convolutional neural networks were constructed and trained. Deep learning models have been assessed for efficiency using performance criteria. Accuracy of 87.12% were attained by the deep learning model.

Using smartphones and artificial intelligence approaches, S. H. Elgohary et al. [12] sought to create a remote, non-invasive, standardized solution that allows a fast scan to find hemoglobin levels. The conjunctiva is automatically generated as a Region of Interest from an image of the eye during the process. After that, characteristics are taken out of it in order to teach a machine learning system whether or not the patient is anemic. 200 participants participated in the study, which had 85% accuracy, 86% precision, and 81% recall rate.

P. Appiahene et al. [13] concentrated on pallor analysis and employed machine-learning algorithms to identify anemia using pictures of the eye's conjunctiva. The research employed a publicly accessible dataset of 710 pictures of the eye's conjunctiva. With the use of Gaussian Blur, Logistic Regression, and Convolutional Neural Networks, two methods for the detection of conjunctiva and anemia were created. These models were then deployed on a Fast API server that was coupled to a front-end React Native mobile application. The created model possessed 90% sensitivity and 95% specificity. It is integrated into a mobile application that can identify anemia with 92.50 accuracy by obtaining and analyzing the patient's conjunctiva.

Today, artificial intelligence technologies are developing rapidly in the field of health, as in every field. The eye is an organ where not only eye diseases but also diseases such as anemia and diabetes can be detected. Invasive methods such as blood sampling are used to diagnose such disorders. Instead of these, it is a more comfortable method to detect diseases using eye images. In the study, three different regions are detected using eye conjunctiva images. According to our research, no academic study has been found on the detection of regions in conjunctiva images. In cases where the number of experts is insufficient, automatic discrimination of the conjunctiva regions of the eye is beneficial.

The rest of the research is arranged as follows. In the second section, the technical background is given, including the data set used, the enhancement technique applied to the images, and the method applied to find the regions. In the third section, the suggested technique is presented. In the fourth section, the experimental results are given. Lastly, the fifth section gives the discussion and conclusions of the study.

2. Technical Background

Technical background information is important for understanding the proposed methodology. This section indicates the dataset used in the training and validation processes. The data is developed by pre-processing before being given as input to the method. YOLOv8 (You Only Look One-v8) is used to detect the regions

2.1. Dataset

The data set used was taken from the Roboflow site. The dataset named Eye conjunctiva contains 218 eye images. Of these, 152 are reserved for training, 32 for verification, and the remaining 34 for testing. The data is labeled in three categories. These; palpebral, forniceal and forniceal palpebral. Some example images of the dataset are displayed in the Figure 2 [14].

2.2. Improving the Image: DNCNN

In order to restore the original image by suppressing related noise, image denoising is an essential pre-processing step. Since noise depends on the high frequency composition of the image, denoising is a complicated process [15]. The primary goal is to strike a balance between minimizing information loss and lowering noise as much as feasible. Image noise can be eliminated by applying filter-based techniques like wavelet, mean, and median. The increased hardware capabilities of computers have led to a rise in the adoption of modern techniques

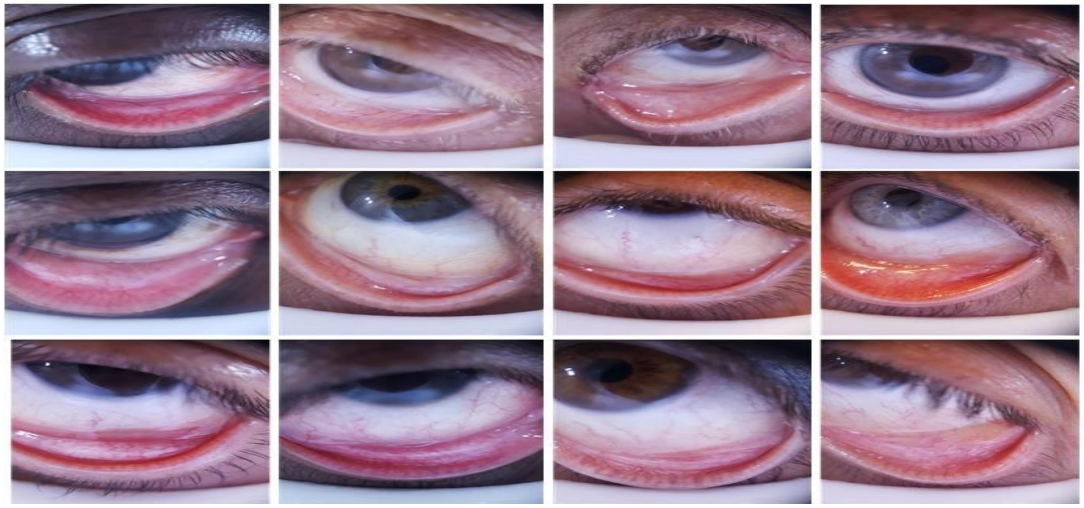


Figure 2. Some images from eye conjunctiva dataset [14]

The Deep CNN Residual Learning (DNCNN) [16] method is applied in this work to eliminate noise. Several attempts have been made by deep neural networks to address the noise removal issue. Thanks to advancements in deep learning techniques and the accessibility of access to large-scale datasets, CNN has recently shown exceptional efficacy in completing a range of vision tasks. In addition to solving challenges with image denoising, JPEG deblocking, and super-resolution, DNCNN, a model designed to handle a variety of low-level assignments, can also carry out blind reconstruction in the absence of any knowledge about the input image. Figure 3 displays the DNCNN network's design.

The DNCNN architecture consists of three different kinds of layers. (1) Convolution + ReLU: In the first layer, 64 filters with a size of 3 x 3 x x are utilized to create 64 feature maps. These maps are subsequently adjusted for nonlinearity using linear units (ReLU, max (0, ·)). In this case, x stands for the number of image channels; that is, x = 3 for a color image. (2) Conv+BN+ReLU: 64 filters with a size of $3 \times 3 \times 64$ are employed for $2 \sim (D - 1)$ layers, and batch normalization is introduced in between the convolution and ReLU. (3) Conv: To rebuild the output for the final layer, c filters with a size of $3 \times 3 \times 64$ are employed. Through hidden layers, DNCNN can gradually extract the image framework from noise observation by combining convolution with ReLU. The DNCNN model employs the notion of residual learning. DNCNN predicts the residual image using just one residual unit, as opposed to the residual network's many residual units. Additionally, faster and more reliable training as well as better denoising performance can be achieved by combining batch normalization and residual learning with DNCNN [16].

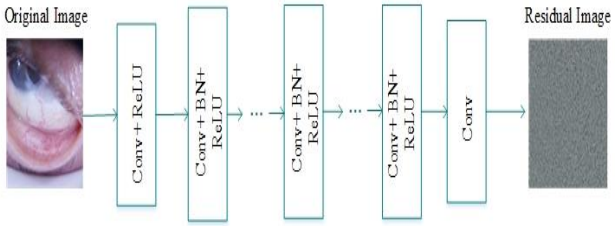


Figure 3. The structure of DNCNN

2.3. Multiple Object Detection: YOLOv8

The YOLOv8 algorithm's structure is first described in this section, which serves as a foundation for algorithm construction. The four components of the YOLOv8 algorithm are the entrance, spine, neck, and head. Figure 4 depicts the YOLOv8 algorithm's structure.

Backbone: In the YOLOv8 algorithm, the backbone network's job is to extract the target's general features. Three modules make up this backbone network: Conv, C2f, and SPPF. The Conv module uses autopad (k,p) to generate fill effects and is composed of three functional modules: Conv2d, BN, and SiLU activation function. The C2f module's design, which draws inspiration from ELAN and the C3 module, enables YOLOv8 to offer greater features while offering lightweight features that enable the acquisition of gradient flow information. SPPNet served as an inspiration for the redesign of the SPPF module. The SPP module gradually adopts several small-size pooling cores in place of a single large-size pooling core, maintaining its original functionality [17], 18].

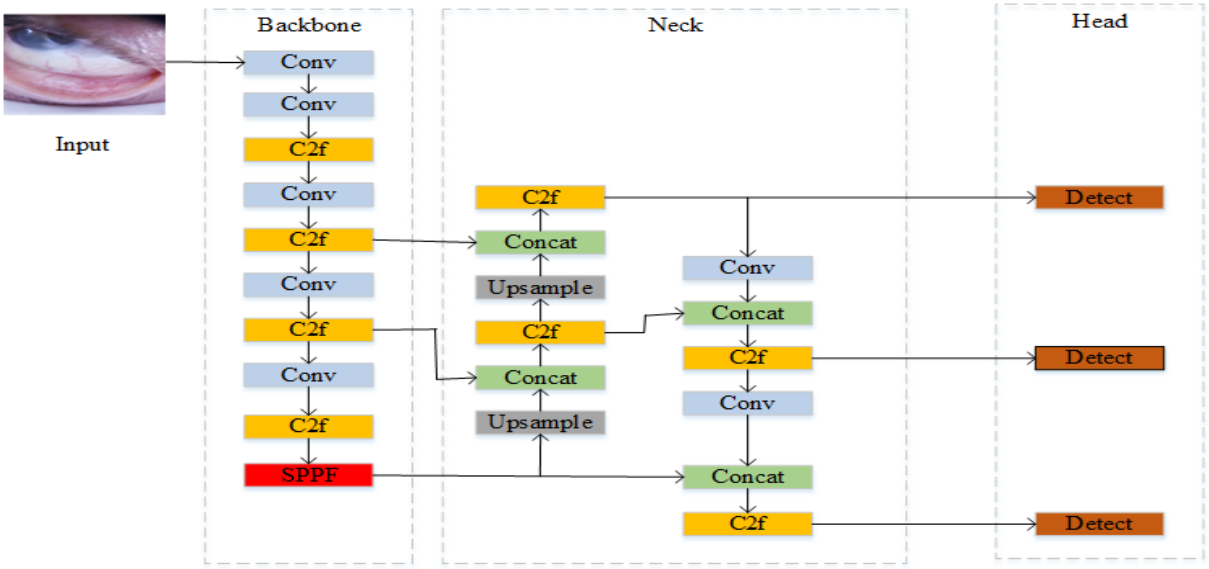


Figure 4. The architecture of YOLOv8

Neck: PAN-FPN module is used in the neck of YOLOv8 for efficient feature fusion at various scales. With the use of FPN and PAN architectures, this module employs a multi-scale fusion technique in which lower layers maintain precise position details while top layers acquire more information [19].

Header: The head of the YOLOv8 algorithm uses the Split Head structure, which is a common parser head structure. This structure uses the concept of DFL (Distributional Focal Loss) to distinguish between headings for classification and detection. Targets are predicted and regressed using the three detecting layers in the head of the YOLOv8, which correspond to three distinct kinds of anchors with various aspect proportions extracted from the Neck. The YOLOv8 algorithm's junction boxes are adaptable to the dataset and may automatically modify their corresponding junction boxes based on various datasets.

3. Proposed Method

The study aims to detect the palpebral, forniceal palpebral and forniceal regions in eye conjunctiva images. The proposed method is carried out in two stages. In the first stage, the DNCNN algorithm was applied to the images to be used to perform the method and the images were enhanced. The noisy signal is the DNCNN's input. DNCNN maps the remaining learning to train and utilize it to forecast the latent clean signal, in contrast to discriminative denoising, which aims to learn to map the noisy signal to the genuine signal [20].

In the second step of the study, the improved images were given to the YOLOv8 algorithm and the regions were detected. Simple block drawing of the proposed pipeline is as in Figure 5.

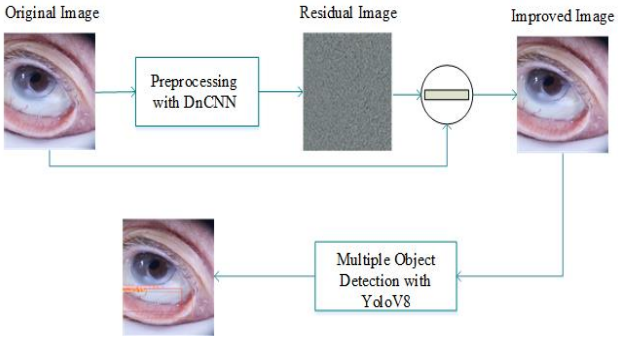


Figure 5. Simple block drawing of proposed pipeline

4. Experimental Results

In this article, it is aimed to detect the regions using eye conjunctiva images. First of all, the images to be used with deep learning technology are pre-processed and developed. Then, the network first produces images in the data set. The system is then tested with images to distinguish palpebral, forniceal palpebral, and forniceal regions. The application was run on a computer with an i7 processor, GPU card and 16GB RAM.

4. 1. Object Detection Metrics

Many metrics are utilized to test the effectiveness of algorithms for object detection. Frequently used metrics are Intersection over Union, Precision, Recall, Average Precision, Mean Average Precision (mAP) and F-score [21].

Intersection over Union (IoU): The relationship between a predicted and the ground truth bounding boxes is quantified using a metric called IoU. It is essential to assessing how accurate object localization is.

Average Precision (AP): The accuracy and recall value of the method are combined into a single value by AP, which computes the area under the precision-recall curve.

Mean Average Precision (mAP): The AP notion is expanded upon by mAP, which computes the average AP values among various object classes. This aids in providing a comprehensive evaluation of the model's effectiveness in situations involving multiple classes of item detection.

Precision: Tells how precise our model is. In other words, it tells you how many of the total palpebral regions detected are actually palpebral. It is the proportion of the overall number of palpable predictions the algorithm makes to the genuine positive, as shown in equation (1).

$$Precision = \frac{Correct\ Predictions}{Total\ Predictions} = \frac{TP}{(TP+FP)} \quad (1)$$

Recall: It tells us how good the model is at remembering classes from images, that is, how many of the total palpebral in the input image the model can detect. It is the proportion between the true positive generated by the model and the total of the true positive and false negative as shown in Equation (2).

$$Recall = \frac{Correct\ Predictions}{Total\ GroundTruth} = \frac{TP}{(TP+FN)} \quad (2)$$

F1 Score: The F1 Score is a comprehensive evaluation of a method's performance that considers both false positives and false negatives. It is calculated as the harmonic mean of precision and recall. The F1 Score equation is provided by equation (3).

$$F1\ Score = 2 * \frac{(Precision*Recall)}{(Precision+Recall)} \quad (3)$$

The F1 score provides a balance between precision and recall. This balance is very important, especially in data sets with class imbalance. For example, if a model provides very high precision but ignores sensitivity, this means that it misses true positives.

4.2. Results Obtained using Original Images

In the first part of the study, regions were detected with YOLOv8 using eye conjunctiva images. Yolov8m weight was used when training the model. Epoch is taken as 100 and batch 4. The confusion matrix of the model is as presented in Figure 6.a. The number of locations in the confusion matrix, sometimes referred to as the error matrix, where the true label and the predicted label coincide is represented by diagonal elements, while the points that the classifier incorrectly identified are represented by off-diagonal elements. Indicative of many accurate predictions, the confusion matrix's diagonal values increase with increasing complexity. Figure 6.b gives the normalized confusion matrix of the method.

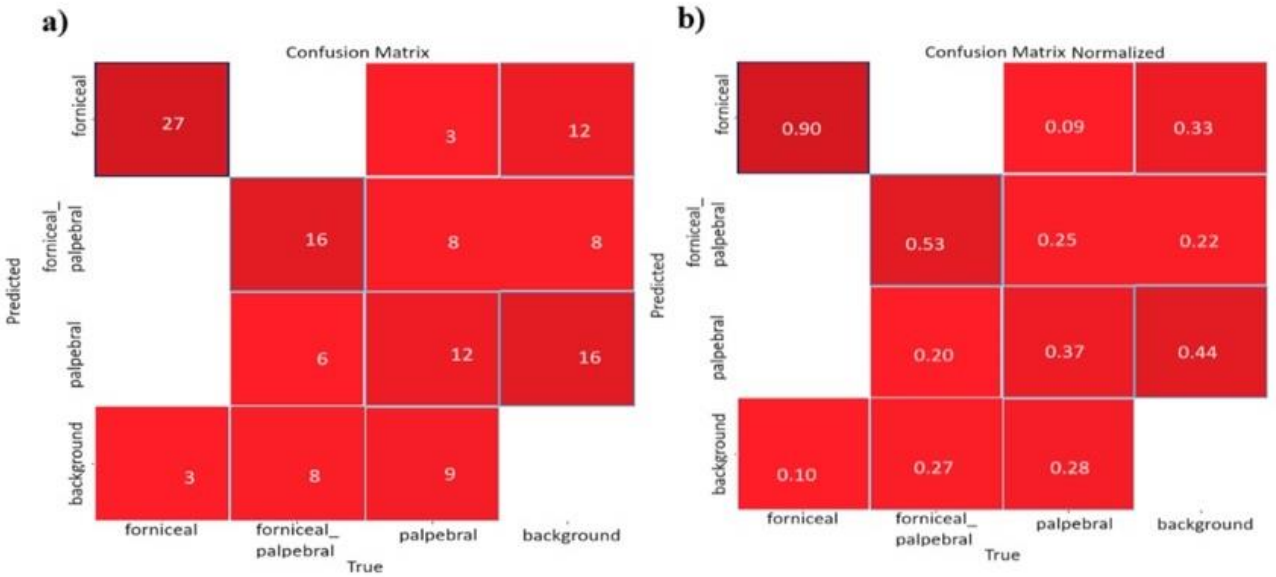


Figure 6. a) Confusion Matrix and, **b)** Normalized Confusion Matrix

The F1-Confidence curve and P curve of the method are given in Figure 6.a, and Figure 6.b, separately. When looking at the F1-Confidence curve, a value of 0.97 was reached for all classes. A 73.4% confidence interval is provided in the P-curve, which expresses the likelihood of acquiring test outcomes that are at least as extreme as the actually acquired results.

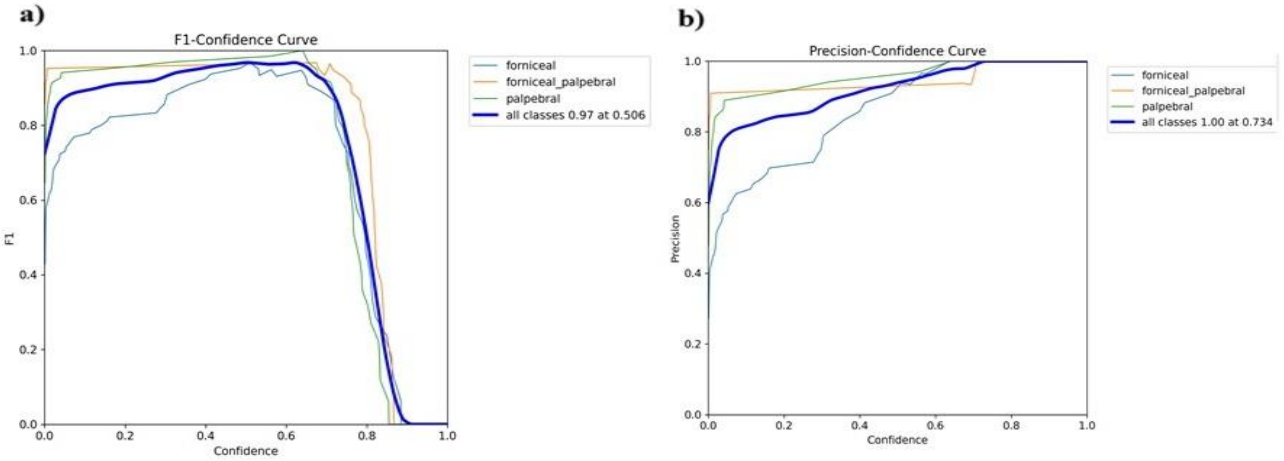


Figure 7. a) F1-Confidence Curve and, **b)** Precision-Confidence Curve

Figure 8.a, and Figure 8.b display the model's Precision-Recall and Recall-Confidence curves, respectively. The trade-off between recall and precision for different threshold levels is displayed on the precision-recall curve. High precision and recall are shown by a high area under the curve. Here, forniceal, forniceal palpebral, and palpebral values were found to be 0.990, 0.991, and 0.995, respectively. Finding the area under the precision-recall curve is the general definition of average precision, or AP. The mean of AP is called mAP. In certain situations, AP is determined for every class and averaged to provide mAP. They mean the same thing in other contexts, though. mAP establishes a balance between the precision (sensitivity) and recall of the results that the model correctly detects in different equivalent values. It is expressed that the model is not only examined by making correct detections, but also by keeping false positives low. mAP also evaluates the accuracy of the confidence values that the model assigns to its predictions. Incorrectly high confidence scores can lead to false positives and cause serious problems in practice. Recall is the quantity of accurate positive predictions generated relative to all possible positive predictions. Recall gives an indicator of missed positive predictions, in contrast to precision, which only comments on correct positive predictions out of all positive predictions. The mAP value for all classes is 0.992. When we look at the Recall-Confidence curve, it is seen as 1 for all classes.

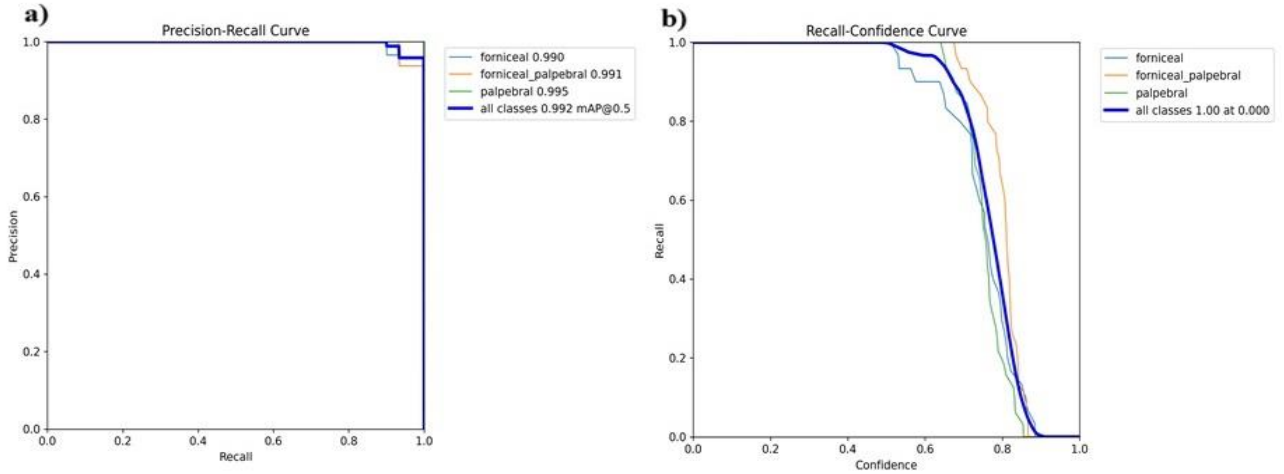


Figure 8. a) Precision-Recall Curve and **b)** Recall-Confidence Curve

Training results of the method are presented in Figure 9. 34 images were used for testing. Some of the results obtained from these are as given in Figure 10.

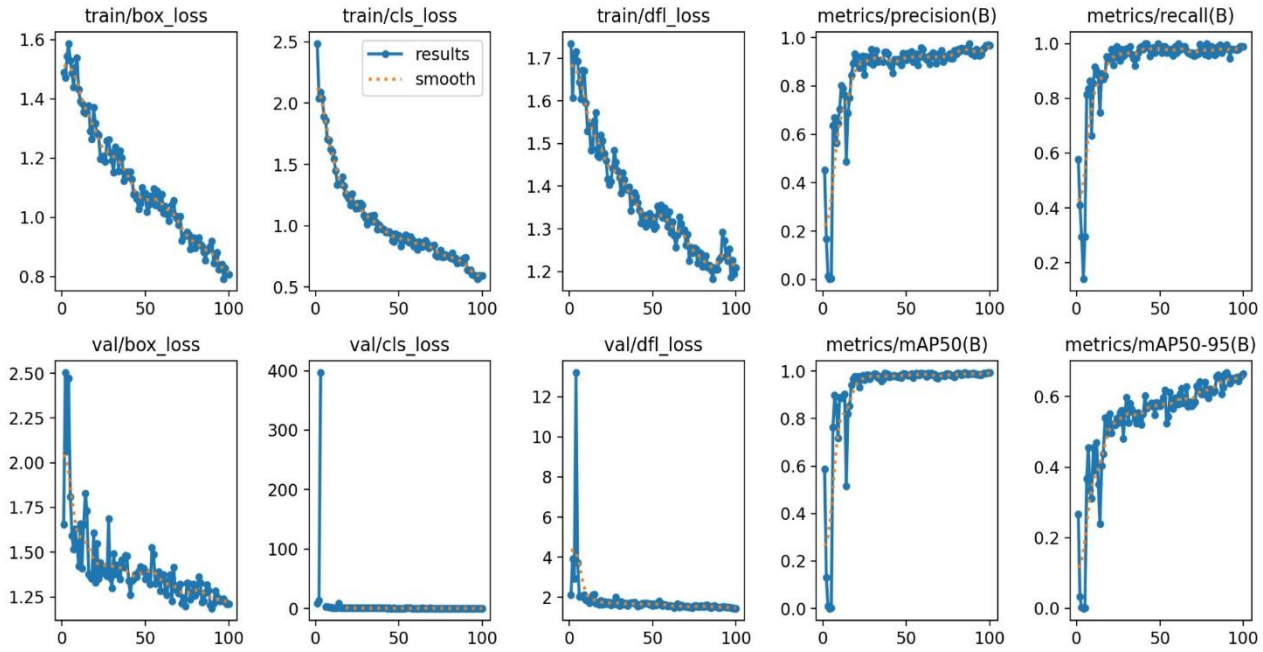


Figure 9. Training Results of the YoloV8m Model on the Eye Conjunctiva Dataset

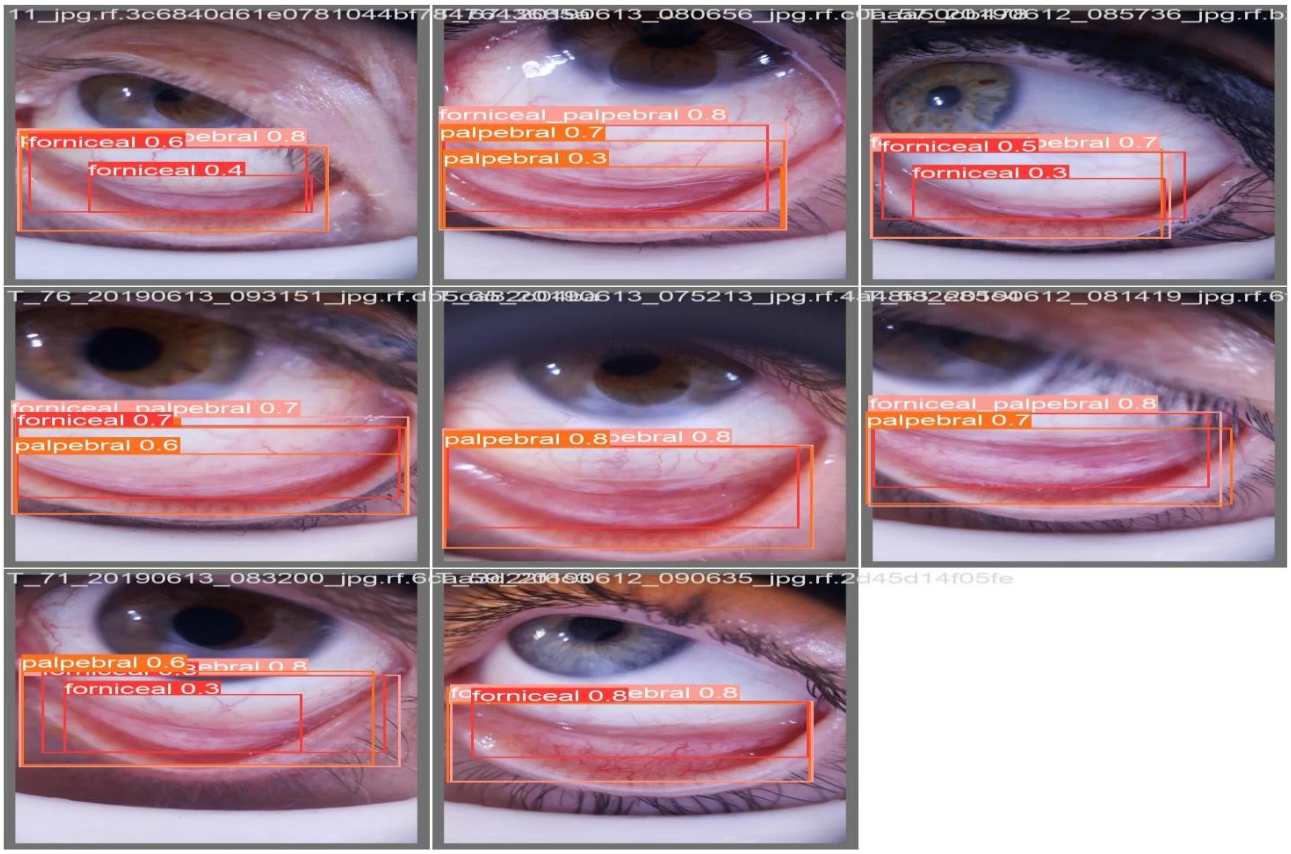


Figure 10. Test results of the YoloV8m Model on the Eye Conjunctiva Dataset

4.3. Results Obtained using Images Enhanced using DNCNN

In the study, it was observed that the proposed image enhancement method positively affected the performance results. Figure 11.a gives the relevant confusion matrix. Figure 11.b presents the normalized confusion matrix.

F1-Confidence and Precision-Confidence curves are given in Figure 12.a, and 12.b, respectively. When we look at the curves, the precision-confidence value was 0.734 in all classes in the method performed with raw data, and 0.772 in all classes when the data obtained with DNCNN was used.

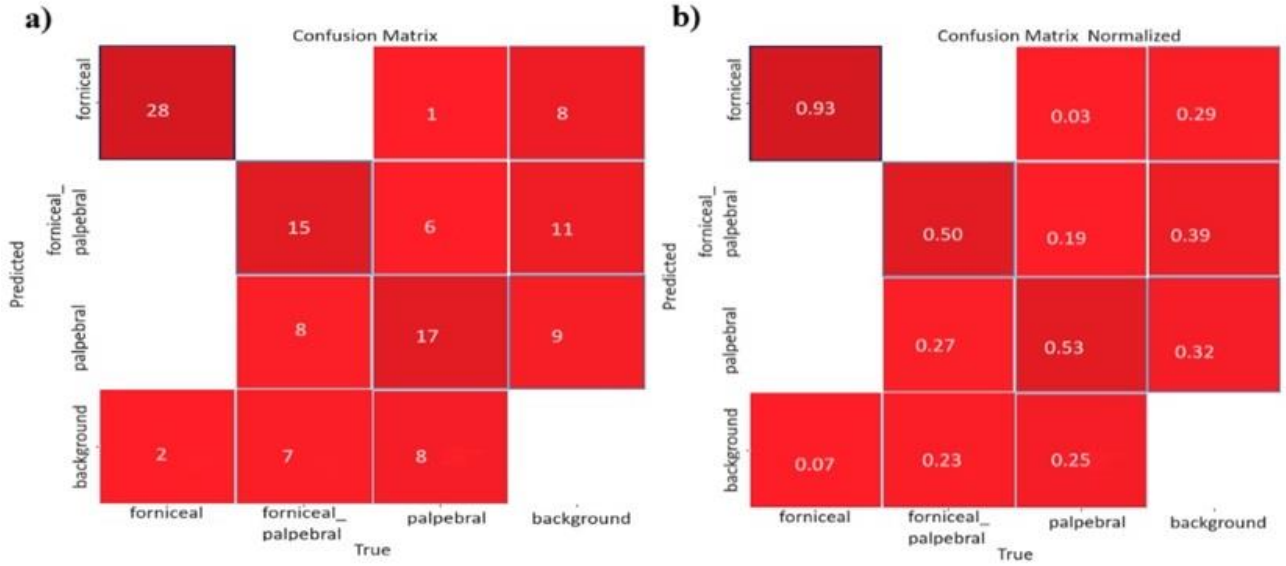


Figure 11. a) Confusion Matrix and, b) Normalized Confusion Matrix

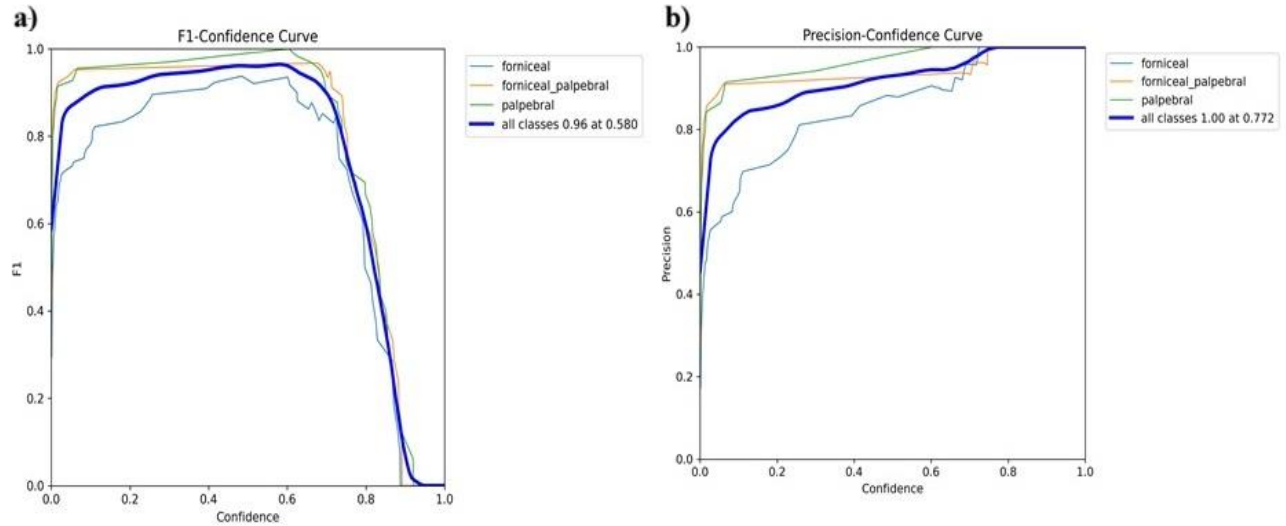


Figure 12. a) F1-Confidence Curve and, **b)** Precision-Confidence Curve

Figure 13.a gives the Precision-Recall curve. Forniceal 0.973, forniceal_palpebral 0.983, palpebral 0.995 and mAP value in all classes were obtained as 0.984. The Recall-Confidence curve is given in Figure 13.b. The training results of the model trained with DNCNN-improved images are presented in Figure 14. The results of some of the test images are presented in Figure 15.

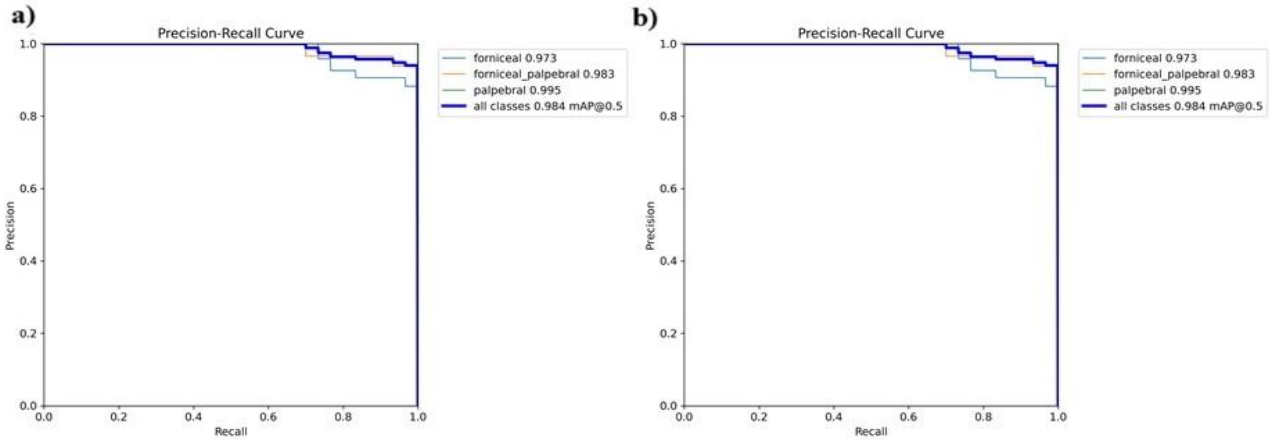


Figure 13. a) Precision-Recall Curve and **b)** Recall-Confidence Curve

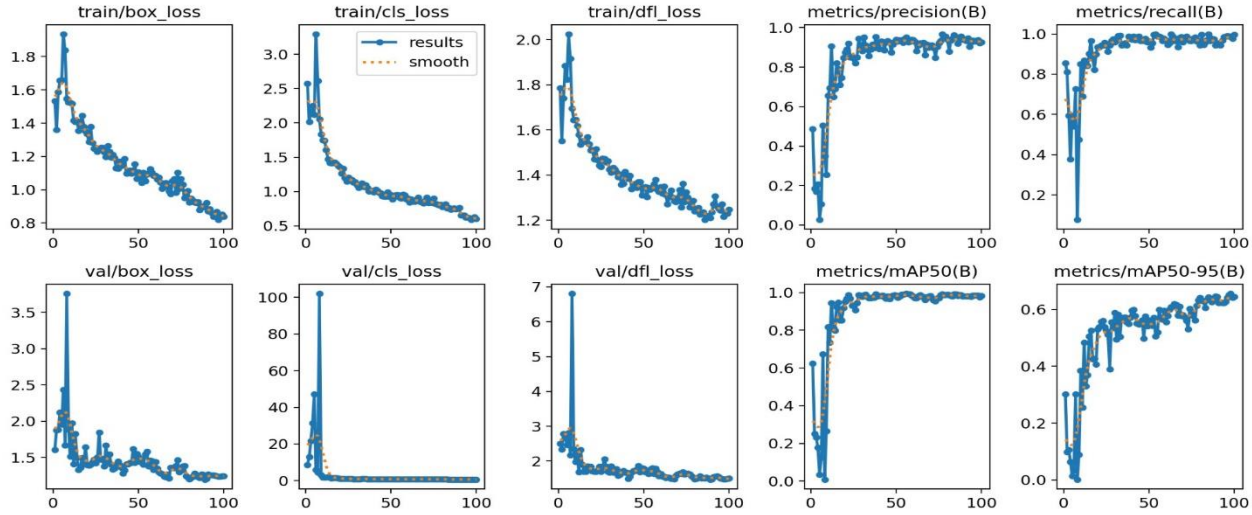


Figure 14. Training Results of YoloV8m Model on DNCNN-improved Eye Conjunctiva Dataset

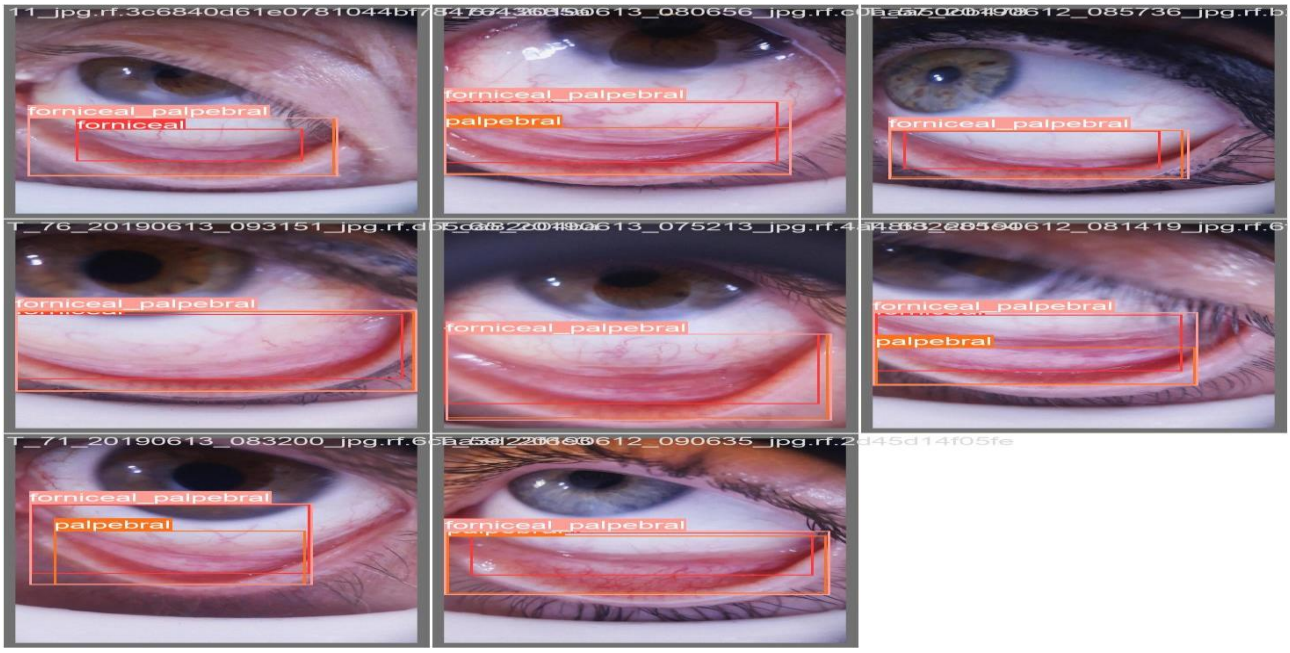


Figure 15. Test results of the YoloV8m Model on enhanced images

Table 1. Studies related using eye conjunctiva images

Authors/Year	Task	Dataset	Model	Results
S. Dhalla et al.[5]/2023	Conjunctiva Segmantation	2592 images	LinkNet	94.17 % Acc.
E. Purwanti et al. [6]/2023	Anemia Classification	654 images	ResNet-50	97.94 % Acc.
X. Li et al. [7]]/2022	Diabet Classification	611 images	HMT-Net	75.17 % Acc.
E. R. Ghugare et al. [8]/ 2023	Diabet Classification	611 images	CNN	96 % Acc.
B. B. Pallavi et al. [9]/2023	Anemia Classification	300 images	CNN	97 % Acc.
B. B. Pallavi et al. [9]/2023	Anemia Segmantation	300 images	CNN	92.2 % IOU
G. Dimauro et al. [10]/2017	Anemia Classification	102 images	SMOTE+KNN	98% Acc.
S. Wei et al. [11]/2023	Bulbar Conjunctival Injection Rating	1401 images	CNN	87.12 % Acc.
S. H. Elgohary et al. [12]/2022	Anemia Classification	318 images	KNN	84.6 % Acc.
P. Appiahene et al. [13]/2023	Anemia Classification	710 images	CNN+Logistic Regression+ Gaussian Blur	92.50 % Acc.
This paper	Multi-region Detection of Eye Conjunctiva	218 images	Yolov8	0.992 mAP 0.734 Precision Confidence 0.97 F1 Confidence
This paper	Multi-region Detection of Eye Conjunctiva	218 images	DNCNN+Yolov8	0.984 mAP 0.772 Precision Confidence 0.96 F1 Confidence

5. Discussion

In the study, the locations of the conjunctiva images, including palpebral, forniceal-palpebral and forniceal, were determined. Palpebral regions of the eye are used to detect disease. By taking a picture of the palpebral conjunctiva, doctors can determine if a blood sample is required or if the patient even needs to notify them. This helps to narrow down the pool of potential blood sample donors. Additionally, it can draw attention to possible anemia and make it possible for many people to be screened for anemia, especially in environments with low resources.

A thorough understanding of the architecture of the conjunctiva, particularly the eyelid, is necessary to build a treatment that bolsters the anemia diagnosis. The conjunctiva is a mucous membrane that surrounds the arches but does not touch the cornea. It stretches from the inner palpebral borders to the eyeball. The plentiful abundance of microvessels ensures a high degree of vascularization, which generates the palpebral conjunctiva a great nominee for physical examination observation. Compared to the bulbar conjunctiva, the palpebral conjunctiva more clearly displays the vascularization of the underlying region, highlighting even the smallest variations in blood color.

Many studies have been conducted using images of the eye conjunctiva. Table 1 lists some studies in the literature using eye conjunctiva images. Looking at the table, it can be seen that these images can be used to classify diseases such as anemia and

diabetes. When the accuracy rates are examined, it is seen that the methods need to be improved. According to our research, no academic study has been found on the detection of regions in conjunctiva images. Therefore, it was not included in the comparison table. In addition, the results of the realization of raw images with the YOLOv8 algorithm are given. Looking at the table, we see the effect of pre-processing images with the deep learning method on performance.

6. Conclusion

Using eye conjunctiva images, diseases such as anemia and diabetes can be detected. Automatic disease detection with artificial intelligence is frequently preferred in the health field in terms of patient comfort and the convenience it provides to the specialist. The proposed method enables the detection of regions in eye conjunctiva images with acceptable high accuracy rates. The use of the method in detecting regions in eye conjunctiva images will provide convenience to specialists. Nonetheless, it's critical that the AI techniques applied in the medical industry deliver excellent results. One stage that will be used to identify the disorders is the identification of the conjunctiva regions. As a result, the approach needs to be improved going forward

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] C. L. Shumway, M. Motlagh, and M. Wade, "Anatomy, head and neck, eye conjunctiva," 2018.

[2] R. Ram and M. B. Goren, "Palpebral Conjunctiva," in *Encyclopedia of Ophthalmology*, U. Schmidt-Erfurth and T. Kohnen, Eds., Berlin, Heidelberg: Springer, 2014. doi: 10.1007/978-3-642-35951-4_845-1.

[3] B. Çuvadar and H. Yılmaz, "Non-invasive hemoglobin estimation from conjunctival images using deep learning," *Medical Engineering & Physics*, vol. 120, p. 104038, 2023.

[4] A. K. Bitto and I. Mahmud, "Multi categorical of common eye disease detect using convolutional neural network: A transfer learning approach," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 4, pp. 2378–2387, 2022.

[5] S. Dhalla et al., "Semantic segmentation of palpebral conjunctiva using predefined deep neural architectures for anemia detection," *Procedia Computer Science*, vol. 218, pp. 328–337, 2023.

[6] E. Purwanti, H. Amelia, M. A. Bustomi, M. A. Yatijan, and R. N. Putri, "Anemia Detection Using Convolutional Neural Network Based on Palpebral Conjunctiva Images," in *2023 14th International Conference on Information & Communication Technology and System (ICTS)*, pp. 117–122, IEEE, 2023.

[7] X. Li et al., "Identifying diabetes from conjunctival images using a novel hierarchical multi-task network," *Scientific Reports*, vol. 12, no. 1, p. 264, 2022.

[8] E. R. Ghugare, R. Patil, and Z. B. Khan, "CNN based Non-Invasive Technique for Diabetic Detection from Conjunctival Image," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–5, IEEE, 2023.

[9] B. Pallavi et al., "A Deep Learning-based System for Detecting Anemia from Eye Conjunctiva Images Taken from a Smartphone," *IETE Technical Review*, pp. 1–13, 2023.

[10] G. Dimauro et al., "Detecting clinical signs of anaemia from digital images of the palpebral conjunctiva," *IEEE Access*, vol. 7, pp. 113488–113498, 2019.

[11] S. Wei et al., "Developing a Deep Learning Model to Evaluate Bulbar Conjunctival Injection with Color Anterior Segment Photographs," *Journal of Clinical Medicine*, vol. 12, no. 2, p. 715, 2023.

[12] S. H. Elgohary et al., "A Machine Learning Method to Screen Anemia From Conjunctiva Images Taken by Smartphone," in *2022 10th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC)*, pp. 125–128, IEEE, 2022.

[13] P. Appiahene et al., "Detection of anemia using conjunctiva images: A smartphone application approach," *Medicine in Novel Technology and Devices*, vol. 18, p. 100237, 2023.

[14] Eye conjunctiva detector, [Online]. Available: https://universe.roboflow.com/eyeconjunctivadetector/eye-conjunctiva-detector. [Accessed: Aug. 2, 2024].

[15] M. A. Günen and E. BEŞDOK, "Effect of denoising methods for hyperspectral images classification: DnCNN, NGM, CSF, BM3D and Wiener," *Mersin Photogrammetry Journal*, vol. 5, no. 1, pp. 1–9, 2022.

[16] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[17] M. Ma and H. Pang, "SP-YOLOv8s: An improved YOLOv8s model for remote sensing image tiny object detection," *Applied Sciences*, vol. 13, no. 14, p. 8161, 2023.

[18] C. Wang et al., "Strawberry Detection and Ripeness Classification Using YOLOv8+ Model and Image Processing Method," *Agriculture*, vol. 14, no. 5, p. 751, 2024.

[19] F. Solimani et al., "Optimizing tomato plant phenotyping detection: Boosting YOLOv8 architecture to tackle data complexity," *Computers and Electronics in Agriculture*, vol. 218, p. 108728, 2024.

[20] A. Yapıcı and M. A. Akcayol, "Derin Öğrenme ile Görüntülerde Gürültü Azaltma Üzerine Kapsamlı Bir İnceleme," *International Journal of Advances in Engineering and Pure Sciences*, vol. 34, no. 1, pp. 65–90, 2022.

[21] R. Padilla, S. L. Netto, and E. A. Da Silva, "A survey on performance metrics for object-detection algorithms," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 237–242, IEEE, 2020.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1194-1204 YEAR: 2024
DOI: 10.17798/bitlisfen.1540373

DentaGAN: GAN-Based Synthetic Individual Dental Data Generation in Radiographic Images

Buse Yaren KAZANGİRLER[1*], Caner ÖZCAN[2]

[1] *Karabuk University, Department of Computer Engineering, Karabük, Türkiye*
[2] *Karabuk University, Department of Software Engineering, Karabük, Türkiye*
(ORCID: 0000-0002-8690-2042) (ORCID: 0000-0002-2854-4005)

Keywords: Generative adversarial networks, Dental image generation, Synthetic data augmentation, Two-stage neural network, Panoramic images

Abstract

Panoramic radiographs are a low radiation exposure type often used as a data source for many deep learning algorithms. On the other hand, the operational structure of a traditional deep learning algorithm requires a large amount of data, which is a major problem for many researchers. It is aimed to overcome this problem through deep GAN models, many versions of which have been developed recently. The main purpose of the study is to generate a two-stage GAN model for data with the same image dimensions. The study is carried out in the form of inputting panoramic images containing a whole view, as well as single tooth data whose performance is desired to be measured, to the architecture. The generator model created for each tooth object in all panoramic radiographs generates new tooth objects that the model has yet to encounter in the dataset. Fréchet Inception Distance was used as a performance metric by measuring the distance for the Inception-v3 activation distributions for the real samples in the generated and training set. Thus, the statistical similarity of these two groups obtained from the experimental results was observed in the part of the experimental results. The cropped individual tooth classes were much more successful than the entire panoramic dataset.

1. Introduction

Dental radiographs, widely used in medical imaging, are continuously used to detect tooth loss, tooth material loss, and many health problems that are not detected by visual examination. Various digital radiographs and scan results containing data are essential in medicine, especially in forensic dentistry [1]. In addition, digital imaging has become the most widely used imaging technique in dentistry, the most common field of digital radiography [2]. Although intraoral and extraoral radiographic techniques have advantages or disadvantages relative to each other, both types of imaging contribute negatively to the health of the patient. Radiographs have extremely impressive impacts on radiation, whether a device placed in the mouth or imaging the whole oral region. Therefore, apart from using radiographs to treat the patient during the examination, there should be a limited acquisition process for the datasets intended to contribute to the academic literature [1]. The use of various Artificial Learning (AL) techniques is more appropriate when sufficient data cannot be obtained. If the amount of data in the studies mentioned above is small, it would only be appropriate for the individuals who will use the data to reproduce it manually with an expert. It is also possible to misinterpret the types of radiographs taken in clinics and healthcare institutions depending on the quality of the device and the need for more expert interpretation [2]. It is obvious that the learning-based world goes beyond manual programming and produces solutions to these problems [3]. For this reason, the age of Artificial Intelligence (AI) has rapidly developed in medicine, health, science, and many other important fields [4].

Deep Learning (DL), which has made a name for itself as a sub-branch of AI, makes things much

*Corresponding author: tekinbuseyaren@gmail.com
Received: 29.08.2024, Accepted: 06.12.2024

easier compared to classical machine learning techniques. In the last decade, it has been observed that the frequency of use of DL models has increased significantly, especially in the field of health and medicine, due to their use in the detection of important diseases such as cancer, tumor cells, and pneumonia [5]. The function of using AI in dentistry has been inevitable due to the reasons explained and the development of the technology. In the literature, many qualified scientific studies, both old and new, have taken their place in this field. One of the DL methods, Convolutional Neural Networks (CNNs), is that many conventional studies in dentistry identify teeth by non-manual detection or segmentation [6].

CNNs, which are very effective learning and recognition models, include the building blocks of algorithms used to detect objects in data [7]. Thanks to the use of CNN models, it is possible to detect teeth in radiographic images, but sometimes satisfactory performance is not achieved when the existing images are limited. Augmenting the existing images externally or with an automatic AI model is necessary. Although data augmentation for preprocessing is done in classification studies, it has been observed that the purpose of targeted data duplication has yet to be achieved. The need for more data is a disadvantage, especially for deep learning studies with a small dataset. For this reason, synthetic data generation is provided using the Generative Adversarial Networks (GANs) introduced by Goodfellow et al. [8]. GANs are models that can generate data using two different Artificial Neural Networks (ANNs) that compete in their internal structure. In particular, GAN models have been widely used for synthetic data generation in dental data. Thus, small datasets will be expanded by GAN generation. Therefore, these models have been created to eliminate the data generation problem and generate synthetic data that is very similar to reality [9]. The purpose of generative models is to analyze the given training examples to produce the most similar fake data and, thus, to analyze the probability distributions [8]. It is concluded that the performance of a Machine Learning (ML) based algorithm positively depends on the size, cleanliness, quality, and diversity of the input data [10], [11], [12],[13]. The most significant reason for choosing the model predicted for this study as GAN is the lack of an adequate dataset and the fact that radiographic data harms human health, so X-ray data cannot be obtained again.

The literature has many types of neural networks and different data generation and augmentation theories. Although the performance of the data increased by transformation with CNN models is acceptable, the augmented versions of the cropped and zoomed data do not lead to generating different data. A great deal of the previous research into the small sample problem has focused on the "generative classification" paradigm, proposing a semi-supervised framework [7], [8]. While most studies have used GAN models on intraoral and extraoral radiographs, some studies have used high-quality dental computed tomography data. The study by Hu et al. [12], it was stated that more increased quality results would be obtained because the noise, such as low-dose artifacts and blurring found in dental computed tomography data compared to other panoramic data is less. While it is mentioned in the study that low-dose artifacts and soft tissues are detected with high quality, it should be noted that dental objects in extraoral radiographs will be used to produce high-quality data.

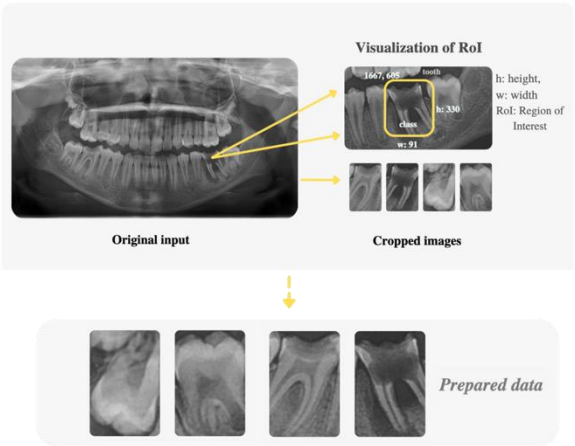


Figure 1. Automatically according to the coordinates of objects in radiographs.

Figure 1 shows that dental objects in a whole panoramic radiographic data are evaluated separately by cropping from their individually labeled bounding boxes. Although GANs can generate new objects by using the features in data, the most remarkable reason for using them individually is because each tooth class exhibits very different features in complex data, such as a molar, premolar, and incisor. The method proposed in the DentaGAN is passed through convolutional filters for conventional deep learning models and generates fake images with a two-stage neural network model. The data generated by the generator network is discriminated by the discriminator and passed through the two-stage structure. The accuracy of the discriminator network in separating spurious samples depends on optimizing the model during its natural training. Therefore, it is ensured that the discriminator neural network is sufficiently trained so that the model can distinguish between real and fake samples. The tooth classes for

the common notation used by dentists were determined and cropped, and each was given to the GAN model as a separate class. The main contribution of our work can be summarized as follows:

- This paper presents a GAN-based approach, which is an innovative algorithm in this field, especially for radiographic images containing individual dental data. This method offers synthetic data as an alternative, especially when real data sets are limited. In addition, this study is the first to implement conventional GAN networks in the field of dental radiography, making it an innovative contribution to the literature. In the literature, GANs have generally been used in medical image processing to expand limited datasets. However, these studies have generally focused on general medical imaging fields such as brain images, eye vessel to eye fundus images in MRI, CT, etc. data.
- The dataset used in the study was evaluated as a total of three separate studies, including all uncropped panoramic images, cropped tooth data from panoramic images, and a separate class according to Dental Federation Notation (FDI) each of which was evaluated.
- The proposed model has effectively increased the diversity and quality of dental datasets. This strategy is crucial for improving the accuracy of dental image analysis models.
- The study showed that traditional GANs are more challenging for generating dental objects than other object classes.
- Since the new radiographic data is close to real-world data and can be used in different neural networks, all caries, damaged, or healthy dental objects are included in the training set.

2. Material and Method

2.1. Generative Adversarial Networks

GANs provide an AL method without the need for annotation files and pieces of information developed as an alternative to learning image spaces. Today, the unique data produced by GANs are used in many different applications, such as image classification, object detection, style transfer, and semantic processing. Thanks to the two different neural network structures it contains, it ensures that the image data produced is statistically indistinguishable from the images in the training set and is new and similar [13]. GANs are also known as generative algorithms, which fall under both generative and discriminative algorithms in ML. For this reason, it has become prevalent nowadays [14]. In the GAN architecture concept, random latent vectors (noise) are passed through the generator to synthesize new data from scratch. Note that there is no direct image input, as the generator only uses the latent vector structure to generate images. Furthermore, since the new images generated from this structure are noise-based, the images in the dataset are only used to check the discriminator for real and fake data. During training, the discriminator is taught to accurately discriminate between real and generated images, ensuring that the generator gets better at generating more realistic images. These two networks are trained in tandem, with the generator continuously adapting to fool the discriminator as the discriminator gets increasingly better at identifying synthetic data.

2.2. Image Generation

Image generation means generating data similar to the input data by adding noise. As shown in Figure 2, it consists of five block layers. The semantic feature vector extracted from the input data is generated in the last dense layer. When the architecture of the generator is examined, the tanh activation unit is used to normalize the last layer from [-1, +1]. Then, the generated vector is given to the next block, the discriminator network [15].

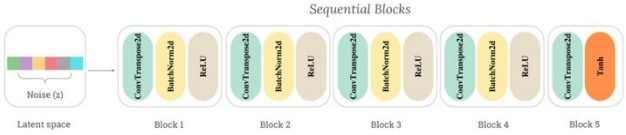


Figure 2. Generator module architecture responsible for image generation for GANs.

In the generator module, the images are passed through certain pre-processing steps to be given to the neural network in the same format. Images are given to the model architecture as “Single Tooth” classes. This architecture input only points to an example model. Otherwise, the model is given per class such as “13”, “48”, or the entire panoramic data.

2.3. Image Discrimination

In the discriminator module, the similarity of the data to the original data is calculated by a discriminator as a result of the combination of real and fake data. In this case, while the generator neural network generates the data in the decoder structure, the generated data is in the decoded position while being transmitted to the discriminator neural network. The crucial task is the discriminator, which tries to reach

the maximum likelihood of predicting real and fake data [16]. As seen in Figure 3, it consists of five dense layers. The sigmoid function was found suitable for the activation function, which will be used in the probability estimation in the last layer. In addition, the reason for using Leaky ReLU activation instead of Rectified Linear Unit (ReLU) is due to the faster convergence of the discriminator network. Thus, we have taken precautions to avoid overfitting the model.

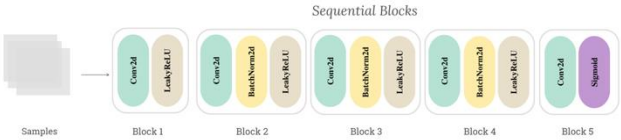


Figure 3. The architecture of the discriminator module responsible for the fidelity of images generated for GANs.

The loss value needs to be backpropagated for the weights updated in the predictions made by the discriminator. Hence, the layers are frozen, and the backpropagation is done with the generator in the weight update. The main method of training the network in GANs is to select the appropriate generator to enable the discriminator to perform classification with maximum accuracy [17]. During the training of the generator network, the optimization phase of the value function in [1] is included [18].

$$V(D,G) = E_{x\sim px}[logD(x)] + Ez\sim pz\left[\log\left(1D\left(G(z)\right)\right)\right] \quad (1)$$

where px and pz indicate original data (x) and generated data distributions (z), respectively [17]. The generator is represented by mapping from the noise space to the data space. $D(x)$ indicates the likelihood that input x comes from data rather than generator G. At the same time, G is trained to minimize $[log(1 - D(G(z)))]$.

2.4. Dental dataset and preparation

The dental dataset used in the study, 565 panoramic radiographs, was annotated by dentists to determine the bounding boxes of dental objects. DentiAssist [19] was used as an object annotation software specifically designed for dental radiographic data. Panoramic radiographs for this study were initially $G(x)$ to be generated as a "Single Tooth" class. Then, considering that the tooth structures in panoramic radiographs are very different, they were cropped by looking at the spatial information in the label files.

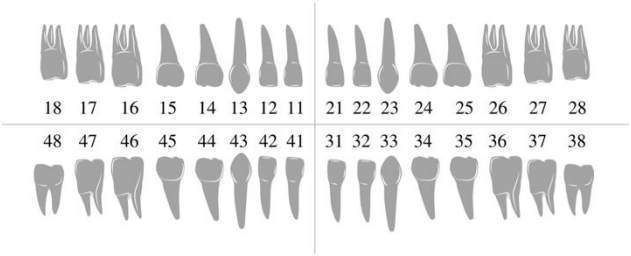


Figure 4. Teeth numbering notation in data labeling and preparation step.

In the FDI notation shown in Figure 4, the superset digits are ordered from left to right, and the subset digits are from right to left. The first digits are named (1 - 2 - 3 - 4) and consist of four separate digits. The reason for using this notation is that it is a universal system of notation in the world. These 32 classes, also labeled in the panoramic radiographs, were cropped from their bounding boxes, as shown in Figure 1. In this way, it also provided the individual evaluation of the teeth. The cropped classes were prepared as "Single Tooth" to be given as input to the generator in the GAN model. In addition, the 32-class tooth categories in the GAN model were considered separately in a complete panoramic class and the single tooth class. Figure 5 shows the data numbers of individual dental datasets in 32 classes. In Figure 5, the names of the individual tooth classes are indicated by the number (11). The percentages written below the numbers represent their contribution to the data set. The difference in the data numbers in the figure is that the molar teeth with the second digit of eight (impacted) are very likely to be extracted in individuals. From this conclusion, the clusters of teeth belonging to the 18-28-38-48 classes are expected to be slightly lower than the discriminator networks. Apart from these situations, the small number of data compared to other datasets is due to the dental structures of the patients who underwent radiographs.

This study uses test data for evaluation after the training phase. A specifically allocated validation cycle needs to be allocated during the training process. It means that the test set is only used for post-training evaluation. Considering that the generator and discriminator performance is continuously evaluated against each other, no direct validation and test sets are defined in the study. On average, 10% of each dataset is reserved for testing and 90% for training. For example, 509 train data were allocated for the training of the panoramic dataset, while 56 were selected for testing. For instance, for a single tooth, this was 13,497 train data and 1,499 test data, while for an example, individual class (class 47) was 421 train data and 46 test data.

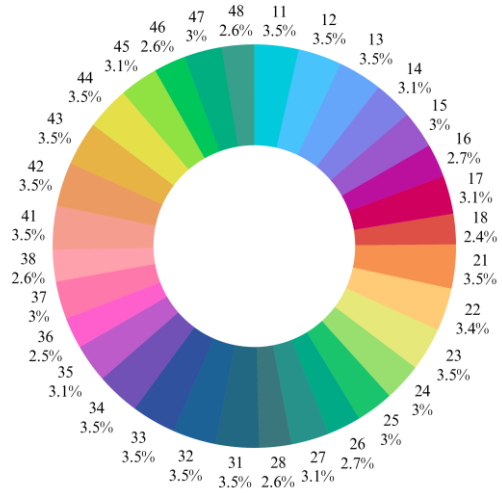


Figure 5. Data distribution statistics for different dataset groups.

During image preprocessing in the generator, batch size 16 was selected, and the image size was 64x64. The scaling, size reduction, and normalization processes are completed before the input data is given to the generator architecture. Then, it passed the relevant data loader and augmentation stage and duplicated the data with certain pre-processing steps.

2.5. Dental dataset and preparation

The architecture of the two-stage neural network structure in dental radiographs is shown in Figure 6. As the first step in the architecture, all panoramic radiographs were given input, while simultaneously cropped panoramic images in 224x224 dimensions were used as input. The original input images are passed through the appropriate pre-processing steps, and the data generated by the generator through latent space and added noises are given to the discriminator. The generated fake and real images of the data set are controlled by a discriminator using the binary cross-entropy loss function. During the training procedure, the output unit should be a number between 0-1 when calculating the loss value for the generator and the discriminator. For this reason, we will calculate the loss value with the binary cross-entropy calculation [20]. The purpose of the function at this stage is to determine whether the generated image is real or fake. The binary cross-entropy loss minimizes the mean probability error between the target and the estimation label for all pixel values found [19].

As seen in the [2], $log(y\hat{}'x)$ for $logD(x\hat{}i)$ wants likelihood close to 1 by capturing satisfactory predictions in real images with gradient ascent. The y in the equation represents the original images, while y' tries to be 1 in this case. Also, for $1 - D(G(z\hat{}i))$, the fake images must be well-estimated, so a probability close to 0 is intended [17].

$$L(w) = -y^i \log\left(y'^{(i)}\right)(1 - y^{(i)})\log\,(1 - y'^{(i)}) \qquad (2)$$

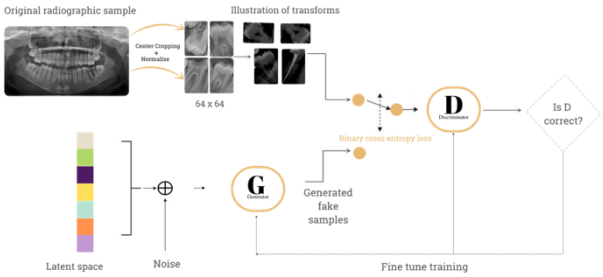


Figure 6. Two-stage generation and discrimination architecture of dental images [8].

Gradient descent in neural networks descends on a static loss surface. In GANs, on the other hand, each downhill step changes the entire surface by a certain amount. The training of the model takes place under difficult conditions, as it is a dynamic system that searches for the experiment between two forces instead of searching for minimum values [17]. Therefore, it was necessary to carefully consider the parameters of the GAN model to be created. In the selected parameters, different learning rates were chosen as the size of the hidden vector z that is, 64 as the size of the generator input, 128 as the size of the feature maps in the builder, ndf: 64 as the size of the feature maps in the discriminator, glr: 0.001 (learning rate of the generator) and dlr: 0.0005 (learning rate of the discriminator) for the optimizers. The same parameters were chosen for comparison in the appropriate epoch training of all the datasets mentioned in the study. One of the hyperparameters that should be used for Adam optimizers, ($beta_1$) was determined as 0.5. In addition, the real label value is 0.9, and the fake label is 0. Following the determined parameters, the original images in the dataset are trained with the discriminator, and the weights are updated during the training. In the figure, starting from the first layer 512 of the created generator neural network, the 2-dimensional transpose convolution layer is used to the last layer 32. Figure 7 is an example of the generator and discriminator pipeline on whole panoramic images. Other datasets are also given as inputs to the generator and discriminator networks. The three output channels from the ConvTranspose2d unit of the generator network are given to the discriminator network in the next section. Then it is transmitted to the 64, 128, 256, and 512 layers, respectively, as shown in the figure. Both weight updates and feature extractions take place in this step. Pixel distance and feature distance are compared in images for diversity and fidelity. Therefore, evaluating GANs is a challenging task [21], [22].

Algorithm 1. Pseudocode of the training procedure for the proposed approach

Input: *maxepoch, dataloader,generator,discriminator, classifier,cof ,adw*
for *epoch in range(maxepoch)* **do**
for *(realdata, reallabel, embedding)* in *dataloader* **do**
batchsize = len(realdata)
#step 1: generate fake images and train the discriminator with real data
genimg=generator(random((batch, 128, 1, 1)), embedding)
predreal=discriminator(realdata), BCE(predreal, ones((batchsize, 1))).backward()
#step 2: train the discriminator with generated data
predfake=discriminator(genimg), BCE(predfake, zeros((batchsize, 1))).backward()
#step 3: train the generator and opt. discriminator
predgen=discriminator(genimg), BCE(pred−gen, zeros((batchsize, 1))).backward()
#step 4: train the classifier with real data
outreal=classifier(realdata), classification−loss(outreal, reallabel)
#step 5: train the classifier and data generated in the sample control
fakeprob=classifier(genimg)
outlabel=argmax(fakeprob, dim = 1)
if *softmax(fake)[reallabel]* $\geq$ *cof* **then**
loss'=adw $*$ *classificationloss(selectout, selectlabel)*
loss'.backward()
end
end
end
Output: Generated synthetic data (radiographic image)
Input: *maxepoch, dataloader,generator,discriminator, classifier,cof ,adw*
for *epoch in range(maxepoch)* **do**
for *(realdata, reallabel, embedding)* in *dataloader* **do**

The algorithm of the training procedure for the GAN model is as follows. The input image and the label are generated for each epoch, creating fake images in the specified batch size. The data loader parameter is central to loading the training data. It provides iterability over the dataset with the shuffle parameter: "True" option. When n epochs are performed for Algorithm 1 the images created by the generator are kept. Then, the discriminator model takes the real data and provides a backward prediction for binary cross-entropy loss. The total error is propagated back and updated. The generator and the discriminator are trained for original and fake samples that provide sequences of zeros and ones, the result to be detected in the output [22]. Thus, predicted data is generated. The classifier in steps 5 and 6 of Algorithm 1 represents a general neural network. Therefore, the classifier refers to both the generator and the discriminator networks. Also, a classical GAN structure has only a generator and a discriminator. However, this algorithm has an additional classifier. The task of the classifier is to support the learning process of the model by classifying real and generated data. In other words, the generator generates fake images using random input and embedding, the discriminator is the network that tries to distinguish whether these images are real or fake, and the classifier is the structure that helps the model to better learn the difference between real and fake data. As a result, the classifier acts as an additional control mechanism for the GAN. A classification process is performed between the input images and the new images generated, and they are checked with softmax condition. Accordingly, the loss value is updated. Because GAN models mostly consist of two models that can be tuned to each other, this training demonstrates the struggle between the two neural networks [23]. In summary, we maximize D(discriminator) and $log(D(x)) + log(1 - D(G(z)))$ by maximizing $log(D(G(z)))$ G(generator) is updated. Evaluation is an open area in generator model research. The FID metric in [3] is the most popular for evaluating the success of the data generated by the generator. The Inception-v3 network

measures the performance of the fake images generated. The purpose of using this initial network is to extract features from the intermediate. The data is distributed using a multivariate Gaussian distribution with mean and covariance for statistical purposes [16], [23].

$$FID = |\mu_1 - \mu_2| + T_r(\sigma_1 + \sigma_2 - 2\sqrt[2]{\sigma_1 * \sigma_2}) \quad (3)$$

where Tr sums all the diagonal elements, $\mu 1$ and $\sigma 1$ are the mean and covariance of the training data, and $\mu 2$ and $\sigma 2$ are the mean and covariance of the test data [16]. It calculates the distance of the curves between the fake and the real embeddings. The closer these statistics are to each other, the closer the fake embeddings model is to the real embeddings. Therefore, a smaller value means that features in the reals and fakes are more similar, so the lower the FID, the better and closer distributions.

3. Experimental Results and Discussion

To obtain the experimental results, a device with NVIDIA Geforce RTX 3050 graphics card support was used in the GAN model training. GANs with the same parameters and hyperparameters were generated for 34 dental datasets mentioned in the paper. The experimental findings obtained under the specified conditions are given in Table 1. While training with the generator continues, the high threshold value set by the discriminator ensures that the generator obtains a thriving fake image. When the table is examined, it is revealed that different datasets are used. Taking into account the possible errors that can occur when generating panoramic images, one of the main objectives of the study was to properly trim the data from the bounding boxes. One of the most significant reasons for this is that although GANs generate successful images, radiographic data is more complex than simple data has a disadvantage. As a first step in the study, panoramic data were generated, and due to the high FID value, even if changed the parameter, it was necessary to find another solution. Thus, the teeth were cropped according to the FDI notation, and all the crops were used as a given dataset. In this study, 500 epoch training sessions were performed on dental radiographs and single tooth images to ensure a fair approach. The average training time of the generative and discriminative networks simultaneously per dataset is 7466 seconds, about 124 minutes for the individual class (41 classes). The single tooth class of combined teeth trained in the study: 14,996 images, and the panoramic dataset contains 565 images. Also, the number of individually cropped tooth images varies depending on the panoramic images, as shown in Figure 5. This situation is mainly since sometimes cases such as impacted teeth and molars are not included depending on the age of the extracted individual.

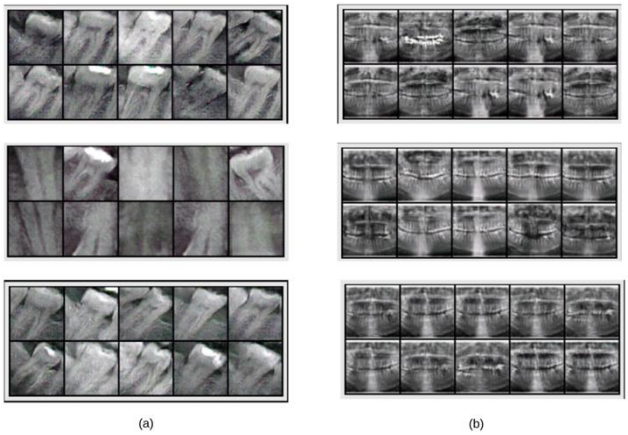


Figure 7. Fake images generated because of training GANs. (a) the results of cropped images, (b) the result of an entire panoramic image.

Nevertheless, the total amount of data generated for all classes reached 10,000. Considering the results in Table 1, the single tooth class has a large amount of data, which leads to a lower FID result. The training time for the single tooth class was about 3,420 minutes, or 57 hours, because it contains a lot of data. In addition, other individual tooth classes produced different results depending on their number and complexity in the dataset. The panoramic dataset contains 565 images, but it also contains a wider range of mouth views and patterns, which increases the time for the neural networks to scan the data. Therefore, producing the entire panoramic image takes 9,300 seconds, or 155 minutes. During the training of this data, the loss of the generator reached 2.9414, while the loss of the discriminator reached 1.2484.

The fake images generated by neural network are shown in Figure 7, and according to this figure, it was observed that the individual tooth classes were generated by GAN much better than all the panoramic images. However, more was needed to restrict all images generated to one class due to the differences in the mouth's molar, premolar, and incisor classes. For these reasons, one-sided label softening was performed while adjusting the parameter. For these reasons, one-sided label softening was performed when adjusting the parameter. An excessively high confidence value can cause some problems with deep neural networks. If the discriminator depends on the small feature map while recognizing the actual images, the generator generates these features to

benefit from the discriminator. In this case, the optimization performed can be very greedy. To avoid such problems in real image estimation, the discriminator is penalized when it exceeds $0.9(D(real > 0.9))$. The target label value has been updated to 0.9 instead of 1.0.

Since FID measures the distributional similarity between the generated and real images, the test set provides a realistic example of this distribution. The results of the test data are shown in Table 1 below. In contrast, the discriminator loss value is excessively high, causing overfitting due to the imbalance between the generator and the discriminator. Overfitting is recognized as one of the major challenges in training GANs. Suffering from such difficulties, the experimental results in this study may not have achieved the expected performance in some individual tooth classes. However, the overall panoramic view was worse in all conditions than in other individual tooth classes. On the other hand, it is expected that the single tooth class created with samples from single tooth classes will have high performance due to the high diversity of data.

Table 1. Findings of neural networks generated and discriminated for different dataset.

Generated class	Loss G	Loss D	Fréchet Inception Distance	Generated class	Loss G	Loss D	Fréchet Inception Distance
Single tooth	2.9414	1.2484	124.8310	Tooth 31	3.4235	0.5093	102.8283
Tooth 11	4.7493	0.4203	187.8658	Tooth 32	3.8033	0.4238	114.4682
Tooth 12	5.0410	0.3903	178.6688	Tooth 33	4.2498	0.6059	144.1210
Tooth 13	3.3994	0.4371	127.4457	Tooth 34	4.0968	0.3835	153.1053
Tooth 14	4.3447	0.3905	146.7246	Tooth 35	4.0497	0.4592	158.1447
Tooth 15	4.6472	0.4127	173.7276	Tooth 36	4.1037	0.5133	213.4831
Tooth 16	3.8004	1.1397	160.5275	Tooth 37	3.1822	0.4515	127.4164
Tooth 17	6.7436	0.6928	297.8457	Tooth 38	4.0272	0.4591	175.5649
Tooth 18	3.8478	0.4241	138.1094	Tooth 41	4.9409	0.3572	151.7190
Tooth 21	3.8511	0.4163	169.2506	Tooth 42	3.1018	0.7503	190.0691
Tooth 22	3.8585	0.3945	181.6382	Tooth 43	4.1703	0.6607	131.0531
Tooth 23	4.2090	0.7617	126.5523	Tooth 44	5.3700	0.3606	150.3511
Tooth 24	5.1406	0.3366	179.1189	Tooth 45	3.1822	0.4515	127.4164
Tooth 25	5.2565	1.4472	215.2094	Tooth 46	4.8492	0.9850	171.0522
Tooth 26	3.0496	1.7622	163.5350	Tooth 47	4.4303	0.4255	257.3137
Tooth 27	4.4674	0.6183	215.6518	Tooth 48	4.6258	0.3531	226.9374
Tooth 28	4.8688	0.7251	135.4039	Panoramic	3.9198	0.4527	223.4762

On the other hand, our study adapts existing GAN algorithms to dental radiography, enabling the synthetic generation of individual tooth images. However, datasets with more complex and individual characteristics, such as dental structures, are being studied. This requires more detailed modeling than general medical imaging studies. As a result, in terms of the applicability of GANs to radiographic images and the solution to data limitations in dentistry, this study makes an important contribution to the literature. In particular, it is necessary to accurately generate anatomical details and anomalies that may be present in radiographic data, such as periodontal structures, root canals, or differences due to anomalies such as caries and restorations. However, the current architecture of conventional GANs is insufficient to maintain such a precise level of detail.

4. Conclusion and Suggestions

This paper implements a novel method for generating and augmenting synthetic individual dental data in radiographic images using GANs. The results show that the recommended strategy is effective in increasing the diversity and quality of dental datasets, both of which are necessary to improve the accuracy of dental image analysis models. Synthetic data alleviates the problems caused by the lack of real-world datasets and provides a reliable alternative for building robust DL models in dental radiography. Results have shown that the AI generated by the DentaGAN model produces synthetic data that is as high quality as actual dental radiographs, with the added benefit of being adaptable to a range of clinical settings. This capability has significant potential in the field of dental informatics as it provides larger and more

representative training datasets, leading to more robust models and, therefore, better patient outcomes. However, when the FID findings for each class were analyzed, it was observed that evaluating alternative GAN methods in future research may be appropriate. The findings suggest that the performance can be improved in the future by experimenting with different GANs with larger-sized images.

In future studies, although the individual tooth classes are balanced in the dataset, the data of the tooth classes that are difficult to learn will be increased. As a result of experimental studies and findings, we have concluded that generating dental objects with the traditional GAN model is more challenging than other object classes. The results of the study indicate that conventional GANs are more difficult to generate dental objects than other object classes. This finding suggests that experimenting with different GANs on larger images may improve performance in the future. Therefore, in the next stage of the study, it will be possible to increase the possible performance by working with different types of GAN models with large image sizes.

Contributions of the authors

B. Y. Kazangirler: Methodology, visualization, wrote the paper, performed the experiments, interpretation of the data
C. Ozcan: Conceived and designed the analysis, investigation, wrote the paper, revised the manuscript

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] J. Park *et al.*, "Deep learning on time series laboratory test results from electronic health records for early detection of pancreatic cancer," *J. Biomed. Inform.*, vol. 131, p. 104095, Jul. 2022, doi: 10.1016/j.jbi.2022.104095.

[2] S. Vinayahalingam *et al.*, "Automated chart filing on panoramic radiographs using deep learning," *J. Dent.*, vol. 115, p. 103864, Dec. 2021, doi: 10.1016/j.jdent.2021.103864.

[3] M. J. Cardoso, N. Houssami, G. Pozzi, and B. Séroussi, "Artificial intelligence (AI) in breast cancer care - Leveraging multidisciplinary skills to improve care," *Breast Off. J. Eur. Soc. Mastology*, vol. 56, pp. 110–113, Dec. 2020, doi: 10.1016/j.breast.2020.11.012.

[4] B. Y. Tekin, C. Ozcan, A. Pekince, and Y. Yasa, "An enhanced tooth segmentation and numbering according to FDI notation in bitewing radiographs," *Comput. Biol. Med.*, vol. 146, p. 105547, 2022.

[5] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[6] D. Frejlichowski and R. Wanat, "Application of the Laplacian Pyramid Decomposition to the Enhancement of Digital Dental Radiographic Images for the Automatic Person Identification," in *Image Analysis and Recognition*, A. Campilho and M. Kamel, Eds., Berlin, Heidelberg: Springer, 2010, pp. 151–160. doi: 10.1007/978-3-642-13775-4_16.

[7] Y. Lin *et al.*, "DHI-GAN: Improving Dental-Based Human Identification Using Generative Adversarial Networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 9700–9712, Dec. 2023, doi: 10.1109/TNNLS.2022.3159781.

[8] S. Tian *et al.*, "DCPR-GAN: Dental Crown Prosthesis Restoration Using Two-Stage Generative Adversarial Networks," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 1, pp. 151–160, Jan. 2022, doi: 10.1109/JBHI.2021.3119394.

[9] R. Havale, B. S. Sheetal, R. Patil, R. Hemant Kumar, R. T. Anegundi, and K. R. Inushekar, "Dental notation for primary teeth: a review and suggestion of a novel system," *Eur. J. Paediatr. Dent.*, vol. 16, no. 2, pp. 163–166, Jun. 2015.

[10] N. K. Singh and K. Raza, "Medical Image Generation Using Generative Adversarial Networks: A Review," in *Health Informatics: A Computational Perspective in Healthcare*, R. Patgiri, A. Biswas, and P. Roy, Eds., Singapore: Springer, 2021, pp. 77–96. doi: 10.1007/978-981-15-9735-0_5.

[11] "U-Patch GAN: A Medical Image Fusion Method Based on GAN | Journal of Imaging Informatics in Medicine." Accessed: Nov. 29, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s10278-022-00696-7

[12] Z. Hu *et al.*, "Artifact correction in low-dose dental CT imaging using Wasserstein generative adversarial networks," *Med. Phys.*, vol. 46, no. 4, pp. 1686–1696, Apr. 2019, doi: 10.1002/mp.13415.

[13] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018, doi: 10.1109/MSP.2017.2765202.

[14] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3313–3332, Apr. 2023, doi: 10.1109/TKDE.2021.3130191.

[15] A. Borji, "Pros and cons of GAN evaluation measures: New developments," *Comput. Vis. Image Underst.*, vol. 215, p. 103329, 2022.

[16] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, Accessed: Aug. 24, 2024. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2016/hash/8a3363abe792db2d8761d6403605aeb7-Abstract.html

[17] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," Jan. 07, 2016, *arXiv*: arXiv:1511.06434. doi: 10.48550/arXiv.1511.06434.

[18] A. Figueira and B. Vaz, "Survey on synthetic data generation, evaluation methods and GANs," *Mathematics*, vol. 10, no. 15, p. 2733, 2022.

[19] A. Karaoglu, C. Ozcan, A. Pekince, Y. Yasa, B. Tekin, and D. Ozdemir, "Automatic dental segmentation module supported by artificial intelligence for dentistry students education," *Artif Intell Theory Appl*, vol. 1, pp. 180–190, 2021.

[20] U. Ruby and V. Yendapalli, "Binary cross entropy with deep learning technique for image classification," *Int J Adv Trends Comput Sci Eng*, vol. 9, no. 10, 2020, Accessed: Aug. 24, 2024. [Online]. Available: https://www.researchgate.net/profile/Vamsidhar-Yendapalli/publication/344854379_Binary_cross_entropy_with_deep_learning_technique_for_Image_classification/links/5f93eed692851c14bce1ac68/Binary-cross-entropy-with-deep-learning-technique-for-Image-classification.pdf

[21] N. Shibuya, "Understanding generative adversarial networks," *Retrieved Internet*, 2017.

[22] M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet, "Are gans created equal? a large-scale study," *Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, Accessed: Aug. 24, 2024. [Online]. Available: https://proceedings.neurips.cc/paper/2018/hash/e46de7e1bcaaced9a54f1e9d0d2f800d-Abstract.html

[23] J. Liu, C. Gu, J. Wang, G. Youn, and J.-U. Kim, "Multi-scale multi-class conditional generative adversarial network for handwritten character generation," *J. Supercomput.*, vol. 75, no. 4, pp. 1922–1940, Apr. 2019, doi: 10.1007/s11227-017-2218-0.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1205-1215 YEAR: 2024
DOI: 10.17798/bitlisfen.1540768

Environmental Problems in Mining Quarries and Theoretical and Practical Comparison of Dust Emissions

Volkan SEREN[1*], Ahmet KILIÇ[2]

[1] *MERS Çevre Madencilik Mühendislik Proje Danışmanlık iş Sağ. ve Güv. İnş. San. ve Tic. Ltd. Şti., Ankara, Türkiye*
[2] *Aksaray University, Engineering Faculty, Department of Environmental Engineering, Aksaray, Türkiye*
(ORCID: 0009-0001-8413-5441) (ORCID: 0000-0002-2365-3093)

Keywords: Mine Quarry, AERMOD, PM10, Settled Dust, Emission.

Abstract
This study reveals that the development of the country is only possible through meticulous analysis of regional potentials and sustainable use of natural resources. The aim of this study is to evaluate the environmental impacts arising from the site selection of mining operations and to put forward the necessary measures for the efficient and controlled use of natural resources within the scope of sustainable development. In this context, the importance of factors such as expropriation, topography, environmental conditions and energy and water access in the site selection of mines is emphasized.
In the study, the AERMOD model developed by the United States Environmental Protection Agency for air quality modeling was used to estimate the environmental impacts of dust emissions that may occur during mining activities. In the modeling process, data on pollutant sources such as mass flow rate, source height and gas outflow rate were combined with meteorological data and calculated according to the Regulation on the Control of Industrial Air Pollution. In addition, in-plant PM10 measurements were performed and analyzed by gravimetric method.
As a result, the modeling studies and the data obtained serve as a basis for the environmental permitting process of the facilities that will be put into operation and provide guidance to decision makers for air quality management. The study also emphasizes the importance of assessing the cumulative air quality impacts of existing facilities and demonstrates that verifying these impacts with reliable laboratory studies will contribute to sustainable development targets.

1. Introduction

The environment is an environment in which living things maintain their relations during their lifetime in mutual interaction and is defined as the environment in which living things sustain their lives. At the same time, the environment is directly and indirectly affected by air, soil and water pollution that occurs because it is in social, economic, cultural as well as physical, chemical and biological interaction. Changes in air composition can disrupt the environmental balance and threaten the lives of humans and other living things [1]. Environmental pollution can be caused by a wide variety of reasons and has negative effects on the ecosystem. There are many pollution sources that create complex and multidimensional pressure on the ecosystem. Urbanization, industrialization, mining and agricultural activities are among these sources [2].

When environmental problems reach crisis proportions, negative impacts on human health occur. Environmental crises occur when the saturation capacities of receiving environments are exceeded and the balances in the natural environment are disrupted. These balances are especially disrupted by the unconscious consumption of natural resources by humans [3]. With the increase in these problems day by day;

*Corresponding author: vlknsrn@hotmail.com
Received: 29.08.2024, Accepted: 30.11.2024

two ways of taking measures against environmental problems have emerged. The first one is to eliminate the problems after they occur (restorative policies) and the second one is to prevent them before they occur (preventive policies) [4].

Strategic documents (strategies, policies, legislation) and specific procedures for decision-making for specific activities have been developed by many countries since the 1980s [5]. However, environmental assessment systems have some problems because they are limited to the study scale [6].

Environmental Impact Assessment (EIA) is used to comprehensively determine the positive or negative impacts that a planned activity may have on the environment. The EIA aims to prevent or minimize potential damages with the help of assessments. At the same time, the goal of EIA is to ensure that all negative environmental impacts that may be caused by planned activities are identified in advance and necessary measures are taken without hindering economic and social development [7]. This method has been widely used especially in industrialized countries since the 1970s [8].

EIA reports are prepared to prevent environmental pollution before and during the operations of the investments within the scope of the regulation. In this context, these reports help investors to develop methods to prevent environmental pollution [9]. Environmental management, which is the key to sustainable production in mining and industry, although costly at first, provides a competitive advantage to mining companies and reduces the environmental impacts of mining activities with the help of legislation [10].

Mines that start to operate without taking the necessary precautions into consideration pose a great danger to public health and environmental quality [11]. The magnitude of the problems defined in this context varies according to some variables (types of minerals extracted, mining methods used, types of waste materials to be generated, geological structure of the selected area, soil structure and vegetation in the mining area, size and depth of the sites, hydrological and hydrogeological characteristics and climatic conditions, etc.) [12].

According to the 2023 data of the General Directorate of Mining Affairs of the Ministry of Energy and Natural Resources, a total of 14.763 mining licenses were granted, of which 4.693 were exploration licenses and 10.070 were operating licenses [13]. Mining operations create various negative impacts that lead to the depletion of non-renewable natural resources, pollution of air, water and soil, noise and vibrations, deterioration of topography and destruction of vegetation [14].

The extent of harmful effects on the environment in open pit mines varies depending on hydrology-hydrogeological characteristics, geological structure, capacity (size and depth) of the mining areas, existing soil properties, vegetation of the region, climatic conditions, etc. [15]. Open pit mining method can be generally defined as the removal of the top layer consisting of topsoil and stripping material (rock mass ore-free zone) over the ore underground or in outcrop and bringing the ore to the surface [16]. In this method of operation, the noise-vibration problem caused by construction machinery and step drilling-blasting operations in production and dust emissions released during pickling (top cover) operations can be considered as other major environmental problems [17].

The effects of pollutant emissions, which can cause serious health problems in living organisms and pose environmental problems, can manifest themselves as respiratory irritation, allergic reactions such as asthma and, in the long term, serious diseases such as cancer. These health problems can vary depending on the type of dust exposure and the risk is particularly high when small particles enter the body. The health risks are directly linked to particle size, and particles smaller than 10 μm in particular are known to cause serious health problems by penetrating the lungs and entering the bloodstream. Similarly, studies have found that dust contact with plant leaves and dust accumulation have negative effects on plant growth and inhibit photosynthesis, respiration and evaporation from leaves. Quarry dust negatively affects the functionality of the ecosystem due to loss of vegetation and damage to agricultural and forest lands [18].

Various legal regulations have been developed in our country and other developed countries in order to prevent and control the pollution that occurs during the production activities of industrial facilities and causes environmental pollution by exceeding the limit values when they are not adequately controlled. These regulations include measures to reduce environmental impacts.

The Environmental Law (No. 2872), which entered into force in 1983 within the scope of environmental legislation, is the first legal regulation for the protection of the natural

environment and the prevention of environmental pollution in our country. This law imposes the obligation to treat and dispose of wastes generated as a result of production and consumption activities in accordance with the standards specified in the relevant regulations. According to the Environmental Permit and License Regulation (EPLR), mines above certain capacities must obtain an environmental permit and are obliged to take necessary measures to control dust emissions. Accordingly, according to the EIR [19], quarries with a mine production capacity of 150 tons/day or more and quarries where minerals in group I (a and b), group II including limestone, group IV and group V, and mines where explosive materials are used are subject to environmental permits. The quarries within this scope are obliged to take dust preventive measures for dust emissions generated during the operation phase and to meet the limit values mentioned in the Regulation on the Control of Industrial Air Pollution (RCIAP) and to obtain an environmental permit for emissions.

This study covers the comparison of the emission results of the gypsum stone quarry and crushing and screening plant within the borders of Güblüce village, Zile district, Tokat province during the planning and operation phase and the ways to be followed according to the results.

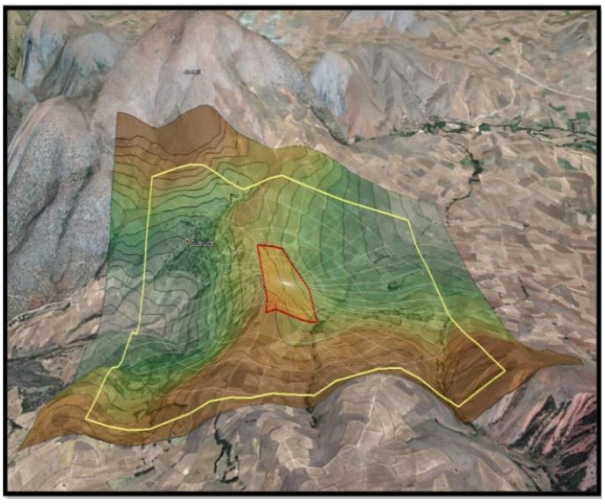


Figure 1. Facility Impact Area Map (40,140576-35,667532)

This research is a comparison of the data of the Project Introduction File made in 2012 with the regulation limit values for 2024.

2. Material and Method

In this study, the emission measurement results of a blasted gypsum quarry and crushing and screening plant, which operates 8 hours a day and has a production capacity of 62.53 tons per hour, were evaluated. The facility is located in Güblüce Village, Zile district of Tokat province. The quarry operates with open pit method and production operations are carried out by blasting and drilling methods. Air quality control can be considered at three different levels: global, regional and local. The most important of these is local air quality control and the air quality management applied for this purpose. [20]. The concept defined as air quality management covers all detection, monitoring and improvement works carried out to ensure that emissions released into the atmosphere without treatment are kept at levels that will not harm human and environmental health or reduced to these levels. This process aims to implement the necessary measures to control emissions and protect air quality [21]. In our country, limit values for 13 pollutants (SO_2, PM_{10}, NO_X) related to air quality management are determined by the Air Quality Assessment and Management Regulation (AQAMR) [22].

Modeling is used to express an event mathematically using numerical data in terms of the factors on which the event depends [23]. In modeling, many factors affecting dispersion and transport and possible effects of pollutants are calculated. There are multiple models that can be used to calculate atmospheric pollutants [24, 25]. These modeling methods are useful and economical for the analysis of dispersion and dilution behavior [25]. One of these models, AERMOD, was developed in 1991 as a result of a joint effort by AMS and EPA (Environmental Protection Agency) scientists. In 2005, the AERMOD model was officially recognized by the EPA. AERMOD is a steady-state dispersion model designed for short-range (maximum 50 km) dispersion of air pollutant emissions from stationary industrial sources [26].

AERMET is a three-stage meteorological data preprocessor. In the first stage, surface and upper layer data are extracted, processed together and quality checked. In the second stage, all data are merged into a single file with a 24-hour period. In the final stage, this combined meteorological data and the calculated layer data are prepared for use in the AERMOD dispersion model. The PFL and SFC files produced by AERMET are used by AERMOD. With the help of these data, atmospheric parameters are included in the simulation [26].

Emission factors (given in Table 1 of Annex-12 of the same regulation) were used to

calculate the mass flow rates of dust emissions (resulting from the operations to be carried out at the mine site) in accordance with the article "Hourly mass flow rates of emissions discharged from the facilities to the atmosphere are determined by taking measurements from the stacks for existing facilities and by using emission factors for non-stack sources and new facilities to be established." in Annex-2 (Article 1, subparagraph a) of the Regulation on the Control of Industrial Air Pollution. [27].

Table 1. Emission Flow Rate of Dust According to Source Definitions

Source	No	Process	Controlled Dust Emission flow rate (kg/hour)	Controlled Dust Emission flow rate (kg/hour)
Vegetal Soil Excavation	1	Excavation	0.04	0.08
	2	Loading	0.016	0.0325
	3	Transportation	0.018	0.036
	4	Unloading	0.0007	0.0014
	5	Storage	0.0145	0.029
		TOTAL	**0.09**	**0.18**
Quarry Activities (Ore Production)	6	Excavation	0.78	1.56
	7	Loading	0.312	0.625
	8	Transportation	0.35	0.7
	9	Unloading	0.312	0.625
Quarry Activities (Minepass Production)	10	Loading	0.031	0.0625
	11	Transportation	0.03	0.06
	12	Unloading	0.031	0.0625
		TOTAL	**1.846**	**3.695**
Mineral Storage	13	Storage	0.07	0.14
		TOTAL	**0.07**	**0.14**
Waste Storage	14	Storage	0.0676	0.135
		TOTAL	**0.0676**	**0.135**
Final Transportation	15	Transportation	0.5075	1.015
		TOTAL	**0.5075**	**1.015**
Explosion	16	Explosion	0.7425	62.53
		TOTAL	**0.7425**	---
TOTAL		**TOTAL**	**2.5**	**5**

AERMOD View program developed by USEPA was used in the air quality modeling study. The AERMOD Modeling System is a flexible modeling tool for linear and steady-state plume modeling with Gaussian plume characteristics. It can be applied to various emission sources, such as point, volumetric and areal source types. AERMOD evaluates new or steady-state dispersion scenarios with sophisticated algorithms that include vertical profiles of winds, turbulence effects, penetration of rising inversion layers, Plume rise, surface levels and heat at buoyancy sources, and terrain effects. The program optimizes its performance according to the characteristics of the terrain and different emission sources, allowing to obtain the most realistic results at high concentrations. The AERMOD modeling system consists of two core components, AERMAD for processing terrain information and AERMET for processing meteorological data. It is also designed to model various types of sources such as multiple point sources, areal, linear and volumetric sources, structures, concentrations and deposition. [28].

The areas where the AERMOD Modeling System is used are as follows:

- Complex Industrial Sources: It can be applied on single or multiple point, areal, linear and volumetric sources. However, structures and impacts from structures are not considered in this model.
- Particulate and Gas Deposition: Used to assess the accumulation of gases and particles released into the atmosphere.
- Constant and Variable Emissions: Both constant and time-varying emission sources can be analyzed.
- Rural and Residential Areas: Applicable in both rural and urban areas.
- Transport Up to 50 km from Source: The transport of emissions up to 50 kilometers from the source can be modeled.
- Concentration Estimation for All Terrain: Concentration estimates are made by considering different terrain structures. [28].

The modeling study for estimating the values of dust and gaseous pollutants in ambient air by mathematical methods includes the following steps:

- Determination of the Dispersion Region of the Sources: In the first step, the dispersion region where the pollutant sources will be effective is defined.
- Mapping the Dispersion Region: The defined dispersion region is divided into squares of certain dimensions in the Cartesian coordinate system and latitude, longitude and elevation data are obtained. Alternatively, in the polar coordinate system, the region can be mapped angularly according to latitude, longitude and altitude values.
- Collecting Information on Pollutant Sources: In this step, basic information such as mass flow rate of pollutant sources, source height, source diameter, gas outflow rate, gas temperature are collected.
- Collection of Meteorological Data: Hourly meteorological data (such as hourly temperature values, wind speeds and directions) for a representative year are collected
- Analysis of Meteorological Data: Using the collected meteorological data, hourly stability classes are determined and mixing heights are calculated. [29].

After transferring the data listed above to the modeling program, the program is run to estimate the hourly, daily, monthly and annual probable ground level concentration values of the pollutants in the ambient air. This modeling study was carried out to estimate the air pollution that may occur in the quarry operation. In the study, particulate matter - dust (PM) pollutant parameter resulting from quarry operations was modeled. Before the plant starts operations, the environmental permitting process for emissions of mines is evaluated within the scope of the RCIAP and the EPLR. The purpose of environmental permits is to encourage activities and facilities that may have a polluting impact on the environment to take the necessary measures to protect the environment during their operations. Environmental permit is a tool to prevent environmental pollution and provides legal intervention of the state against polluting activities. [30].

In the study, PM 10 measurements were made at 6 points within the facility. In addition, the modeling results were examined and settled dust measurements were made for a total of 2 months in 1-month periods at two points within the investigation area of the facility with high model results.

Table 2. Emission sources and measured parameters

No/ Code	Emission Source	Parameter						
					Powder			
		CO	NO_x	SO_2	PM10	Settling Pewder	VOC	Other
1	Sieve Side	-	-	-	X	-	-	-
2	Crusher Side	-	-	-	X	-	-	-
3	Stock Area	-	-	-	X	-	-	-
4	Quarry (Blasting Moment)	-	-	-	X	-	-	-
5	Bunker Side	-	-	-	X	-	-	-
6	On-site Road	-	-	-	X	-	-	-
7	Facility Inspection Are (2 Points)	-	-	-	-	X	-	-

Emission sources are indicated in Table-2. It has been determined what kind of measurements will be made at these emission sources.

PM10 measurements are made by gravimetric measurement method (TS EN 12341) by keeping particles smaller than 10 microns in diameter on filter paper. For these measurements, MCZ LVS 1 measuring device is used because the air vacuum time and volume can be adjusted. First, the filter papers to be used in the method are weighed by conditioning at 20 °C (±1 °C) temperature and 50% (±5%) relative humidity for 48 hours. The weighing results are recorded and placed in clean petri dishes to avoid exposure to air, moisture and contaminants until the sampling point. The measuring device, which is transported to the appropriate sampling point, is placed on a flat area (at least 30 cm away from the obstacle) where there is no obstacle that may prevent sampling from the air to be analyzed, and measurement and sampling are performed according to the relevant procedure. After the sampling process is completed, the filter paper is placed in a clean petri dish with the help of tweezers and sent to the laboratory for weighing and the weighing results obtained are recorded.

Determination of settled dust is carried out by gravimetric method in the plant inspection area and with a four-way ambient air sampling device in accordance with TS 2342 standard. The sampling system, consisting of a base plate (1 piece), tripod (1 piece) and dust collection container (4 pieces), collects settled dust from 4 main directions and can determine the source of dust at a certain point.

In the study, Air Quality Dispersion Model was carried out during the production planning phases of the gypsum quarry and crushing and screening plant by applying the blasting method with a capacity of 62.53 tons per hour, in accordance with the Air Quality Dispersion Model in Mining Projects guideline published in 2018 by the General Directorate of EIA Permit and Inspection of the Ministry of Environment, Urbanization and Climate Change [31]. After the facility became operational, dust emissions causing air pollution were taken under control within the scope of SDGPLR and the dust measurement results were measured to be the basis for the environmental

permit on emission required to be obtained in accordance with the EPLR.

3. Results and Discussion

In mines, dust is generated at different stages of the production process, during drilling, blasting, machine excavation of the main mass, transportation and storage. These dust emissions adversely affect air quality by disturbing the composition of the air as an areal source. AERMOD model was used for the dust emissions to be generated within the scope of the research and PM10-Precipitated Dust emissions were calculated as annual-monthly-hourly in the Long-Short Term based on controlled and uncontrolled operations during blasting and production phases. The results were evaluated within the scope of Air Quality Assessment and Management Regulation.

After the site became operational, PM10 was measured at 6 points with MCZ LVS 1 PM 10 particulate matter measurement device within the scope of environmental permits and settled dust was determined by Gravimetric Method with Four-Way Ambient Air Sampling Device within the facility area. The results were evaluated within the scope of Air Quality Assessment and Management Regulation.

As a result of the evaluations, the data in Table 3 were obtained by using the AERMOD model in the study phase (theoretical) and operation phase (practical) of dust emissions generated outside the blasting operations.

Table 3. Comparison of PM10 and settled dust results outside blasting operations

Parameter	**Period**	**Model Result Value [µg/m³]**		**Limit Values by Year**		
				2018	**2019-2023**	**2024 onwards**
PM 10 (AERMOD)	STL Daily	Scenario 1	32	**60**	**50**	**50**
		Scenario 2	233			
	LTL Anual	Scenario 1	6	**44**	**40**	**40**
		Scenario 2	45			
PM 10 (Measuring)	STL Daily	Field Measured	92,65	**60**	**50**	**50**
	LTL Anual	Field Measured	5,33	**44**	**40**	**40**
Parameter	**Duration**	**Limit Value [mg/m² day]**		**Limit Values by Year**		
				2018	**2019-2023**	**2024 onwards**
Settling Dust (AERMOD)	STL Monthly	Scenario 1	298	**390**	**390**	**390**
		Scenario 2	2168			
	LTL Monthly	Scenario 1	24	**210**	**210**	**210**
		Scenario 2	178			
Settling Dust (Measuring)	STL Monthly	Field Measured	287,06	**390**	**390**	
	LTL Anual	Field Measured	47,85	**210**	**210**	

Short Term Value (STV): The maximum daily average values or, statistically, the value that corresponds to 95% of the measurement results when all measurement results are ranked according to their numerical magnitude.

Short Term Limit Value (STLV): It is the limit value that statistically corresponds to 95% of all measurement results when the maximum daily average values or numerical values are arranged in order of magnitude and should not exceed the value specified in of Annex-2 of the RCIAP

***Long Term Value (UVV)*:** The value calculated as the arithmetic average of all measurement results.

Long Term Limit Value (*LTL*): It is the arithmetic average of all measurement results and is the limit value that should not exceed the value specified in Annex-2 of the RCIAP.

The calculations required for the AERMOD model are made according to the guidelines published by the Ministry of Environment, Urbanization and Climate Change. Certain formulas are applied to calculate PM10 (Particulate Matter less than 10 micrometers)

values based on field measurements. These formulas allow the data obtained under real conditions to be converted into standard conditions. The following formulas are utilized in the calculation of PM10 values.

Formulas:

$V_{std} = V_{act} \times (P_{act}/P_{std}) \times (T_{std}/T_{act})$

$V_{act} = 60 \times Q_{act} \times t_{hr}/1000$

$PM_{std} = M_{PM}/ V_{std}$

$PM_{act} = M_{PM}/ V_{act}$

P_{std}: Standard Pressure (760 mmHg)
P_{act}: Actual Pressure (mmHg)
T_{act}: Actual Temperature (Kelvin)
t_{hr}: Measurement Duration (hours)
PM_{std}: Standard PM10 Concentration
PM_{act}: Actual PM10 Concentration
V_{act}: Actual Volume
T_{std}: Standard Temperature (273 Kelvin)
V_{std}: Standard Hacim
M_{PM}: Mass of Particulate Matter Collected on Filter

When the modeling and measurement results in Table 4 are examined, it is seen that PM and settled dust concentrations resulting from production activities are below all limit values. However, in the modeling study, it is understood that there are cases where PM concentrations exceed 50 µg/m³, which is the limit value of the RCIAP. The RCIAP states that PM10 "cannot be exceeded more than 35 times in a year".

As can be seen in Table 3, modeling studies for dust emissions generated outside of blasting operations are controlled dust modeling called "Scenario-1" and uncontrolled dust modeling called "Scenario-2". Based on the controlled dust modeling, the Daily STL PM10 value is 32 µg/m3 and the annual LTL value is 6 µg/m3. However, the Daily STL PM10 value measured in the field is 92.65 µg/m3 and the annual LTL value is 5.33 µg/m3. As a result, when the results measured in the modeling study and in the field are compared, it is observed that the PM10 Annual LTL values are very close to each other and work in harmony. However, it is observed that there is approximately a 3-fold difference in the model and measurement results for the daily PM10 STL value. Similarly, when the results of the controlled dust modeling are compared with the settled dust values, the Monthly Short Term Limit (STL) value is 298 mg/m² and the Annual Long Term Limit (LTL) value is 24 mg/m² per day. In the field measurements, the Monthly STL value was recorded as 287.06 mg/m² and the Annual LTL value as 47.85 mg/m².

Due to the limited number of modeling results and field measurement values, a small variation was added to the measurement data to calculate the estimated correlation coefficient. This triangulation was done in order to enable correlation analysis in a limited data set.

4. Results

The correlation coefficient for PM10 and settled dust data is as follows:

1. PM10 Daily (STLV): r=-1.0r = -1.0r=-1.0r=-1.0
2. PM10 Annual (LTL): r=1.0r = 1.0r=1.0
3. Settled Dust Monthly (STLV): r=1.0r = 1.0r=1.0
4. Settling Dust Annual (LTL): r=-1.0r = -1.0r=-1.0r=-1.0

These results show high positive or negative correlation due to the limited number of data. The negative correlation in PM10 daily and annual values shows that the model results have an inverse relationship with the field data, whereas the positive correlation in PM10 annual and monthly values shows that there is a consistent relationship between modeling and measurement data.

In a study by Demirel et al. (2019), the differences between dust emission modeling and field measurements were discussed [32]. In this study, it was observed that modeling results were in good agreement with field measurements with annual average values, but model results were lower than field values in short-term, daily measurements. Similarly, in this study, deviations between model results and measurements were found in short-term values. In another study by Kocak and Cetin (2020), dust emissions at mining sites were estimated using the AERMOD model [33]. In Kocak and Çetin's study, it was observed that the modeling results did not exceed the daily limit values, but these values were exceeded in field measurements. This is similar to the limitation of modeling in not being able to fully reflect the variability in field conditions, as in the present study. Yılmaz and Kara (2018) compared controlled and uncontrolled dust modeling scenarios [34]. In the controlled scenario, it was reported that short-term PM10 concentrations were reduced below the limit values by measures such as irrigation systems and anti-dust coatings. Similarly, in this study, lower PM10 and settled dust concentrations were obtained in the controlled dust modeling scenario. The findings of Yılmaz

and Kara emphasize the effectiveness of control measures and are consistent with the results of this study.

As a result, when the results of the modeling study and the results measured in the field are compared, it is observed that the monthly CVS values of the precipitated dust are very close to each other and work in harmony. However, it is observed that there is approximately a 2-fold difference in the model and measurement results for the annual settled dust UVS value. At the same time, the theoretical and practical dust emissions caused by production in the mine were compared and it was determined that some values of the results were very close and there were differences in some values.

5. Conclusion and Suggestions

In order to minimize the damages that may be caused by the pollutants and emissions generated during the production phase of the mines to the topography, vegetation, living creatures and habitats, surface waters, groundwater, water resources, settlements, atmosphere, the planning of the mines should be done perfectly and after the production activities are started, they should be inspected according to whether they are working properly according to the planned methods.

In this context, as can be seen in Table 3, when we consider the modeling studies in dust emissions other than blasting operations, the Daily Short-Term Limit Value PM10 value is 32 µg/m3 and the annual Long-Term Limit Value is 6 µg/m3, while the Daily Short-Term Limit Value PM10 value measured in the field is 92.65 µg/m3 and the annual Long-Term Limit Value is 5.33 µg/m3.

Likewise, when controlled dust modeling results are compared with settled dust values, the Monthly Short Term Limit Value is 298 mg/m² and the Annual Long Term Limit Value is 24 mg/m² per day. In the field measurements, the Monthly Short Term Limit Value was 287.06 mg/m² and the Annual Long Term Limit Value was 47.85 mg/m². As a result, when the results of the modeling study and the results measured in the field are compared, it is observed that the Monthly Short-Term Limit Values of the Precipitating dust are very close to each other and work in harmony. However, it is observed that there is approximately a 2-fold difference in the model and measurement results for the annual Long-Term Limit Value.

Considering the results of the research, the wastes generated after the start of the activity should be minimized and disposal should be ensured in accordance with the laws and regulations in force.

In this context, the relevant institutions and organizations should work more sensitively and work to increase awareness by organizing sectoral training and seminars.

It is possible to make predictions based on receptor models for facilities that have not started operations and therefore cannot be measured, and by using the data of dispersion models used for the operating facilities in the region, it is possible to make predictions about the air pollution that will be created together (cumulative effect) by the facilities in the region (active or not yet operational). With the help of modeling, it is possible to detect the pollution coming to the region from long distances. With these features, models are of great importance in determining air quality management and provide significant support to decision-makers.

The differences between the selected models and the fact that the same model produces different predictions under different tuning and validation conditions make the reliability of the results questionable. Guidance in the legal regulations in Annex-2 of the Regulation on Control of Industrial Air Pollution can have a significant impact on this reliability.

In air quality management, model data are very significant sources for competent authorities in determining the extent to which existing conditions and planned new activities will affect air quality. It is very critical to determine the accuracy and reliability of the model data and to determine the representation rates of the pollution that will be caused by the activities in the current area. It is possible that the data to be obtained by determining the boundaries of the models to be selected correctly represent the actual measurement data. The use of model data with low margin of error in the calculation of total air pollution will increase the decision-making speed of the competent authorities and the accuracy of these decisions. For this aim, it is necessary to increase the use of appropriate models, to use the relevant models by accredited organizations and to carry out the necessary studies to increase the precision of the data.

It is estimated that a modeling unit will be established in the Air Quality Assessment Units by the Ministry of Environment, Urbanization and Climate Change, Provincial Directorates of Environment, Urbanization and Climate Change, and a database will be created by evaluating the submitted files down to mining groups, production

methods, capacities, provinces, districts and villages, and the files submitted will be compared in the databases created and definite results will be reached about their accuracy and validity.

If this system is implemented, it will play a major role in the cumulative evaluation of the effects of the facilities on air quality throughout the country and in achieving the target air quality standards in the coming years.

It will also be valid for the dust emission measurement data carried out during the aforementioned study, and in the same way, it will be possible to establish a database and ensure self-control by establishing a laboratory within the Provincial Directorates of Environment, Urbanization and Climate Change or by working in cooperation with the University laboratories in the relevant province to carry out confirmation measurements and to compare the accuracy and reliability of the results.

Currently, the Ministry of Environment, Urbanization and Climate Change and Provincial Directorates play an important role in the planning and implementation of environmental projects. However, the lack of a systematic inventory for monitoring and evaluation of planned projects reduces the effectiveness of projects and makes the reliability of results questionable.

The current situation does not allow for the evaluation of planned projects based on concrete data. This negatively affects the success rate of environmental projects and reduces the effectiveness of future planning. Therefore, the establishment of a project inventory system will provide more accurate and concrete data, which will increase the validity of environmental projects.

For the effective management of environmental projects, the Ministry of Environment, Urbanization and Climate Change and the relevant Provincial Directorates need to ensure the monitoring and evaluation of planned projects by establishing a systematic inventory. This will contribute to increasing environmental sustainability and achieving more effective results. The findings of the present study are in line with the model and measurement discrepancies observed in similar studies in the literature, the negative impacts of uncontrolled dust emissions and the effects of particulate matter such as PM10 on short and long-term limit values. Studies in the literature emphasize that field measurements may not always be in full agreement with theoretical models and the influence of environmental factors. It is also emphasized in the literature that discrepancies in short-term values, variability in field conditions and modeling limits should be taken into account, as in the present study.

Contributions of the Authors

Volkan SEREN: Methodology, Data Collection, Data Analysis, Writing - Review & Editing.
Ahmet KILIÇ: Conceptualization, Methodology, Supervision, Data Analysis, Writing - Review & Final Editing

Acknowledgment

I would like to thank Dr. Rıfat YILDIRIM.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] J. İbadullayeva, K. Jumaniyazova, S. Azimzadeh, S. Canıgür, and F. Esen, “Effects of Environmental Pollution on Human Health,” *Turkish Medical Students Research Journal*, vol. 1, no. 3, pp. 52–58, 2019.

[2] D. Gülcan, “An Ideological Debate on the Role of the State in the Ecological Crisis,” *Journal of International Relations*, vol. 15, no. 59, pp. 49–63, 2018.

[3] A. Celik et al., *SMEs in the Face of New Strategies*. Kahramanmaraş: T.S.O. Publication, 1997.

[4] M. Yücel, “Questioning the Necessity of the EIA-Regulation,” in *Kuşadası Landscape Values*, 2018, pp. 205–224.

[5] UNDP and REC, “A Guide to Strategic Environmental Assessment,” Tbilisi, 2006, pp. 1–7.

[6] OECD, "Applying Strategic Environmental Assessment," OECD Publishing, 2006.

[7] Ministry of Environment and Forestry, "Environmental Impact Assessment," General Directorate of Environmental Impact Assessment and Planning, Accessed: Jul. 28, 2009. [Online]. Available: http://www.cedgm.gov.tr/ced.htm

[8] R. Keleş and C. Hamamcı, *Environmental Policy*, 5th ed. Ankara: İmge Kitabevi, 2005.

[9] Regional Environmental Center (REC), *Sustainable Production and Consumption Publications-II, Temiz Üretim*, 2011.

[10] [Online]. Available: https://www.surdurulebiliruretim.com/madencilikte-sorumlu-uretim-surdurulebilir-gelecek/

[11] S. A. Silvester, I. S. Lowndes, and D. M. Hargreaves, "A computational study of particulate emissions from an open-pit quarry under neutral atmospheric conditions," *Atmospheric Environment*, vol. 43, pp. 6415–6424, 2009.

[12] H. Ceylan, "Marble Mining and Environmental Law," *CBU Soma Vocational School Journal of Technical Sciences*, vol. 1, no. 9, 2008.

[13] Ministry of Energy and Natural Resources, General Directorate of Mining and Petroleum Affairs (MEPEG), 2024. [Online]. Available: https://www.mapeg.gov.tr/Sayfa/Madenistatistik

[14] Y. Cındık and C. Acar, "Rehabilitation of Quarries to Finished Re-Gaining Activity and The Nature," *Artvin Çoruh University Faculty of Forestry Journal*, vol. 11, no. 1, pp. 11–18, 2010.

[15] İ. H. Hatipoğlu, S. K. Kısakürek, H. Doygun, and M. Gözcü, "The Effects of Open-Pit Mining on Natural and Cultural Landscapes," in *National Symposium on Marble and Quarry Repair Techniques*, Sep. 18–20, 2014, Isparta, Turkey.

[16] R. V. Ramani, "Environmental Planning for Surface Mining Coal," in *Environmental Consequences of Energy Production Problems and Prospects*, The Pennsylvania Academy of Science, USA, 1987.

[17] E. Kahraman and A. M. Kılıç, "Occupational Health and Safety Measures in Drilling and Blasting Operations in Mines," in *1st International Mediterranean Science and Engineering Congress*, Oct. 26–28, 2016, Adana, Turkey.

[18] E. Missanjo, E. Ndalama, D. Sikelo, and G. Kamanga-Thole, "Quarry Dust Emission Effects on Tree Species Diversity in Chongoni Forest Reserve and Vegetation Characteristics in Adjacent Villages," 2015.

[19] Anonim, "Regulation on Control of Industrial Air Pollution," *Official Gazette*, Jul. 3, 2009, no. 27277.

[20] K. Alp and O. Tünay, *Air Pollution Control*, Istanbul Chamber of Commerce Publication, Istanbul, 1996, pp. 31–40.

[21] N. Özkurt, "Research on Air Quality Modeling in Rural Areas (Çan-Bayramiç Region Example)," Ph.D. dissertation, GYTE, Institute of Science and Technology, 2011, pp. 36–40.

[22] [Online]. Available: https://cygm.csb.gov.tr/hkd---genel-bilgiler-i-81545

[23] A. Atımtay, "Air Pollution Modeling and Modeling Techniques," *Chamber of Environmental Engineers Training Seminar Notes*, 2011, pp. 48–49.

[24] E. J. G. Silva, T. Tirabassi, M. T. Vilhena, D. Buske, and R. S. Quadros, "Solution of the Puff Model for Pollutant Dispersion in the Atmospheric Boundary Layer by the Giltt Method," in *22nd International Congress of Mechanical Engineering*, Nov. 3–7, 2013, Brazil, pp. 10384–10391.

[25] A. Mori, "Integration of Plume and Puff Diffusion Models/Application of CFD," *Atmospheric Environment*, vol. 34, pp. 45–49, 2000.

[26] EPA, "User's Guide for AERMOD Meteorological Processor (AERMET)," EPA-454/B-19-028, 2019.

[27] "Regulation on Control of Industrial Air Pollution," *Official Gazette*, Jul. 3, 2009, no. 27277. (Amended: Apr. 13, 2012, no. 28263).

[28] British Columbia Ministry of Environment, Environmental Protection Division, "Guidelines for Air Quality Dispersion Modelling in British Columbia." [Online]. Available: http://www.env.bc.ca/air/airquality/pdfs/aq_disp_model_06.pdf. Accessed: Apr. 25, 2008.

[29] [Online]. Available: https://www.aem.com.tr/aermod-ile-hava-kalitesi-modelleme/

[30] N. B. R. Monteito and E. Aparecida da Silva, "Environmental Licensing in Brazilian's Crushed Stone Industries," *Environmental Impact Assessment Review*, vol. 71, pp. 49–59, 2018.

[31] [Online]. Available: https://webdosya.csb.gov.tr/db/ced/icerikler/hava-kalitesi-dagilim-modeli-kilavuzu-firmalar 20180319135859.pdf

[32] B. Demirel and S. Akduman, "AERMOD Modeli ile Maden Ocaklarındaki Toz Emisyonlarının Değerlendirilmesi," *Hava Kirliliği Kontrolü Dergisi*, vol. 45, no. 3, pp. 123–132, 2019.

[33] M. Koçak and D. Çetin, "Maden Saha Emisyonlarının Modelleme ve Ölçüm Sonuçlarının Karşılaştırılması," *Çevre Bilimleri Dergisi*, vol. 12, no. 1, pp. 56–67, 2020.

[34] S. Yılmaz and A. Kara, "Kontrollü ve Kontrolsüz Toz Modellemesi ile Toz Emisyonlarının Karşılaştırılması," *Maden Mühendisliği Dergisi*, vol. 24, no. 2, pp. 95–105, 2018.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1216-1232 YEAR: 2024
DOI: 10.17798/bitlisfen.1542941

A Novel Approach for Arrhythmia Classification Using CI- 1D-LBP with LSTM, 1D-CNN and GRU Models

Hazret TEKİN[1*], Yılmaz KAYA[2]

[1] *Batman University, Department of Electrical and Electronics Engineering, Batman, Türkiye*
[2] *Batman University, Department of Computer Engineering, Batman, Türkiye*
(ORCID: 0000-0002-9379-721X) (ORCID: 0000-0001-5167-1101)

Keywords: ECG, Center-Independent One-Dimensional Local Binary Pattern (CI-1D-LBP), 1D-CNN, LSTM, GRU.

Abstract

Atrial arrhythmias (ARR) are known as the most encountered cardiac disorders in today's world that have direct or indirect detrimental effect on human health. Therefore, Computer-Assisted Diagnosis (CAD) systems are instrumental in the early detection and diagnosis of diseases, serving a pivotal role in the initial assessment and identification process. In this study, ECG data belonging to four different types of arrhythmias, namely ventricular beat (VB), supraventricular beat (SVB), fusion beat (FB), and an unidentified arrhythmic beat (UB), as well as ECG data showing normal sinus rhythm (NSR) of healthy individuals were classified. The ECG data were sourced from the MIT-BIH database. The Center-Independent 1-Dimensional Local Binary Pattern (CI-1D-LBP), originated from the local binary pattern (LBP) method, proposed as a new approach for deriving the essential features needed for the classification of ECG signals. With this new approach, histograms are generated from the signals, and these histogram data are used as input for classification in 1D-CNN, LSTM, and GRU deep learning methods. The CI-1D-LBP+GRU methodology exhibited superior efficacy in classifying the five-labeled dataset (VB-SVB-FB-UB-NSR) relative to the other applied methods, attaining an impressive accuracy rate of 98.59%.

1. Introduction

The heart works regularly with a heartbeat (pulse) of 60-100 beats per minute, but it may experience arrhythmia because of the disruption of this natural working cycle [1]. Arrhythmia, also known as cardiac rhythm disorder, is caused by the irregularity of heartbeats. Upon the manifestation of these arrhythmias, the heart may present with tachycardia, denoting an abnormally rapid heartbeat; bradycardia, indicating an unusually slow heartbeat; or an irregular rhythm. [2]. Arrhythmias, which are usually seen in people with chronic heart conditions, can also occur in individuals with no pre-existing heart conditions. Therefore, examinations are performed taking into consideration how the arrhythmia is and whether there is any heart disease causing this arrhythmia in patients. Although some arrhythmias may seem harmless at times, some of them can be serious enough to cause life-threatening situations. For example, during an arrhythmia, the heart might struggle to supply sufficient blood to the body, causing symptoms such as shortness of breath, fainting, loss of consciousness, or even sudden death [3].

The most encountered types of arrhythmias in the medical world are ventricular tachycardia (VT), supraventricular tachycardia (SVT), and fusion arrhythmias that arise from the combination of these two arrhythmias [4]. In ventricular tachycardia, the heart rate can increase up to 200-250 beats per minute [5]. As a result of this rapid beating of the heart, the ventricles, which are known as the main pumps that allow blood to circulate throughout the body, may not fill with enough blood, leading to serious health problems [5]. Supraventricular tachycardia is one of

*Corresponding author: hazrettekin@sirnak.edu.tr
Received: 03.09.2024, Accepted: 02.10.2024

the most common types of arrhythmias in infancy and childhood [6]. Especially in infancy, it can manifest itself with symptoms such as restlessness, weakness, and loss of appetite. In children, chest pain and palpitations are among the most prominent complaints, and fainting may occur, albeit rarely [6]. The simultaneous stimulation of the ventricle by both supraventricular and ventricular effects is called a fusion beat or rhythm [7]. Fusion beat arises from the chance encounter of supraventricular and ventricular stimulation in the ventricle [8].

Timely diagnosis of cardiac conditions, such as arrhythmias, coupled with the swift and effective implementation of suitable treatment modalities, mitigates the health risks associated with these disorders. ECG signals, acquired from biomedical electrodes meticulously located at specific anatomical sites, play an indispensable role in the early diagnosis and strategic management of cardiovascular disorders. ECG is a time-resolved signal that meticulously encodes the cardiac electrical activity throughout its cyclical operational phases [9]. The morphological and quantitative characteristics of ECG signals can elucidate intrinsic pathologies, such as arrhythmias, by revealing subtle symptomatic indicators [10]. Substantial quantities of ECG data are procured from an array of modalities, including stationary and mobile ECG devices, holter monitors, [11]. It is essential to establish sophisticated algorithms for the analysis of ECG signals derived from these diverse recording modalities. Prompt and precise diagnosis of cardiac conditions, such as arrhythmias, is critical for enabling swift and effective therapeutic interventions. Therefore, the development of computer-based new technologies is necessary.

The complex and dynamic nature of ECG signals and their susceptibility to noise require a careful selection of signal analysis methods that can provide optimal results in terms of performance [12]. Consequently, it is imperative to advance the development of robust and practical methodologies for signal analysis, feature derivation, and categorization techniques, which can be seamlessly incorporated into Computer-Aided Diagnosis models. The aim of these diagnostic algorithms is to provide the most accurate and fast results while keeping the computational cost and complexity at the lowest level.

In this study, ECG data belonging to four different arrhythmia types, namely ventricular beats (VB), supraventricular tachycardia (SVT), fusion beats (FB), and an unidentified beat (UB), as well as ECG data showing normal sinus rhythm (NSR), were classified.

The data were sourced from the MIT-BIH Arrhythmia Database. In the context of this study, the novel locally binary pattern-based (LBP) Center-Independent 1D-LBP (CI-1D-LBP) methodology was employed to extract the crucial features necessary for the categorization of ECG signals. The CI-1D-LBP method employs a nuanced approach, utilizing binary data derived from the comparison of each of nine discrete points within a window function applied along the ECG signal's length, relative to their neighboring counterparts. Using this method, signal histograms with values ranging from 0 to 255 was generated, and these histogram data were fed as input data to 1D-CNN, LSTM, and GRU deep learning methods. Each deep learning method's hyperparameters were optimized, and their performance results were compared with each other. The proposed method was tested for algorithmic robustness and accuracy performance using ECG datasets (VB-SVB-FB-UB-NSR). In addition, the proposed method was tested on ECG data with different signal lengths obtained as a result of preprocessing steps in the dataset to evaluate its classification performance rates. This allowed us to observe the effect of the length of ECG data on classification performance. When the average accuracy percentages in the application are examined, it is observed that the hybrid CI-1D-LBP+GRU method performs the best in all applications.

Reviewing the literature on the detection of arrhythmia and other cardiac conditions through ECG analysis, it is evident that the process involves a three-stage procedure at its core. During the preliminary stage, ECG signals are purified from noise and other signals caused by hardware effects and external factors by using various filter algorithms. The second stage involves using various mathematical and statistical feature extraction algorithms to determine the features that will characterize the signal and be used for signal classification in the subsequent stage. In the third stage of the process, the features derived from the signal are utilized for classification using suitable machine learning methods with optimized parameters, and the performance is evaluated accordingly. The main aim and motivation of all these processing stages are to achieve high accuracy for the created model while minimizing computational costs.

Thanapatay and colleagues used a hybrid model to classify NSR and four type of arrhythmias. DWT and PCA were employed to derive features from the ECG signal, while SVM were utilized in the subsequent classification phase. The experiments were conducted on 45,686 beats and a reported precision rate of 97.6367% was achieved.[13].

Karpagachelvi and colleagues aimed to classify six different types of heartbeats. They utilized a total of 40,916 beats and applied Discrete Wavelet Transform for extracting features. As classifiers, they employed SVM and ELM and achieved the highest performance of 90.52% using the PSO-SVM classifier [14].

Vijayavanan and colleagues used the morphological attributes of the ECG waveform to extract features. They performed performance analyses using Probabilistic Neural Networks (PNN) in the classification stage. They used Discrete Wavelet Transform (DWT) in the preprocessing stage. They took training periods of 5, 10, 15, and 20 minutes. Approximately 150 records from each category were used for training and 50 records for testing. They found the accuracy of the system to be 96.5% [15].

In their study, Hadiyoso and Rizal employed the Hjorth Descriptor High-Level Complexity to extract features of the ECG signal. The test data consisted of three types of ECG signals: NSR, ARR and Congestive Heart Failure. KNN and MLP (Multilayer Perceptron) were used to classify the features of the signal. By using this method, both MLP and KNN achieved 94% accuracy [16].

Gutiérrez et al. categorized eight distinct types of ECG recordings. They first applied bandpass filtering to the signal and then used Quadratic (2nd order) Wavelet Transform for feature extraction. Probabilistic Neural Networks were used for classification. The overall classification rate varied between 92.746% and 100% [17].

Wu et al. introduced a new approach derived from DIA and YOK for learning ECG arrhythmia features. Gaussian-Bernoulli and Bernoulli-Bernoulli were used to extract the features of ECG data. To enhance the efficacy of the network, a fine-tuning process using backpropagation was performed by appending a softmax regression layer to the hidden representation layer designed for supervised classification. Experimental results showed that the proposed approach achieved successful performance with 99.5% accuracy [18].

Lynn and colleagues proposed the Bi-GRU (Bidirectional Gated Recurrent Units) method as a recurrent neural network (RNN) structure for detecting different types of arrhythmias. The proposed method was compared to LSTM. It was found that the Bi-GRU model, which performed better than LSTM, had an accuracy percentage of 98.55% [19].

Acharya et al. introduced an 11-layer CNN model for the diagnosis of congestive heart failure (CHF). CNN model streamlines the preparation of ECG signals and obviates the necessity for intricate mathematical algorithms in feature extraction This model also reduces the software processing load. The proposed CNN model achieved a high success rate of 98.97% [20].

Han et al. proposed a hybrid method for arrhythmia detection that combines LSTM and 1D-CNN. The proposed architecture was compared to 1D-CNN and Multi-Layer Perceptron (MLP). The accuracy rates were measured as 92.03%, 90.98%, and 86.15% for the hybrid LSTM-1D-CNN, 1D-CNN, and MLP, respectively [21].

Darmawahyuni and colleagues classified ECG data from patients with congestive heart failure (CHF) and healthy individuals. Initially, they employed the Discrete Wavelet Transform to isolate ECG signals from noise. Subsequently, they segmented denoised ECG signals to enhance classification accuracy. Ultimately, the RNN-LSTM architecture was utilized as a deep learning technique for further analysis. The results demonstrated that the proposed method attained accuracy rate surpassing 99% [22].

Zheng et al. applied a pioneering technique, named as Singular Value Decomposition (SVD), for extracting features from ECG signals. CNN and SVM machine learning methods are employed for classification. Optimal average accuracy for both CNN and SVM models was attained when only 3 singular values were selected, with accuracy rates surpassing 96% [23].

In their study, Cinar and Tuncer proposed a novel hybrid deep learning approach for the classification of ECG signals from patients with congestive heart failure, cardiac arrhythmias, and healthy individuals. The presented hybrid AlexNet-SVM model was utilized to classify ECG signals and its performance was benchmarked against other machine learning techniques, including SVM, KNN, and LSTM. Results revealed that SVM achieved an accuracy of 68.75%, KNN attained 65.63%, and LSTM reached 90%. In contrast, the hybrid AlexNet-SVM model, which utilized spectrogram images derived from ECG signals, achieved a remarkable accuracy rate of 96.77% [24].

Mohonta and colleagues introduced an advanced 2D-CNN framework augmented with CWT for the classification of five distinct arrhythmic signal types. This 2D-CNN-based model was adeptly applied to the classification of brief ECG segments (0.225 seconds) utilizing scalograms derived from the continuous wavelet transform. The proposed methodology achieved outstanding performance metrics, with average sensitivity, specificity, and

accuracy rates of 98.87%, 99.85%, and 99.65%, respectively [25].

Madan et al. proposed a hybrid deep learning model for classifying ECG data associated with atrial fibrillation (ARR), congestive heart failure (CHF), and normal sinus rhythm (NSR). They transformed ECG signals into scalogram images to automate noise filtering and feature extraction. The classification was performed using a hybrid CNN-LSTM model, achieving accuracy rates between 98.7% and 99% [26].

Panganiban et al. developed a CNN-based application for arrhythmia classification using spectrogram images of ECG signals. They utilized Google's Inception V3 model to retrain the final layer for this task, classifying ECG datasets into two labeled classes and five labeled classes for training and testing. Their method achieved significant accuracy rates of 98.73% for two-class classification and 97.33% for five-class classification [27].

Salem et al. proposed a method for analyzing and classifying four types of ECG arrhythmias using spectrograms generated from ECG signals via Short-Time Fourier Transform (STFT). These spectrogram images served as input for a hybrid deep learning model that combined DenseNET with SVM, achieving an accuracy of 97.23%. [28]

Rahul and Sharma proposed a hybrid model consisting of 1-D CNN and Bi-LSTM algorithms for the detection of three different types of arrhythmias. A hybrid model composed of stationary wavelet transforms (SWT) and Savitzky-Golay (SG) median filters, which was proposed as a new method for preprocessing ECG data, was used in the preprocessing stage. It was found that the proposed hybrid 1-D CNN-Bi-LSTM classification model had an accuracy rate of 99.44% [29].

2. Material and Method

2.1. Dataset

To test the proposed approach, ECG data for four different types of arrhythmias, including ventricular beats (VB), supraventricular beats (SVB), fusion beats (FB), and an unspecified beat type (UB), as well as ECG data for normal sinus rhythm (NSR) from healthy individuals, were obtained from the MIT-BIH database [30], [31]. The dataset comprises a total of 113,382 ECG recordings, each comprising 187 samples, including 90,589 recordings of normal sinus rhythm (NSR), 7,236 recordings of ventricular tachycardia (VB), 5,179 recordings of supraventricular tachycardia (SVB), 2,339 recordings of fusion beats (FB), and 8,039 recordings of an unidentified beat type (UB). The sampling frequency of the ECG data is given as 125 Hz. The ECG graphs for NSR, VB, SVB, FB, and UB heartbeats are shown in the Figure 1.

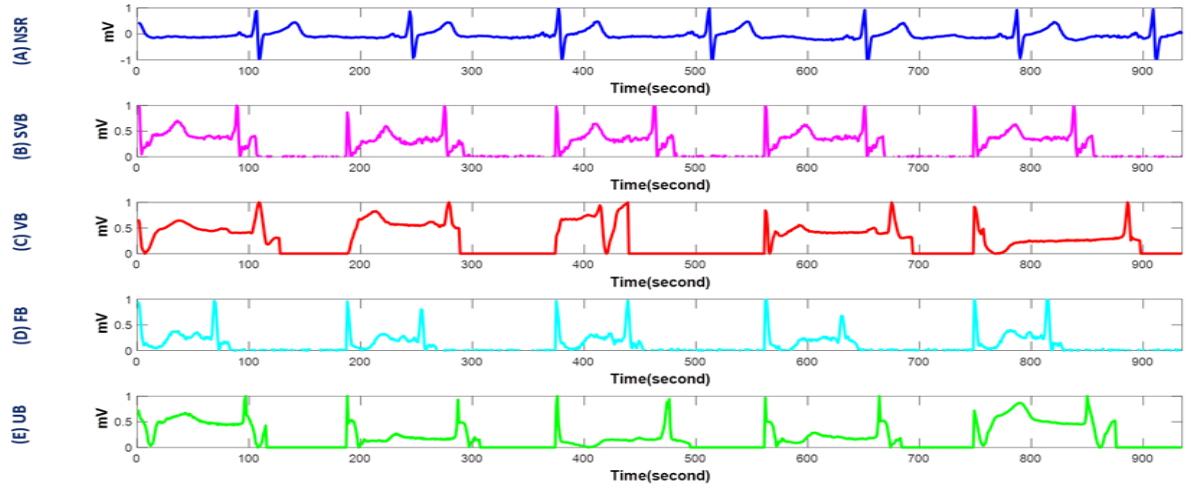


Figure 1. ECG sample signals for (A) NSR, (B) SVB, (C) VB (D) FB, (E) UB

2.2. Proposed Arrhythmia Diagnosis System

The suggested method for extracting features and classifying ECG signals to identify the type of arrhythmia involves four stages as illustrated in Figure 2. Each stage includes a brief explanation of the operations carried out.

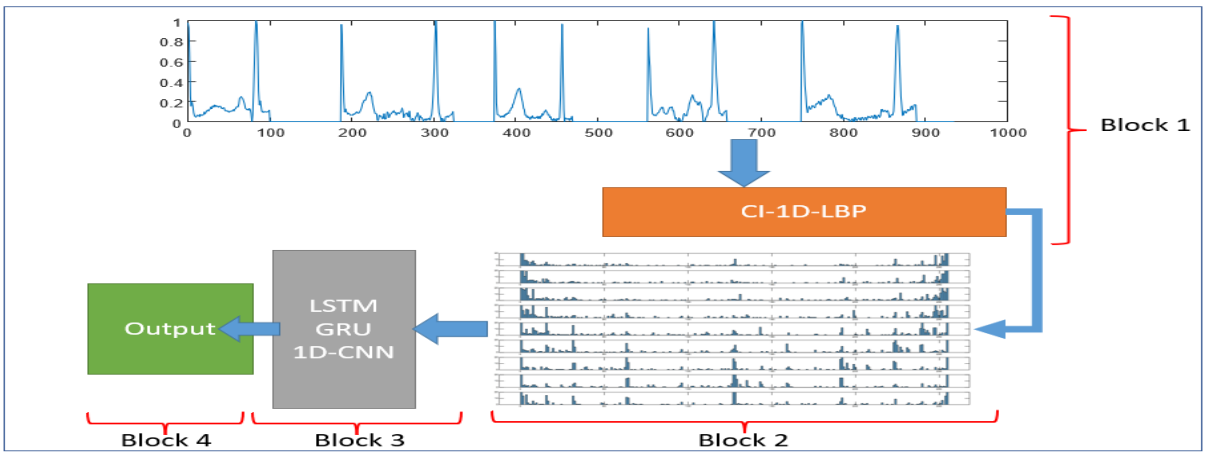


Figure 2. The flow diagram of the proposed arrhythmia classification system

Block 1: In this phase, the CI-1D-LBP method was employed to every individual ECG signal sample within the dataset, acting as a signal transformation procedure. Upon feeding a signal into the CI-1D-LBP method, nine distinct signals are generated as output, where the values of the produced signals exhibit variations between 0 and 255. Figure 3 shows nine different signals obtained by applying the CI-1D-LBP method to an example ECG signal.

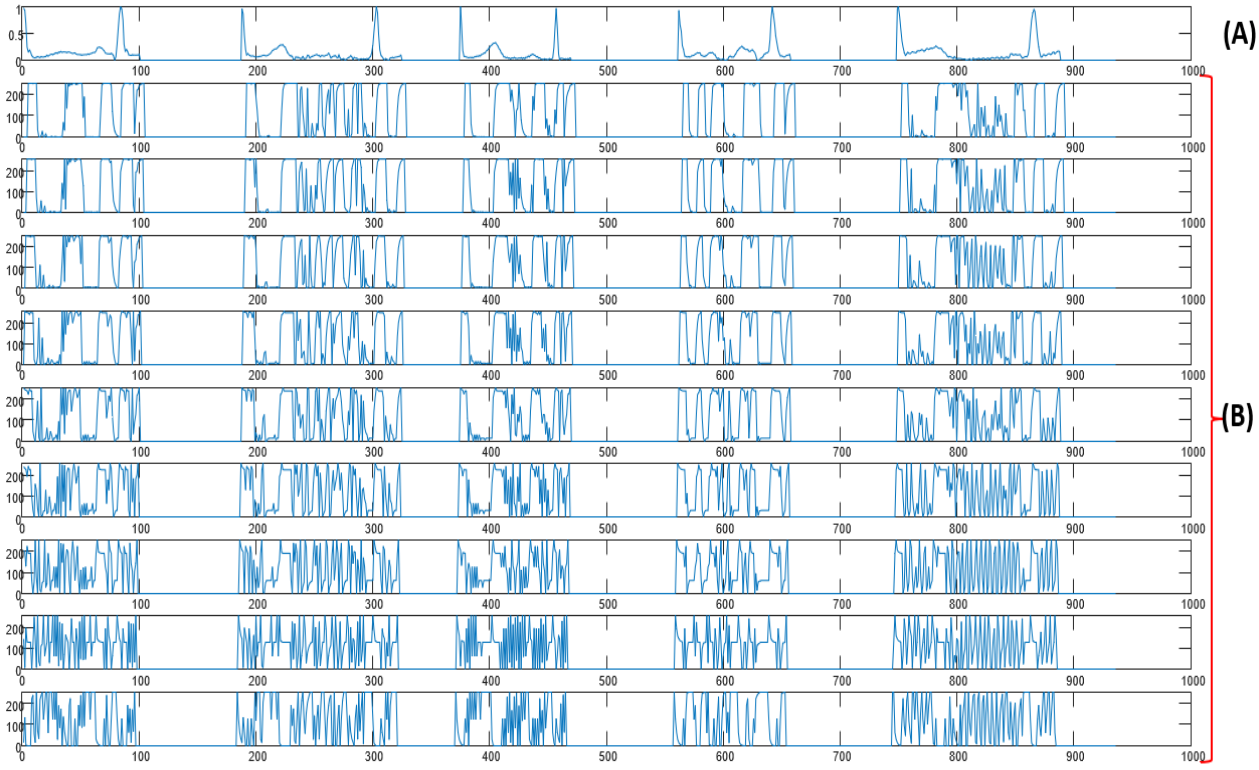


Figure 3. Implementation of the CI-1D-LBP method to a ECG sample. (A) Sample ECG signal, (B) Produced signals

Block 2: In this step, histograms of the signals generated through the implementation of the CI-1D-LBP method. The histograms corresponding to the nine different signals produced by the CI-1D-LBP method are utilized as input to the LSTM, 1D-CNN, and GRU models. The histograms of the newly generated signals also exhibit dissimilarities. Figure 4 displays the histograms corresponding to the nine different signals produced for example ECG signal, highlighting the variations between them. As demonstrated, these dissimilarities contribute to the creation of feature matrices characterizing each ECG signal instance within the dataset.

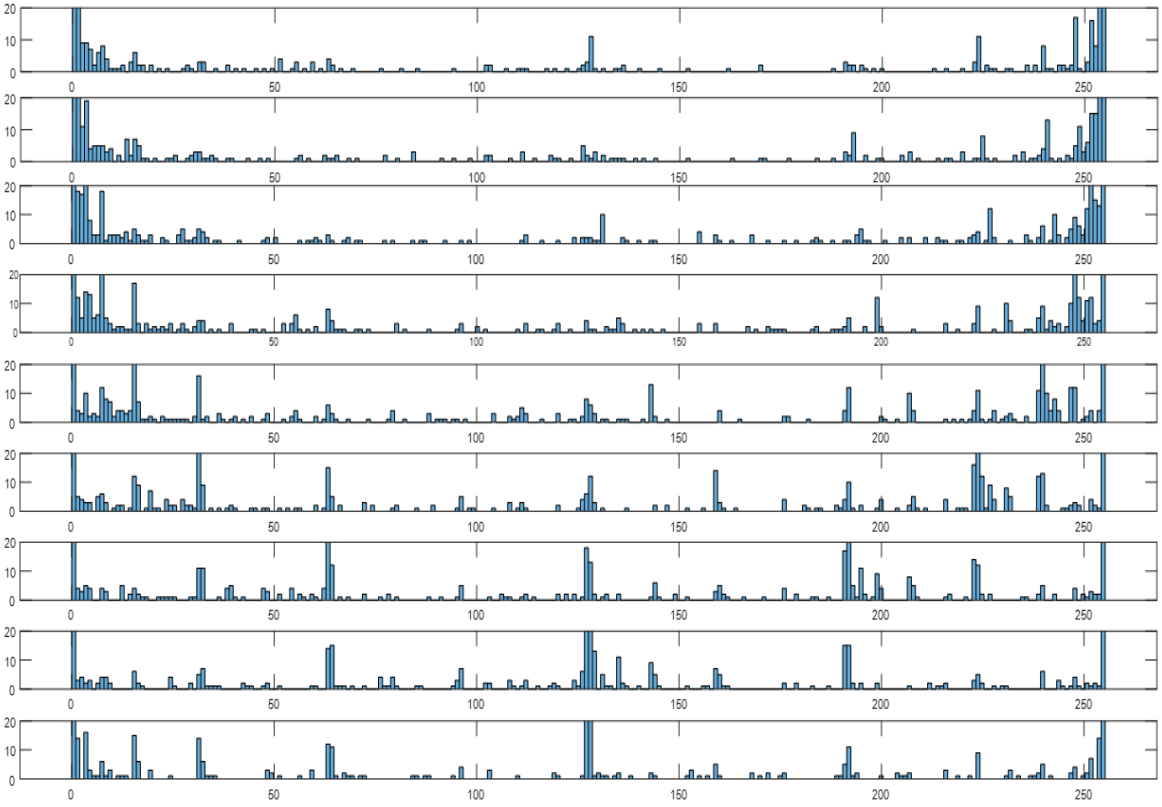


Figure 4: Histograms of the newly generated signals via CI-1D-LBP

Block 3: In this phase, the CI-1D-LBP histograms derived from the preceding stage are input into LSTM, GRU, and 1D-CNN deep learning architectures.

Block 4: This stage pertains to the decision-making process, where the classification accuracy percentage, performance parameters and confusion matrices are presented as means of evaluating the classification performance.

2.3. Center independent 1D local binary patterns (CI-1D-LBP)

The CI-1D-LBP method employs a nuanced approach, utilizing binary data derived from the comparison of each of nine discrete points within a window function applied along the ECG signal's length, relative to their neighboring counterparts. Figure 5 provides a detailed illustration of the CI-1D-LBP method, which operates by traversing the signal pointwise through a window of size WS (Window Size) = 9.

The window length of 9 was chosen in the application of the method to ensure that each selected central point on the signal could be compared with its eight neighboring points, resulting in the formation of 8-bit binary sequences. In the proposed method, the value of nine is defined as the window length(WS) within which comparisons are made. Specifically, nine points are selected within the window, and each point is compared to its eight neighboring points. This window size is critical for balancing feature richness and computational efficiency. The eight comparisons result in an 8-bit feature vector, where each bit reflects whether a particular point in the window is greater or smaller than its neighboring points.

Choosing a window length (WS) of nine ensures sufficient local detail is captured while avoiding unnecessary computational complexity. A smaller window would reduce the granularity of the features extracted, while a larger window would increase the number of comparisons, leading to higher computational costs without significant gains in feature representation. Thus, a window length of nine provides an optimal trade-off between capturing local variations in the signal and maintaining computational efficiency. These binary sequences were then used to generate histograms from the ECG samples, with values ranging between 0 and 255.At any given time, the values that enter the WS=9 window are represented as a vector P={Pc,P1,P2,P3,P4,P5,P6,P7,P8}, where Pc is the central value and P1-P8 are the surrounding values.

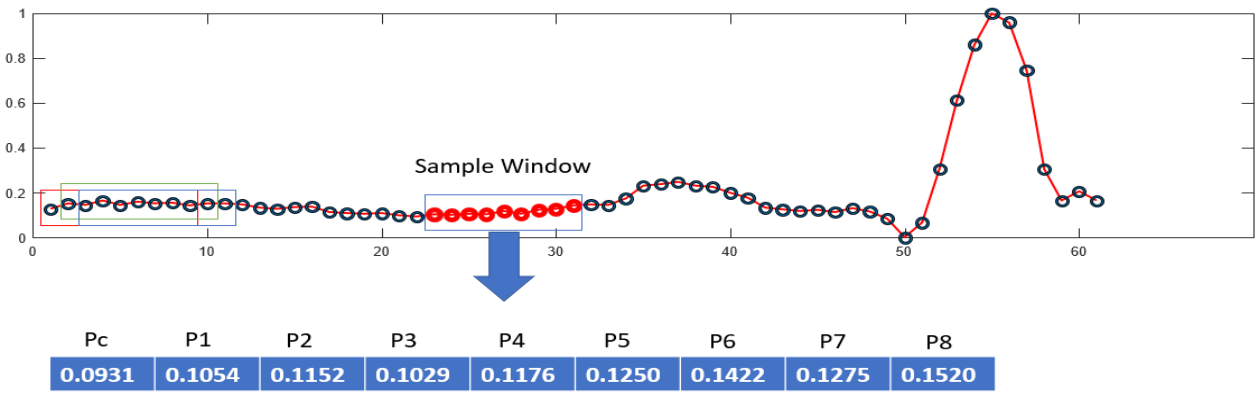


Figure 5. Application of CI-1D-LBP approach to an example ECG signal segment.

The WS window includes 9 values, one of which is Pc (Point Center), and the others are considered as neighboring P={P1,P2,P3,P4,P5,P6,P7,P8} values. The Pc value represents each value on the vector entering the WS window. As shown in Figure 6, there are 9 possible cases within the WS window. In this case, the Pc value is subjected to binary comparisons with the other values using the following equation.

$$\begin{Bmatrix} Pc > Pi \;\; 1, \;\; Pi = \{P1, P2, P3, P4, P5, P6, P7, P8\} \\ Pc \leq Pi \;\; 0, \;\; Pi = \{P1, P2, P3, P4, P5, P6, P7, P8\} \end{Bmatrix} \quad (1)$$

As shown in Figure 6, each of the nine points within the window is sequentially designated as the central point. A difference comparison is then made between the central point and the other eight points within the same window, generating an 8-bit binary value for each point. Consequently, these 8-bit sequences are used to create histograms, characterizing the signal with amplitudes ranging from 0 to 255.

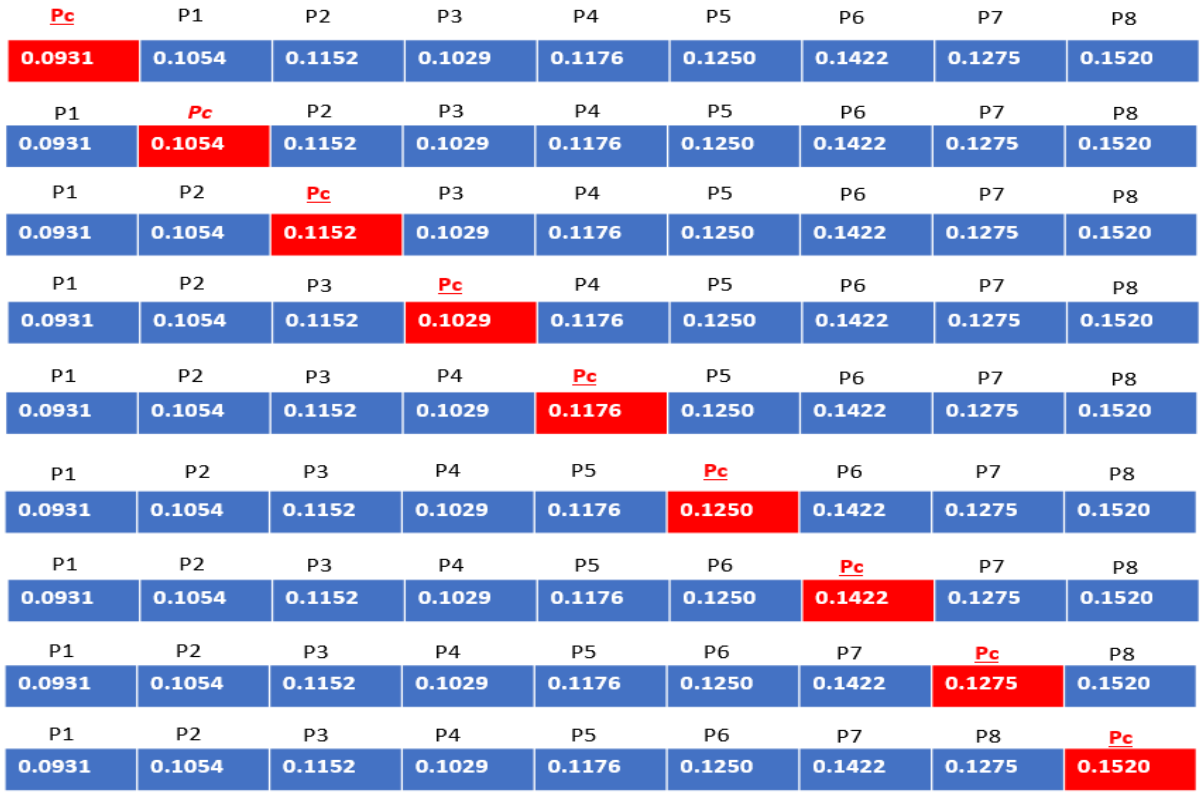


Figure 6: Computation of CI-1D-LBP value according to possible PC values

2.4. 1D-CNN

CNNs are widely employed in areas such as pattern recognition, image interpretation, medical image analysis, and natural language processing [32]. These networks can be transformed into different architectures with changes in network structures and hyperparameters depending on the problem and application [33]. In applications, there are two types of convolutional neural network models, namely 1D-CNN and 2D-CNN. Structurally, the 1D-CNN parallels the 2D-CNN network. 2D-CNNs are chiefly employed for the feature extraction and classification of two-dimensional images, whereas 1D-CNNs are specially engineered for handling the extraction and classification of one-dimensional signals, including ECG, EEG, and audio data [34]. The schematic representation of a conventional 1D-CNN network is shown Figure 7.

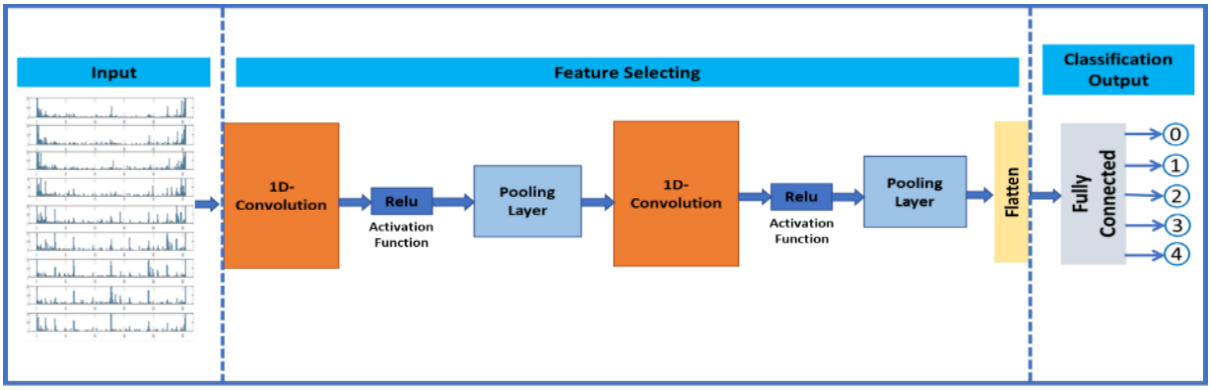


Figure 7. Block diagram structure of a 1D-CNN network

2.5. Long Short-Term Memory (LSTM)

LSTM is a unique form of Recurrent Neural Network (RNN) model specifically designed to address the RNN's shortcomings in learning long-term dependencies [35]. Employed in classification and regression applications, the LSTM architecture offers a solution to the RNN's limitations in capturing long-term dependencies.

The most significant advantage of LSTM lies in its internal sections, commonly referred to as "gates," that enable it to control the data flow throughout the network. These gates consist of four elements: the input gate, forget gate, change gate, and output gate. The gates have the unique ability to coordinate information flow both into and out of the cell, allowing for the addition or removal of information, thus endowing the cell with the capacity to remember values over designated time intervals [36]. The input gate controls when new information should enter the memory. The forget/memory gate regulates the forgetting of existing information and the remembering of new data [37]. The output section controls when the information in the cell should be used at the output of the cell [38]. The block diagram of the LSTM architecture is illustrated below.

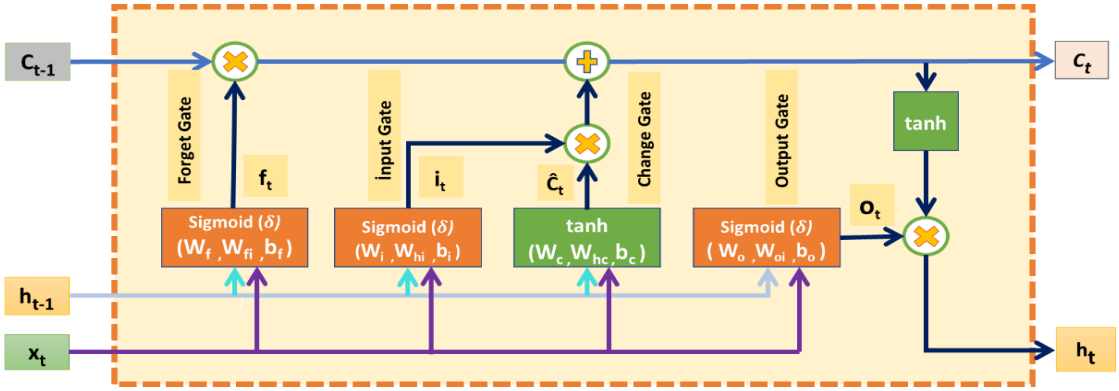


Figure 8. Representation of LSTM blocks and memory cell units.

The LSTM cell contains two states, the long-term (Ct) cell state that represents the intermediate output, and the short-term (ht) cell state. Additionally, LSTM cells contain four trainable gates: input, output, forget, and change gates [40]. These gates have the ability to control the flow of information into and out of the cell by regulating the addition or removal of information, allowing the cell to remember values at certain intervals [34]. The gates are different neural networks that decide which information will be allowed into the cell state. Moreover, during the training process, gates learn which information should be stored or forgotten in the cell state [40]. Let $x = \{x_1, x_2, \ldots, x_n\}$ be the input data at time t. The memory cell "Ct" updates the information using the input (i_t), forget (f_t) and change ($\check{C}_t$) gates. At any time "t", the corresponding gates and layers compute the following functions:

$$i_t = \sigma(W_i x_t + W_{hi} h_{t-1} + b_i) \quad (2)$$

$$f_t = \sigma(W_f x_t + W_{fi} h_{t-1} + b_f) \quad (3)$$

$$o_t = \sigma(W_o x_t + W_{oi} h_{t-1} + b_o) \quad (4)$$

$$\check{C}_t = tanh(W_c x_t + W_{hc} h_{t-1} + b_c) \quad (5)$$

$$C_t = f_t \odot C_t + i_t \odot \check{C}_t \quad (6)$$

$$h_t = o_t \odot \tanh(C_t) \quad (7)$$

Where σ and tanh are sigmoid and hyperbolic tangent activation functions. W_{hc}, W_{hi}, W_f, W_{fi}, W_o, W_{oi}, W_c, W_i are the weight matrices, b_i, b_f, b_o, b_c are the vector of threshold values. ⊙ denotes matrix or vector products. The cell state serves as a transport highway that carries relevant information and the information flow is carefully regulated by the gates. The gates are different neural networks that decide which information is allowed to enter the cell state. Additionally, during training, the gates learn which information should be stored or forgotten. The input gate is a section where a neural network generates the new memory with the help of the tanh activation function and the effect of the previous memory block. The output gate is the section where the output of the current LSTM block is produced.

The LSTM algorithm stands out with its ability to automatically extract features from time series data and learn complex nonlinear relationships [37]. Due to these properties, the LSTM algorithm is remarkable for its ability to automatically extract features from time-series data and learn complex nonlinear relationships. Therefore, the LSTM model is frequently utilized for the extraction and categorization of features from nonlinear signals, such as ECG, EMG, and EEG.

2.6. GRU (Gated Recurrent Units)

The GRU algorithm is an enhanced type of recurrent neural network (RNNs). GRU models, designed by Chu et al. (2014), are created by modifying the LSTM algorithm to mitigate the vanishing gradient problem, which impedes the learning of long-term dependencies. [38]. They use an internal memory feature and update and reset gates to store and filter information. [19]. GRU networks use gate units similar to LSTM networks to model the data flow within the unit, but without a separate memory cell, and they also have the ability to capture long-term dependencies in sequential input data or keep sequential input data in memory without deleting them for a long time, similar to LSTM networks. [41]. In the GRU model, the input and forget gates present in LSTM networks are merged into a single gate called the update gate, and this simpler design of GRU speeds up its training with fewer parameters. The GRU architecture is composed of three gates: the reset gate, the update gate, and the gate that computes the representative output vector. The reset gate governs the extent to which the previous state is to be disregarded, while the update gate determines the proportion of the previous state that should be retained, and the representative output vector gate computes the output representation. The block diagram of the GRU architecture is illustrated below.

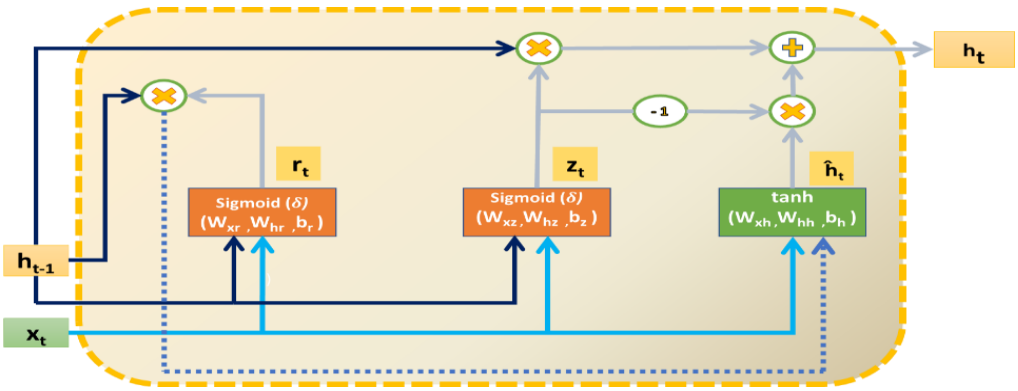


Figure 9. Structures of GRU blocks and cells

As shown in the figure, the network is fed by the current input vector x_t, while the previous layer values are expressed as h_{t-1}. W_{xr}, W_{xz}, W_{xh} ve W_{hr}, W_{hz}, W_{hh} are weight matrices. b_r, b_z, b_h are threshold value vectors. The reset gate is denoted by r_t and the update gate by z_t. In the GRU unit, instead of directly calculating the output vector h_t, a representative $\hat{h}_t$ is first calculated for it. The following equations give the mathematical expressions for the reset gate r_t, the update gate z_t, the representative output gate $\hat{h}_t$ and the output vector h_t. In the equations, the symbol $\odot$ denotes matrix or vector products.

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r) \quad (8)$$

$$z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z) \quad (9)$$

$$\hat{h}_t = tanh(W_{xh}x_t + W_{hh}(h_{t-1} \odot r_t) + b_h) \quad (10)$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \hat{h}_t \quad (11)$$

2.7. Performance Criteria

The accuracy, precision, recall, and f-measure metrics were used to test the performance of the proposed method. The formulas for these metrics are given below.

$$Accuracy = \frac{TP+TN}{TP + TN + FP + FN} x100 \quad (12)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (13)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (14)$$

$$F1 - score = 2x \frac{(Recall\ x\ Precision)}{Recall + Precision} \quad (15)$$

T, F, P, and N represent true, false, positive, and negative, respectively, in these equations. For example, TP represents the number of true positive signals classified, while FN represents the number of false negative signals classified.

3. Results and Discussion

In this study, the CI-1D-LBP methodology was utilized to derive unique features from ECG signals representing four distinct arrhythmia types—ventricular beats (VB), supraventricular beats (SVB), fusion beats (FB), and an indeterminate arrhythmia (UB)—as well as from ECG signals indicative of NSR in healthy subjects. This approach employs binary data derived from the comparative analysis of each of nine discrete points within a window function, applied along the length of the ECG signal, in relation to its neighboring points. Signal histograms with values ranging from 0 to 255, obtained through the aforementioned method, were constructed and utilized as inputs for LSTM, 1D-CNN, and GRU. To evaluate the effectiveness of the suggested methodology, ECG data composed of four different types of arrhythmias (VB, SVB, FB, UB) and normal sinus rhythm (NSR) were used. The hyperparameter settings for the LSTM, 1D-CNN, and GRU employed in the experiment is shown in Table 1.

Table 1 . Hyperparameter values of LSTM, 1D-CNN and GRU models

LSTM		**1D-CNN**		GRU	
Parameter	**Value**	**Parameter**	**Value**	**Parameter**	**Value**
Layers	SeguenceInput LSTM(100) fullyConnected SoftMax Classification	**Layers**	SeguenceInput convolution1dLayer reluLayer layerNormalizationLayer Convolution1dLayer reluLayer layerNormalizationLayer globalAveragePooling1dLayer fullyConnectedLayer softmaxLayer ClassificationLayer	**Layers**	sequenceInputLayer gruLayer fullyConnectedLayer softmaxLayer classificationLayer
Input size	3/256	Input size	3/256	Input size	3/256
Hidden size	100	#filters /filtersize	32-64 / 3	#filters /filtersize	32-64 / 3
Output layer size	3	Output layer size	3	Output layer size	3
Batch size / MaxEpochs	127 / 30	Batch size	127	Batch size	127
Optimizer	Adam	Optimizer	sgdm	Optimizer	rmsprop
Learning rate	0.01	MaxEpochs	70	MaxEpochs	30

The parameters given in Table 1 were fine-tuned to maximize accuracy for each deep learning model. In this study, it was established that a batch size of 127 is optimal for the LSTM, 1D-CNN, and GRU models to attain the most effective classification performance. While the MaxEpoch parameter value was set to 30 for the LSTM and GRU models, it was set to 70 for the 1D-CNN model. The accuracy rates achieved by LSTM, 1D-CNN, and GRU in the classification process for the five-labeled data set (VB-SVB-FB-UB-NSR), are shown in Table 2. The classification tasks were executed for varying training-test ratios.

Table 2. Success rates for 1D-CNN, LSTM AND GRU models

Model	Training-Test	Accuracy (%)	Recall	Precision	F1-score
CI-1D-LBP+1D CNN	70%-30%	97.17	0.953	0.952	0.952
	60%-40%	95.92	0.915	0.893	0.901
	50%-50%	95,48	0.903	0.868	0.892
CI-1D-LBP+LSTM	70%-30%	98.02	0.973	0.941	0.956
	60%-40%	96.21	0.964	0.880	0.917
	50%-50%	95.87	0.895	0.918	0.904
CI-1D-LBP+GRU	70%-30%	98.59	0.975	0.945	0.958
	60%-40%	96.25	0.933	0.904	0.918
	50%-50%	95.81	0.920	0.887	0.898

According to Table 2, high accuracies were observed for all three deep learning methods. However, in general, models with cellular network structures similar to CI-1D-LBP+LSTM and CI-1D-LBP+GRU achieved more successful results than 1D-CNNThe highest level of effectiveness was noted with the CI-1D-LBP+LSTM model as 98.02% for the 70-30 training-test set ratio. In the CI-1D-LBP-GRU model, a success rate of 98.59% was achieved again for the 70-30 training-test ratio. In the CI-1D-LBP+1D-CNN model, a success rate of 97.17% was observed for the 70-30 training-test set ratio. According to the results, the CI-1D-LBP+GRU model has shown the best performance in classifying the five-label ECG data. Confusion matrices of the most successful results of CI-1D-LBP+LSTM, CI-1D-LBP+1D-CNN, and CI-1D-LBP+GRU deep learning models in classifying the five-label ECG data are provided in Figures 10-12. The confusion matrices for both the training and test sets are presented separately. As seen from the confusion matrices, the success percentages of CI-1D-LBP+GRU for each ECG class were higher than CI-1D-LBP+LSTM and CI-1D-LBP-1D+CNN models for both training and test sets.

Figure 10. Confusion matrices obtained with the 1D-CNN model for the training and test sets.

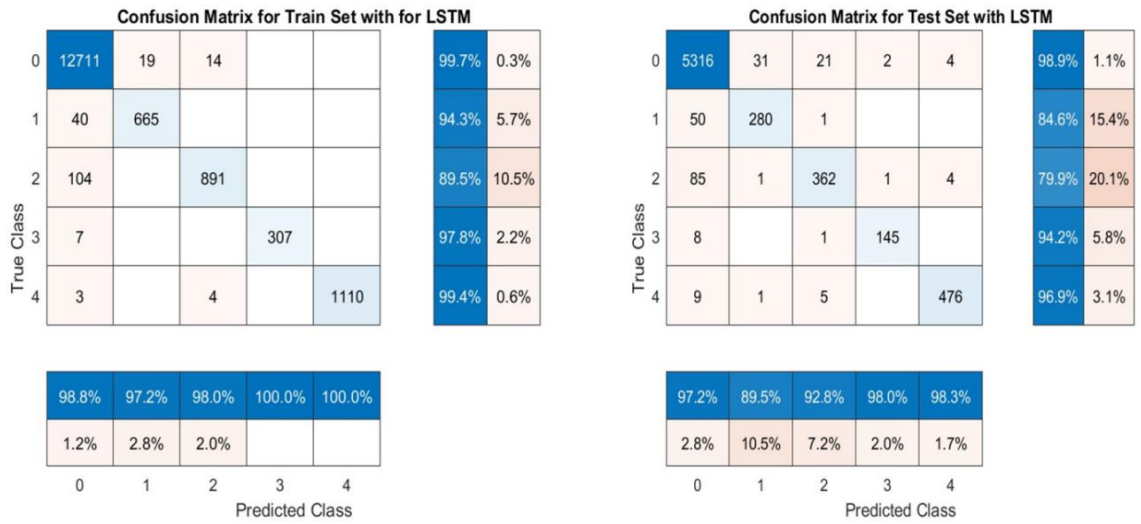


Figure 11. Confusion matrices obtained with the LSTM model for the training and test sets.

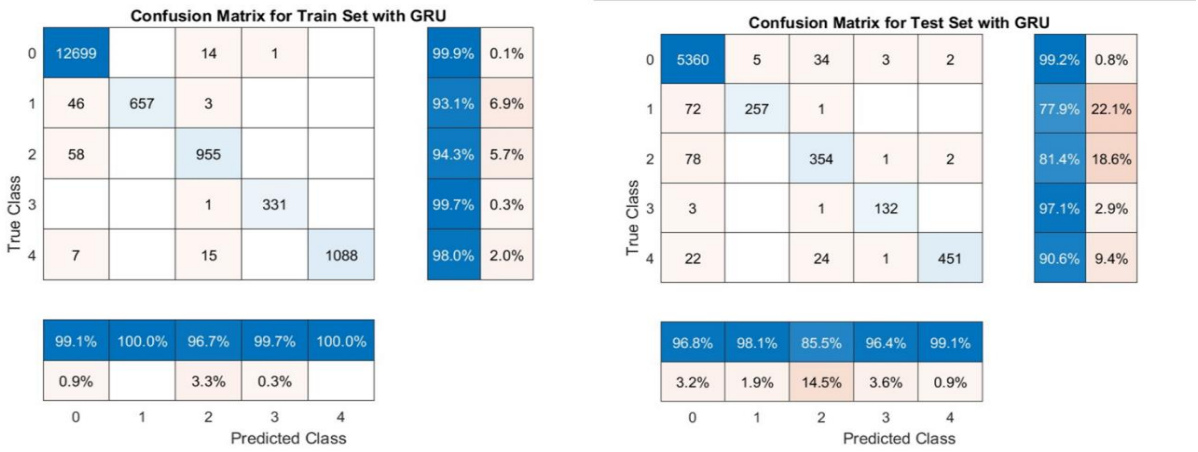


Figure 12. Confusion matrices obtained with the GRU model for the training and test sets

In the application, data sets were used in different combinations and ECG signals belonging to arrhythmia types (VB, SVB, FB, UB) consisting of four labeled data were used to classify arrhythmia types within themselves. Finally, two-labeled datasets consisting of ECG signals from each type of arrhythmia as well as normal sinus rhythm (NSR) were classified separately (i.e., VB-NSR, SVB-NSR, FB-NSR, and UB-NSR). The achieved performance results are presented in Table 3.

Table 3. Classification success (accuracy) rates for different dataset combinations

Datasets	CI-1D LBP+LSTM	CI-1D-LBP-1D+CNN	CI-1D-LBP+GRU
VB-SVB-FB-UB	98.70	97.81	99.08
NSR-VB	99.12	98.23	99.87
NSR-SVB	99.20	98.19	100
NSR-FB	99.47	98.48	99.71
NSR-UB	99.66	98.32	100

It can be observed that the CI-1D-LBP+GRU and CI-1D-LBP+LSTM models outperform the CI-1D-LBP+1D-CNN model in terms of accuracy. Moreover, the CI-1D-LBP+GRU model exhibits the best performance in all two-labeled and four-labeled ECG dataset classifications. The classification performance of ECG data with different signal lengths was tested thanks to the preprocessing steps applied to the dataset. Thus, the effect of ECG data length on classification performance was observed. Upon examination of the results, it can be observed that the classification performance and accuracy of

the proposed models increases as the length of the signal increases. Table 4 presents the classification accuracy rates for the five labeled datasets consisting of NSR and four different arrhythmia types (NSR, VB, SVB, FB, UB) with varying signal lengths.

Table 4. Performance rates (%) according to different signal lengths.

Signal length(Samples)	CI-1D-LBP-LSTM	CI-1D-LBP-1D CNN	CI-1D-LBP-GRU
187	91.20	89.10	91.89
374	92.11	91.32	92.81
561	94.05	93.14	94.76
748	95.24	94.18	95.83
935	96.31	95.91	96.81
1122	97.04	96.33	97.60
1309	98.02	97.17	98.59

The following Table 5 presents the accuracy percentages of the deep learning methods used in previous studies on arrhythmia classification in the literature and the accuracy percentages of this study. Upon analysis of the results, it has been observed that proposed CI-1D-LBP+GRU method, which yielded better results compared to CI-1D-LBP+LSTM and CI-1D-LBP+1D-CNN methods in this study, also outperformed the deep learning methods used in the literature.

Table 5. Evaluating the CI-1D-LBP+GRU in relation to findings from previous research.

Referance	Datasets	Methods	Number of Labeled class in ECG dataset	Performance (Accuracy)(%)
Zheng et al	Three Types ARR and NSR	Singular Value Decomposition (SVD)-CNN-SVM	4	96-97
Murawwat et al	Two types of arrhythmias: Tachycardia and Bradycardia	Multivariate Empirical Mode Decomposition (MEMD)-ANN	3	89.8
Li et al.	NSR and CHF	Distance distribution matrics + CNN	2	81.85
Sharma et al	Four Type ARR, NSR	Fourier-Bessel expansion + LSTM	5	90.07
Çınar and Tuncer	Congestive Heart Failure (CHF), ARR, NSR	Hybrid Alexnet-SVM	3	96.77
Kaouter et al.	Congestive Heart Failure (CHF), ARR, NSR	2D-CNN	3	93.75
Acharya et al	Four Type ARR and NSR	9-layer deep convolutional neural network (CNN)	5	93.50
Kumari et al.	Four Type ARR and NSR	DWT + SVM	5	95.92
This Study	**Four Type ARR and NSR**	**CI-1D-LBP+GRU**	**5**	**98.59**

The Table 5 presents a comparative analysis of various ECG classification methods, highlighting the number of labeled classes and their corresponding performance in terms of accuracy. Methods such as SVD-CNN-SVM, Hybrid AlexNet-SVM, and DWT-SVM show strong results across different datasets, with accuracies ranging from 89.8% to 96.77%. However, the proposed CI-1D-LBP + GRU method demonstrates superior performance, achieving the highest accuracy among the compared approaches. By leveraging the strengths of Local Binary Patterns (LBP) for feature extraction and GRU for modeling temporal dependencies, this study's method offers a more robust and precise classification framework, effectively surpassing traditional and deep learning-based methods.

4. Conclusion

ECGs are extensively utilized for the prompt and precise examination of cardiovascular conditions. In this study, the CI-1D-LBP methodology was introduced for deriving features from ECG signals. When applied to an ECG signal, this method produces 9 different signals. The histograms of these signals, which have values ranging from 0 to 255, were used as inputs to 1D-CNN, LSTM, and GRU deep learning methods.

Computer-assisted diagnostic systems depend on the efficacy of features extracted from ECG signals to ensure accurate heart disease diagnosis. The LSTM, 1D-CNN, and GRU deep learning architectures have been identified as particularly effective classification models, owing to their proven success in analyzing time series data. The ability of deep learning methods to perform feature extraction and classification of data within a single model structure is particularly useful for analyzing high-dimensional, non-stationary data such as ECG.

The CI-1D-LBP+GRU method applied in this study has superior features due to its capacity to model extended temporal dependencies in time series data and retain data in memory for a long time without discarding, which is especially advantageous for the analysis of high-dimensional data such as ECG. Moreover, the CI-1D-LBP+GRU methodology exhibits enhanced capabilities due to its proficiency in identifying and maintaining prolonged temporal dependencies in time series data and preserving data within memory for extended periods without degradation. This attribute proves especially beneficial for the analysis of non-stationary and high-dimensional datasets, such as ECG signals.

The comparison of various ECG classification methods highlights the unique advantages and originality of the proposed CI-1D-LBP+GRU approach in this study. Unlike previous works that rely on conventional feature extraction techniques or deep learning architectures, this method introduces a novel combination of 1D Local Binary Patterns (LBP) for capturing intricate local signal variations, paired with GRU for effectively modeling temporal dependencies. This dual approach allows for a more refined feature representation, enabling superior classification performance across different types of arrhythmias. While several established methods have shown promising results, the integration of LBP with recurrent neural networks offers a fresh perspective that addresses both the spatial and temporal complexities inherent in ECG signals, demonstrating the potential for more accurate and reliable diagnosis in practical applications.

In addition, the CI-1D-LBP+GRU method has lower cellular network complexity compared to the CI-1D-LBP+LSTM model, which has a similar network structure. These superior features of the CI-1D-LBP+GRU have also been demonstrated in the application results. The CI-1D-LBP+GRU method outperformed CI-1D-LBP+LSTM and CI-1D-LBP-1D+CNN in the classification of the five-labeled data set (VB-SVB-FB-UB-NSR), and an accuracy rate of 98.59% was achieved.

Building on the success of the CI-1D-LBP+GRU methodology, future research could explore its application to larger and more diverse ECG datasets to further validate its robustness across different populations and conditions. Additionally, integrating the method with real-time monitoring systems could facilitate early detection of arrhythmias in clinical settings. Enhancing the model by incorporating advanced techniques such as attention mechanisms or hybrid deep learning frameworks could also improve interpretability and performance. Furthermore, expanding this approach to detect and classify additional cardiac anomalies would broaden its potential impact in the field of healthcare diagnostics.

Author Contribution Statement

Conceptualization: HT
Method: YK
Manuscript writing reviewing editing: YK, HT
Supervision: HT

Conflict of Interest

The authors declare no competing interest.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] H. E. Fürniss and B. Stiller, "Arrhythmic risk during pregnancy in patients with congenital heart disease.," *Herzschrittmacherther Elektrophysiol*, vol. 32, no. 2, pp. 174–179, 2021.

[2] F. A. Elhaj, N. Salim, A. R. Harris, T. T. Swee, and T. Ahmed, "Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals," *Comput Methods Programs Biomed*, vol. 127, pp. 52–63, 2016.

[3] H. V Huikuri, A. Castellanos, and R. J. Myerburg, "Sudden death due to cardiac arrhythmias," *New England Journal of Medicine*, vol. 345, no. 20, pp. 1473–1482, 2001.

[4] X. Xu, S. Jeong, and J. Li, "Interpretation of electrocardiogram (ECG) rhythm by combined CNN and BiLSTM," *Ieee Access*, vol. 8, pp. 125380–125388, 2020.

[5] F. Miao, B. Zhou, Z. Liu, B. Wen, Y. Li, and M. Tang, "Using noninvasive adjusted pulse transit time for tracking beat-to-beat systolic blood pressure during ventricular arrhythmia," *Hypertension Research*, vol. 45, no. 3, pp. 424–435, 2022.

[6] M. J. Curtis *et al.*, "The Lambeth Conventions (II): guidelines for the study of animal and human ventricular and supraventricular arrhythmias," *Pharmacol Ther*, vol. 139, no. 2, pp. 213–248, 2013.

[7] M. A. Arias, M. Pachón, and C. Martín-Sierra, "A regular wide QRS complex tachycardia with fusion beats?," *J Arrhythm*, vol. 36, no. 6, p. 1100, 2020.

[8] S. Ayub and J. P. Saini, "ECG classification and abnormality detection using cascade forward neural network," *International Journal of Engineering, Science and Technology*, vol. 3, no. 3, 2011.

[9] A. Çalışkan, "A new ensemble approach for congestive heart failure and arrhythmia classification using shifted one-dimensional local binary patterns with long short-term memory," *Comput J*, vol. 65, no. 9, pp. 2535–2546, 2022.

[10] S. Sahoo, M. Dash, S. Behera, and S. Sabut, "Machine learning approach to detect cardiac arrhythmias in ECG signals: A survey," *Irbm*, vol. 41, no. 4, pp. 185–194, 2020.

[11] Y. Kaya, F. Kuncan, and R. Tekin, "A new approach for congestive heart failure and arrhythmia classification using angle transformation with LSTM," *Arab J Sci Eng*, vol. 47, no. 8, pp. 10497–10513, 2022.

[12] A. S. Eltrass, M. B. Tayel, and A. I. Ammar, "A new automated CNN deep learning approach for identification of ECG congestive heart failure and arrhythmia using constant-Q non-stationary Gabor transform," *Biomed Signal Process Control*, vol. 65, p. 102326, 2021.

[13] D. Thanapatay, C. Suwansaroj, and C. Thanawattano, "ECG beat classification method for ECG printout with Principle Components Analysis and Support Vector Machines," in *2010 International Conference on Electronics and Information Engineering*, IEEE, 2010, pp. V1-72.

[14] S. Karpagachelvi, M. Arthanari, and M. Sivakumar, "Classification of electrocardiogram signals with support vector machines and extreme learning machine," *Neural Comput Appl*, vol. 21, pp. 1331–1339, 2012.

[15] M. Vijayavanan, V. Rathikarani, and P. Dhanalakshmi, "Automatic classification of ECG signal for heart disease diagnosis using morphological features," *International Journal of Computer Science & Engineering Technology*, vol. 5, no. 4, pp. 449–455, 2014.

[16] S. Hadiyoso and A. Rizal, "Electrocardiogram signal classification using higher-order complexity of hjorth descriptor," *Adv Sci Lett*, vol. 23, no. 5, pp. 3972–3974, 2017.

[17] J. A. Gutiérrez-Gnecchi *et al.*, "DSP-based arrhythmia classification using wavelet transform and probabilistic neural network," *Biomed Signal Process Control*, vol. 32, pp. 44–56, 2017.

[18] Z. Wu *et al.*, "A novel features learning method for ECG arrhythmias using deep belief networks," in *2016 6th International conference on digital home (ICDH)*, IEEE, 2016, pp. 192–196.

[19] H. M. Lynn, S. B. Pan, and P. Kim, "A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks," *IEEE Access*, vol. 7, pp. 145395–145405, 2019.

[20] U. R. Acharya *et al.*, "Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals," *Applied Intelligence*, vol. 49, pp. 16–27, 2019.

[21] S. Han, W. Lee, H. Eom, J. Kim, and C. Park, "Detection of arrhythmia using 1D convolution neural network with LSTM model," *IEIE Transactions on Smart Processing & Computing*, vol. 9, no. 4, pp. 261–265, 2020.

[22] A. Darmawahyuni, S. Nurmaini, M. Yuwandini, M. N. Rachmatullah, F. Firdaus, and B. Tutuko, "Congestive heart failure waveform classification based on short time-step analysis with recurrent network," *Inform Med Unlocked*, vol. 21, p. 100441, 2020.

[23] L. Zheng, Z. Wang, J. Liang, S. Luo, and S. Tian, "Effective compression and classification of ECG arrhythmia by singular value decomposition," *Biomedical Engineering Advances*, vol. 2, p. 100013, 2021.

[24] A. Çınar and S. A. Tuncer, "Classification of normal sinus rhythm, abnormal arrhythmia and congestive heart failure ECG signals using LSTM and hybrid CNN-SVM deep neural networks," *Comput Methods Biomech Biomed Engin*, vol. 24, no. 2, pp. 203–214, 2021.

[25] S. C. Mohonta, M. A. Motin, and D. K. Kumar, "Electrocardiogram based arrhythmia classification using wavelet transform with deep learning model," *Sens Biosensing Res*, vol. 37, p. 100502, 2022.

[26] P. Madan, V. Singh, D. P. Singh, M. Diwakar, B. Pant, and A. Kishor, "A hybrid deep learning approach for ECG-based arrhythmia classification," Bioengineering, vol. 9, no. 4, p. 152, 2022.

[27] E. B. Panganiban, A. C. Paglinawan, W. Y. Chung, and G. L. S. Paa, "ECG diagnostic support system (EDSS): A deep learning neural network based classification system for detecting ECG abnormal rhythms from a low-powered wearable biosensors," Sensing and Bio-Sensing Research, vol. 31, p. 100398, 2021.

[28] M. Salem, S. Taheri, and J. Yuan, "ECG arrhythmia classification using transfer learning from 2-dimensional deep CNN features," in 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2018, pp. 1–4.

[29] J. Rahul and L. D. Sharma, "Automatic cardiac arrhythmia classification based on hybrid 1-D CNN and Bi-LSTM model," *Biocybern Biomed Eng*, vol. 42, no. 1, pp. 312–324, 2022.

[30] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE engineering in medicine and biology magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[31] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[32] L. Eren, T. Ince, and S. Kiranyaz, "A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier," *J Signal Process Syst*, vol. 91, pp. 179–189, 2019.

[33] A. Vijayvargiya, R. Kumar, and N. Dey, "Voting-based 1D CNN model for human lower limb activity recognition using sEMG signal," *Phys Eng Sci Med*, vol. 44, pp. 1297–1309, 2021.

[34] T.-H. Hsieh and J.-F. Kiang, "Comparison of CNN algorithms on hyperspectral image classification in agricultural lands," *Sensors*, vol. 20, no. 6, p. 1734, 2020.

[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, 1997.

[36] X. Hu, S. Yuan, F. Xu, Y. Leng, K. Yuan, and Q. Yuan, "Scalp EEG classification using deep Bi-LSTM network for seizure detection," *Comput Biol Med*, vol. 124, p. 103919, 2020.

[37] F. Landi, L. Baraldi, M. Cornia, and R. Cucchiara, "Working memory connections for LSTM," *Neural Networks*, vol. 144, pp. 334–341, 2021.

[38] K. Khalil, O. Eldash, A. Kumar, and M. Bayoumi, "Economic LSTM approach for recurrent neural networks," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 11, pp. 1885–1889, 2019.

[39] K. Smagulova and A. P. James, "A survey on LSTM memristive neural network architectures and applications," *Eur Phys J Spec Top*, vol. 228, no. 10, pp. 2313–2324, 2019.

[40] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[41] K. A. Althelaya, E.-S. M. El-Alfy, and S. Mohammed, "Stock market forecast using multivariate analysis with bidirectional and stacked (LSTM, GRU)," in *2018 21st Saudi Computer Society National Computer Conference (NCC)*, IEEE, 2018, pp. 1–7.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1233-1246 YEAR: 2024
DOI: 10.17798/bitlisfen.1543679

Computational Fluid Dynamics (CFD) Analysis of 3D Printer Nozzle Designs

Rasul HAJILI[1], Mikail TEMIREL[1*]

[1]*Abdullah Gul University, School of Engineering, Mechanical Engineering Department, Kayseri, Türkiye*
(ORCID: 0009-0002-2363-3225) (ORCID: 0000-0002-8199-0100)

Keywords: 3D printing, nozzle design, computational fluid dynamics (CFD), CFD analysis, printing quality, innovation.

Abstract
Additive manufacturing, particularly 3D printing, has gained significant attention recently due to its flexibility, precision, and sustainability. Among the various 3D printing technologies, Fused Deposition Modeling (FDM) stands out as one of the most popular due to its affordability, ease of use, and print quality. However, a major drawback of FDM-based 3D printers is their relatively low print resolution. One of the key factors influencing print quality is the nozzle design, especially its geometry. As a result, numerous studies in literature have focused on improving 3D printing performance by optimizing nozzle design. In this study, we investigated the effects of nozzle geometry from a Computational Fluid Dynamics (CFD) perspective, examining three aspects: die angle, outlet size, and outlet shape. The CFD analysis revealed that the die angle primarily influences the shear stress within the nozzle, while the outlet size has a significant impact on velocity and pressure difference. The outlet shape affects shear stress, velocity, and pressure difference to a lesser extent than the die angle and size.

1. Introduction

3D printing is a type of additive manufacturing technique. This technology creates objects from a computer aided design (CAD). 3D printing process occurs by the layer-by-layer approach to build a 3D object. 3D printing is widely used in the world nowadays and it is a fast-emerging technology[1]. Industries such as aerospace employ 3D printing technology since it is achievable to create desired lightweight structures[2]. However, improvements in this field are necessary to increase the precision of the printing process to avoid additional finishing procedures and ensure efficient manufacturing.

Fused deposition modeling (FDM) is one of the popular branches of additive manufacturing. The FDM method is a common process where a filament of material is fed onto a heated bed using roller mechanism. The material is extruded through a temperature-controlled nozzle[3, 4]. In the FDM process, polymer thermoplastic filaments are commonly used by melting and extruding them through a nozzle, layer by layer, to create the final product[1]. Choosing materials for the filament and nozzle is crucial for 3D printing technology. The most common materials used in FDM are polylactic acid (PLA), acrylonitrile butadiene styrene (ABS), polypropylene (PP), and polyethylene (PE)[1]. According to the literature[5], compared to polylactic acid (PLA), acrylonitrile butadiene styrene (ABS) has higher tensile strength, Young's modulus, elongation at break, and impact strength in fused deposition modeling. Likewise, nozzle materials with high mechanical properties such as durability, low abrasiveness, and most importantly high thermal conductivity are desirable for additive manufacturing.

Properties and the effects of the flowing fluid can be studied by the method called Computational Fluid Dynamics (CFD). This is a branch of fluid dynamics that is widely used in various engineering fields. Since the equations governing fluid motion are usually too complex to be handled effortlessly using

*Corresponding author: mikail.temirel@agu.edu.tr
Received: 04.09.2024, Accepted: 29.12.2024

analytical methods, especially for complicated geometries, numerical approaches are necessary to lower the cost and computational time. CFD combines numerical analysis and computational software to solve complex fluid-related problems[6, 7]. CFD is a science based on the governing laws of fluid dynamics. It produces comprehensive predictions of fluid-flow phenomena. In addition to fluid flow, the problems involving heat transfer, mass transfer, and chemical reactions can be studied using CFD[7, 8]. Methods such as finite-difference, finite-element, spectral, and spectral element are often used in CFD to discretize and solve the system of algebraic Navier-Stokes equations. However, the results from the CFD studies must be interpreted carefully since errors are possible depending on the flow conditions and numerical methods[9, 10]. One of the examples of CFD's extensive use in R&D is in the aerospace field, where it helps to analyze flow's behavior on specified domains with various boundary conditions[11]. Consequently, it allows engineers to detect and fix design errors early on. Beyond aerospace and automotive industries, CFD is also significant in 3D printing industries as it attracts increasing interest for its practical applications[12]. Furthermore, CFD is popular as a cost-effective alternative that delivers highly accurate results compared to experimental methods.

Within the field of 3D printing, CFD has been a valuable tool for testing various printing parameters, including nozzle speed, shear stress, printability, and pressure distribution. Conducting these tests repeatedly through experiments can be both costly and time-consuming, but CFD reduces the number of necessary repetitions, making the research process more efficient. For example, new nozzle design or new filament used in printability experiments is expensive, but CFD allows researchers to study the flow behaviors and material deformation under stress computationally. In one study, COMSOL Multiphysics software was used for a comparative study of a syringe and screw-based 3D food printer to investigate and compare the flow field characteristic of 3D extrusion-based printing units[13]. Other CFD software like ANSYS POLYFLOW, COMSOL, Abaqus, and FLOW-3D have also been employed to investigate additional design parameters and nozzle configurations[14, 15]. In other study, CFD was employed to investigate the print fidelity for polyvinyl alcohol (PVA) material in extrusion dependent 3D printer to investigate the velocity, pressure, shearing rate, and viscosity distributions in nozzle[16]. Another CFD example with Ansys Fluent, focusing on the comparison of the validity of the CFD and experimental data of various parameters of solar still, conducted by Panchal and Patel[17], when the results of CFD were compared with experimental results, the outcome showed that there was a great amount of agreement. CFD is often utilized in nozzle studies, including the rocket nozzles to examine the velocity and pressure of the flowing fluid for the specified region, and the wall shear forces acting on the nozzle surface. Alamu et al.[18] showed that CFD can be applied to the analysis of printed rocket nozzles by simulating flow conditions, thereby enhancing rocket engine performance. CFD helps to characterize and estimate flow types including laminar flows[19], creeping flows (Reynold number <1)[20, 21], turbulent flows[12], non-Newtonian flows[22] as well as heat and mass transfer[8]. This is crucial for designing nozzle geometry. In particular, laminar flow is typical for most 3D printing processes, especially when using materials like thermoplastics in filament-based printers (FDM). Additionally, CFD can be performed to analyze the characteristics of the 3D printed and finalized products that are subjected to internal or external flows. CFD is also used in bioprinting, which is a branch of 3D printing[23-25]. For example, it can be used to analyze the nozzle design for interactions between bioink characteristics and nozzle designs on cell viability. COMSOL was used in a study to compare the impact of shear stress on cell viability between cylindrical and conical nozzles[26]. In the study by Fareez et al.[27], bioprinting research, CFD is performed to optimize the bioprinting by analyzing the properties such as shear stress, diffusivity, and cell viability to aid in the nozzle design and material selection.

Nozzle design can significantly influence the printing quality and durability of the nozzle material. According to Zhang and Sanjayan[28], extrusion resistance affects nozzle design and consequently 3D printing quality. Lower outlet velocity and shorter nozzle designs can enhance printing quality by reducing resistance. Additionally, the nozzle outlet diameter influences 3D printing quality by affecting tensile strength. Smaller nozzle diameters improve strength through finer layers and better adhesion but may increase printing time[29]. Another study by Kaplan et al[30]. reveals that nozzle modifications in FDM printing reduce travel movements while improving printing quality by allowing toolpath optimization, resulting in increased efficiency. Moreover, Xu et al.[31] indicates that nozzle design in 3D printing significantly affects printing speed and the surface texture of the finalized product. Nevertheless, material choice is crucial for mechanical properties.

The use of CFD is increasingly popular in the design and optimization of 3D printer nozzles. Specifically, for FDM-based 3D printers, nozzle design and printing resolution can be optimized using fluid mechanics principles, as the melted filament within the nozzle behaves like a liquid[32]. CFD provides significant benefit to the field by decreasing the cost, increasing the resolution, and fastening 3D printer developing procedure. In a study, color mixing of an FDM 3D printer nozzle was simulated in terms of filament feed with different temperature and experimentally supported[33]. CFD was also employed to simulate different types of melted thermoplastic filaments in 3D printer nozzle. As an example, carbon fiber-reinforced polymer composites were studied to assess clogging problems using CFD and the discrete element method (DEM), focusing on different fiber lengths, fiber volume fractions, and resin viscosities. The analysis revealed that nozzle clogging tends to occur when either the fiber length or the fiber volume fraction is increased[34]. In a separate study, molten TNT/HMX-based explosives were analyzed using CFD to examine the impact of pressure, nozzle diameter, viscosity, and particles on the printing process. The findings showed that increasing the nozzle diameter and inlet pressure enhances the printing velocity, whereas higher viscosity reduces the printing speed[35]. Another CFD study was conducted for the fidelity analysis of the 3D printer using polyvinyl alcohol (PVA) material in two different nozzle diameters. The findings indicated that as the shear rate increased and the gel viscosity decreased, the material could be successfully extruded from both nozzles. Additionally, the larger diameter nozzle demonstrated better performance compared to the smaller one[16]. Alphonse et al.[36] demonstrated that analyzing the outlet flow behavior using CFD can significantly aid in optimizing the performance of a 3D-printed ABS nozzle aerator design. In summary, the use of CFD for optimizing nozzle designs and finalizing 3D-printed nozzle products is a widespread practice.

However, there is a lack of studies specifically focused on simulating the diverse geometry of 3D printer nozzles. For instance, die angle and outlet geometry and their effects on the nozzle performance were not widely analyzed. Furthermore, the combined effect of outlet diameter, die angle, and outlet geometry on essential properties such as pressure drop, outlet velocity, and wall shear stress are yet to be explored. This study aims to address these gaps by utilizing Computational Fluid Dynamics (CFD) analysis and simulating the flow of the molten filament through these nozzles.

In this study, CFD analysis was conducted for four different FDM-based 3D printer nozzle designs, evaluating the varying three key parameters: outlet diameter, die angle, and outlet shape using molten ABS as the primary fluid. Ansys Fluent was used for CFD simulations. The impact of these parameters on wall shear stress, pressure drop, and maximum velocity was examined, and the results were carefully interpreted to understand the effects of these parameters on nozzle performance.

2. Methodology

2.1. Material Selection

Ansys Fluent was used as a CFD solver to simulate the material deposition from the different nozzle designs. In Ansys Fluent, the solid material was set to aluminum by default, and material selection for the solid could not be omitted from the specifications since choosing one is required. Therefore, to accurately represent the model, Brass C37700 was selected as the material for the nozzle. The liquid properties of ABS filament at its melting temperature were chosen as the fluid. The properties of ABS are given in Table 1.

2.2. Nozzle Design

The objective of this study was to optimize the geometry of 3D printer nozzles. Various aspects of nozzle geometry needed to be considered. However, a systematic approach to select designs for simulation is essential for minimizing the number of simulations required along with reducing computational costs during design optimization. SolidWorks (Dassault Systèmes, 2011) was used to create different nozzle designs. Figure 1 illustrates the key design parameters of the 3D printer nozzle. The designs vary in die angle, shape, and size, which significantly affect the nozzle's performance[32, 37, 38]. Consequently, four nozzle designs, as shown in Figure 2, were selected to investigate the influence of these parameters on nozzle's performance. A conventional 3D printer nozzle was chosen as the reference (shown on the left), with modifications to its die angle, shape, and size to create Design-1, 2, and 3. This approach enables us to assess the impact of each parameter, aiding in the design of more efficient and manufacturable nozzles by focusing on optimizing the most influential features.

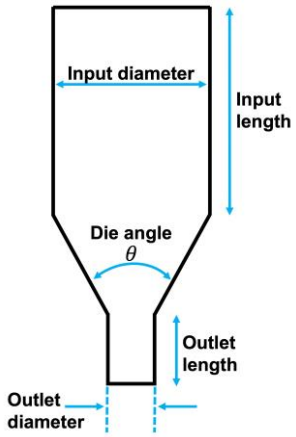


Figure 1. Design parameters of the nozzles.

Table 1. Material properties of the ABS used in the study[39].

Density $(\frac{kg}{m^3})$	Specific Heat Capacity $(\frac{J}{kg} \cdot K)$	Thermal Conductivity $(\frac{W}{m} \cdot K)$	Poisson's Ratio $(-)$	Elastic Modulus (MPa)	Working Temperature $(°C)$	Viscosity $(Pa \cdot s)$
1023	1386	0.2256	0.386	2	240	84

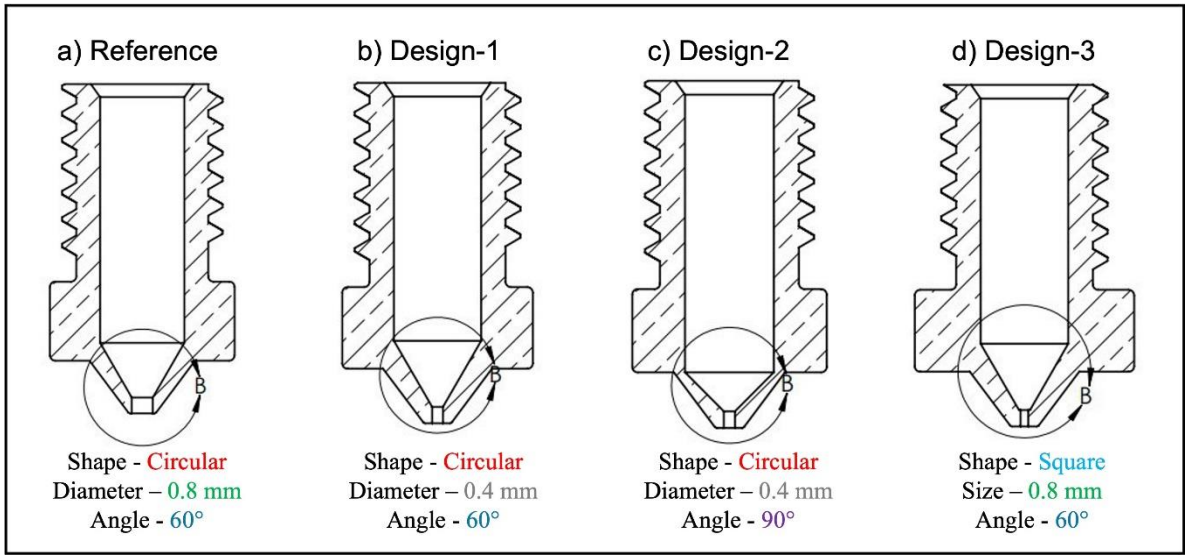


Figure 2. Technical drawing of nozzle designs. a) Reference, b) Design-1, c) Design-2, d) Design-3.

2.3. Simulation Settings

This section discusses the key characteristics of Computational Fluid Dynamics (CFD) analysis that were performed for the fluid flowing through the nozzle. It includes the major stages such as assumptions, meshing, boundary conditions, and simulation setup regarding the analysis.

2.3.1. Assumptions

The operating temperature of the ABS in the liquid state was determined to be 250°C throughout the nozzle. Nevertheless, it is challenging to decide whether non-Newtonian behavior emerges at these elevated temperatures without rheological studies. Further experimental analysis including rheological analysis is required to determine the viscous behavior of the molten ABS under different shear stress values concerning various temperatures. However, based on the observation of maximum shear stress values across the four different nozzle designs, it was assumed that the shear stresses were insufficient to significantly affect the Newtonian behavior of the fluid. Therefore, liquid ABS was considered a Newtonian fluid in this study. The assumption of viscosity being constant with varying shear rates reduces the computational time while introducing a

limitation to the study in terms of accuracy. This limitation can lead to inaccurate representation of the scenarios where shear rates are immense. The potential effect of the significant shear-thinning behavior of the molten filament can result in varying velocity profiles and different pressure distributions throughout the nozzle.

The chosen system for the analysis is an open adiabatic system where only mass flows through the system. Since the temperatures of both the fluid and the solid are constant and equal, there is no heat transfer between the wall and the fluid, making the heat flux equal to zero. This assumption can be considered valid if there is a steady heat source keeping the nozzle's temperature constant, and the filament entering the nozzle is already molten and is in thermal equilibrium with the nozzle.

Flow through the nozzle is considered steady-state flow. The Reynolds numbers for the different outlet designs are calculated based on the following equation:

$$Re = \frac{\rho v d}{\mu} \quad (1)$$

Reynolds number is a dimensionless number that represents the type of flow. In the mentioned equation above, v, ρ, d, and μ are velocity (m/s), fluid density (kg/m^3), outlet diameter (m), and viscosity ($Pa * s$), respectively. According to Cengel and Cimbala[40], flow with $Re < 2300$ is considered to be Laminar. When calculated, the Reynolds Numbers at the inlet in this study for all flows with different nozzle geometries were significantly smaller than 2300. Therefore, it was assumed that the flow throughout the nozzle is Laminar.

In summary of assumptions, the liquid ABS was considered incompressible and Newtonian. Therefore, density and viscosity were constant during the flow throughout the nozzle. The flow was assumed to be in a steady state. Additionally, based on the calculated results of the Reynolds number for each nozzle design, the flow was determined to be laminar. Moreover, the system is in thermal equilibrium given that the temperature of the fluid was set to be equal to the solid's temperature. This condition was achieved by maintaining constant temperatures for both materials. Lastly, the control volume was considered as an open adiabatic system due to no heat exchange across its boundaries. Accurate flow representation limitations that may stem from the assumptions can be attributed to the unknown rheological properties of the molten filament, rather than from the other assumptions mentioned in this section.

2.3.2. Meshing

The meshing process discretizes the object's volume into finite elements. Afterward, the solver software, Ansys Fluent in this study, solves the governing differential equations such as continuity, and Navier-Stokes for each element in the fluid domain. Earlier to meshing, it is crucial to define the boundaries such as inlet, outlet, and wall. These boundaries define the flow domain. Boundaries were specified for three different regions during the internal flow: the inlet surface, the outlet surface, and the wall (Fig. 3). The inlet surface defines the region where the fluid enters the nozzle, and the outlet surface defines the region where the fluid exits. The wall encloses the flow to create an internal flow, with the flowing fluid in contact with the inner surface of the defined wall region. The same meshing technique is applied to the other different nozzle designs in the analysis.

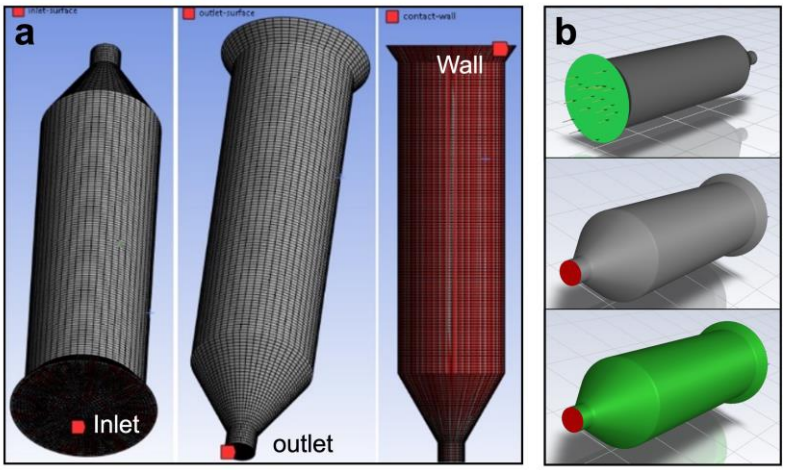


Figure 3. Boundaries of flow in nozzle a) Inlet surface, outlet surface, and wall mesh of the reference nozzle. b) Boundaries of fluid, inlet, outlet, and wall (nozzle inner surface) region.

After defining the boundary regions, meshing was performed with an element size of 8.8314 x 10^{-5} meters. On average, 250,000 nodes were generated during the meshing of each nozzle design. An analysis consisting of 10 different results, including maximum wall shear stress, maximum velocity magnitude, and maximum pressure difference as the number of mesh elements increased, was conducted. Three different figures were plotted to illustrate the fluctuations in results as the number of mesh elements changed. It was observed that significant changes occurred mostly up to 50,000 mesh elements. After 50,000 mesh elements, the results began to stabilize, and the fluctuations minimized. Given that the Ansys Fluent Student version has a limit of 512,000 mesh elements, the 250,000 mesh elements used produced adequate results (Fig. 4).

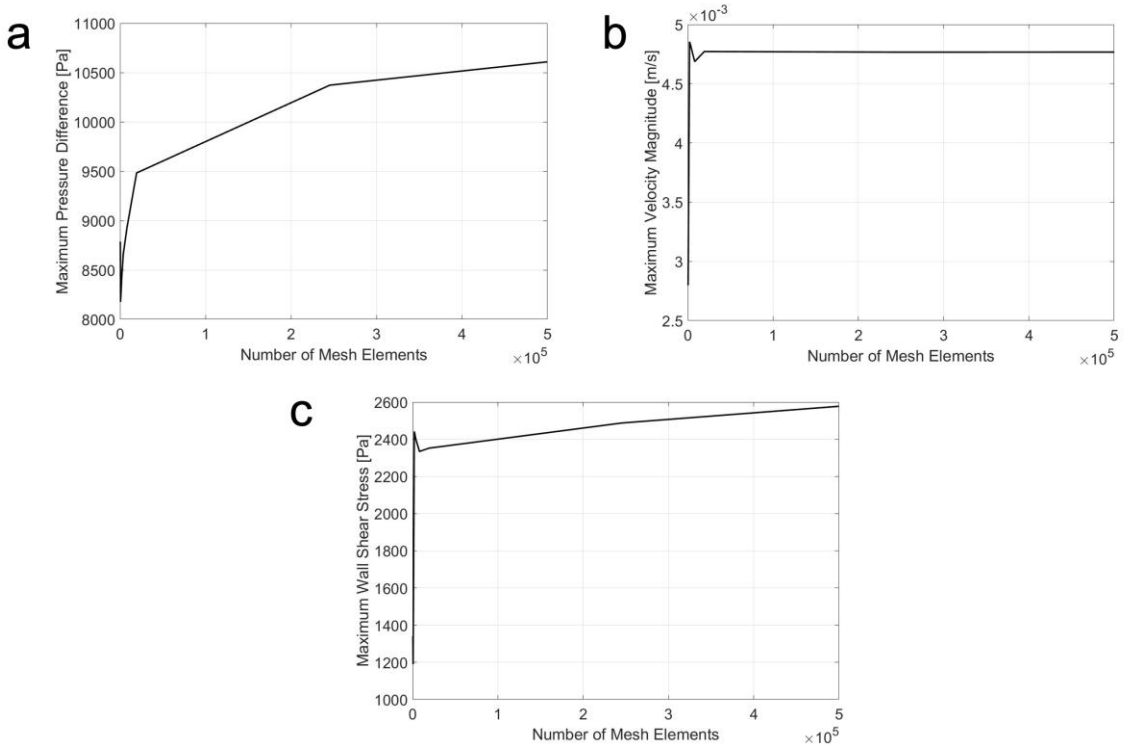


Figure 4. Mesh independence analysis according to the number of elements from 0 to 5 x 10^5. a) Changes in maximum shear stress. b) Changes in maximum velocity magnitude. c) Changes in maximum pressure differences

2.3.3. Boundary Conditions

Defining the boundary conditions in the solver software is one of the fundamental parts of CFD analysis. Boundary conditions involve assigning each solid and fluid material to its corresponding domains and specifying flow and system properties within these defined domains. The boundary conditions applied in the analysis include the conditions at the inlet, outlet, and wall (Fig. 3). Briefly, the definition of this terminology according to the study, Inlet: the region where the fluid enters the nozzle. The flow rate, velocity, or pressure of the entering fluid is specified here. Outlet: the region where the fluid exits the nozzle. The pressure, velocity, or mass flow rate of the exiting fluid is specified here. Wall: the surfaces that enclose the flow, creating an internal flow system. The interactions between the fluid and the wall, such as no-slip conditions, are defined here. These boundary conditions are crucial for accurately simulating fluid behavior and ensuring realistic and reliable results.

The boundary conditions regarding the inlet are specified using a velocity of 0.1 mm/s and a temperature of 503 K. On the other hand, the outlet is set as an outflow with a flow rate weighing 1. Finally, the wall is set as stationary with a no-slip condition and as being adiabatic.

2.3.4. Simulation Setup

During the simulation setup, accuracy was prioritized along with the computational time. Essential software, physical methods, and equations were employed. The solver type was defined as pressure-based, assuming the fluid is incompressible. Discretization methods were selected for accuracy, with strict convergence criteria implemented for continuity and velocities. Ansys Fluent was used as the CFD software package for the simulation. The energy equation was enabled in the physics models

and options of Ansys Fluent. The assumption of no heat transfer between the solid and the fluid implies that conduction heat transfer was not considered, as it was assumed that the solid and fluid were in thermal equilibrium, with negligible heat transfer from the surroundings. The energy equation was enabled to account for any minor changes that might occur during the simulation and to ensure solution stability.

The laminar viscosity model discussed in section 2.3.1 was applied, along with the absolute velocity formulation. The solver type was set to pressure-based with a coupled pressure-velocity coupling scheme. Spatial discretization was performed using a least squares cell-based method for the gradient. Second-order pressure discretization was applied, and both momentum and energy were discretized using a second-order upwind scheme. The time discretization was steady-state, as indicated in the assumptions section. A global time step was used for the pseudo-time method. Residual monitors were set for continuity, x-velocity, y-velocity, z-velocity, and energy, with absolute convergence criteria defined as 0.001 for continuity, x-, y-, and z-velocities, and 10^{-6} for energy.

3. Results and Discussion

This section discusses the results of CFD analysis, including contours and streamlines of various fluid flow properties for each nozzle types. The results also cover wall shear stress, velocity, and pressure values, which will be detailed further in the following sections. As previously mentioned, assumptions, boundary conditions, meshing techniques, convergence criteria, and other simulation settings were kept consistent across all nozzle designs to ensure comparability. Fluid flow properties for each nozzle designs were categorized into five levels: lowest, low, medium, high, and highest. Specifically, minimum and maximum shear stresses on the nozzle's inner wall, the maximum fluid velocity through the nozzle, and the maximum magnitude of pressure drops within the system were analyzed and compared. Before interpreting the results, the way simulations achieved convergence is presented. Figure 5 demonstrates that the solution was considered converged when the residuals dropped below a specific threshold of 10^{-3} for continuity and momentum. Furthermore, the thermal energy monitor remained constant and below the threshold value of 10^{-6} throughout all iterations as anticipated, confirming that no heat transfer analysis was conducted. The convergence criteria were met for all nozzle designs, with the solutions converging on average after 27±8 iterations.

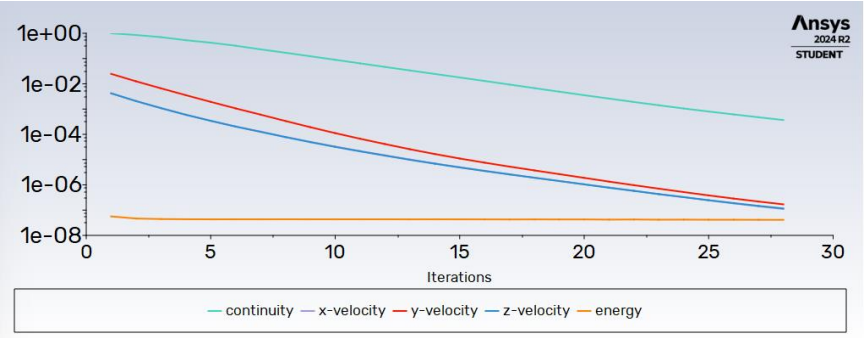


Figure 5. Scaled Residuals Monitor Showing the Convergence for continuity, x-velocity, y-velocity, z-velocity and energy.

Absolute, static, and total pressures were tested in the CFD analysis for all nozzle designs, and as expected, the results for static and total pressures were found to be equivalent. Additionally, it was observed that the pressure differences between the absolute and total pressures were nearly identical. However, the primary goal of including the pressure results in this study is to demonstrate the pressure drop throughout the nozzle, regardless of the pressure type, to accurately observe the effects of nozzle geometry and compare the different designs. Furthermore, the velocity and pressure results are compared to better understand the impact of geometry on performance. Figure 6 illustrates the total pressure values for each region of the various nozzle designs. Negative pressure values can be observed in the Total Pressure Legend. It was anticipated since the initial gauge pressure was set to zero Pascal in the simulation settings. To clarify, when Bernoulli's principle is applied to internal flow within the designed nozzles, it consistently shows that an increase in velocity leads to a decrease in pressure (see Fig. 6). Therefore, observing the negative pressure values in the regions where velocity

increases is understandable. Designs-1 and 2 exhibit higher pressure in the main body compared to the reference and Design-3. This is likely due to their smaller tip diameter (0.4 mm for Design-1 and 2) compared to the 0.8 mm diameter of the reference and size of Design-3. As the flow passes through the nozzle tip in all designs, the pressure gradually decreases until it reaches atmospheric pressure, indicated by the blue color. When comparing Design-1 and Design-2, which have the same outlet diameter, we can observe the effect of the die angle on the pressure drop, as shown in Table 2. Design-2 has a slightly lower pressure drop, despite having a larger die angle. This may be due to the liquid being squeezed through the narrow converging neck in Design-1, which leads to higher pressure.

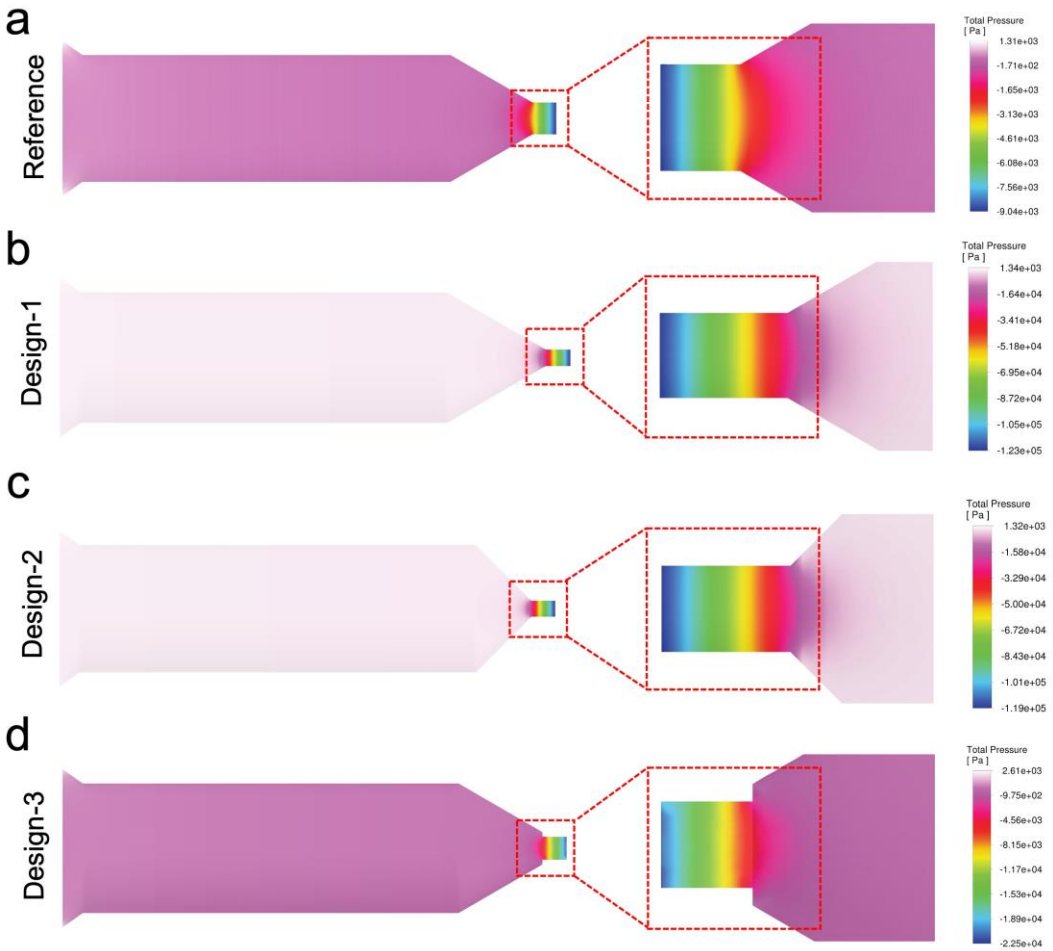


Figure 6. Cross-sectional view of the total pressure magnitude for nozzle designs. a) Reference, b) Design-1, c) Design-2, d) Design-3.

There are no significant differences in the fluid velocity within the main body of all nozzles because the initial velocity of the fluid remains consistent across all designs as seen in the cross-sectional view of the velocity contour results in Figure 7. However, notable changes in fluid velocity occur at the outlet tip, influenced by variations in die angle, outlet diameter, and geometry. For instance, the velocity at the centerline of the reference design is 0.00474 m/s, while it reaches 0.0074 m/s in Design-3, even though both have the same die angle and a very similar outlet area (Table 2). The reference nozzle also maintains the lowest maximum velocity compared to the other designs (Fig. 7), which can be attributed to its largest outlet area. The shear stress at the tip of the surface in the reference design is around $2x10^{-3}$ Pa, whereas in Design-3, it is approximately $1.8x10^{-3}$ Pa, as shown in Figure 8. This suggests that the square shape exerts lower shear stress on the fluid, resulting in higher velocity. Designs-1 and 2 exhibit the same outlet velocity at the centerline, owing to their identical outlet diameters, despite having different die angles. This implies that the die angle does not influence velocity when the diameters are the same. Additionally, Designs-1 and 2 have the highest velocities among the designs due to their smaller outlet areas. Moreover, it is observed that the maximum velocity appears at the centerline of the outlet region for all four nozzle designs.

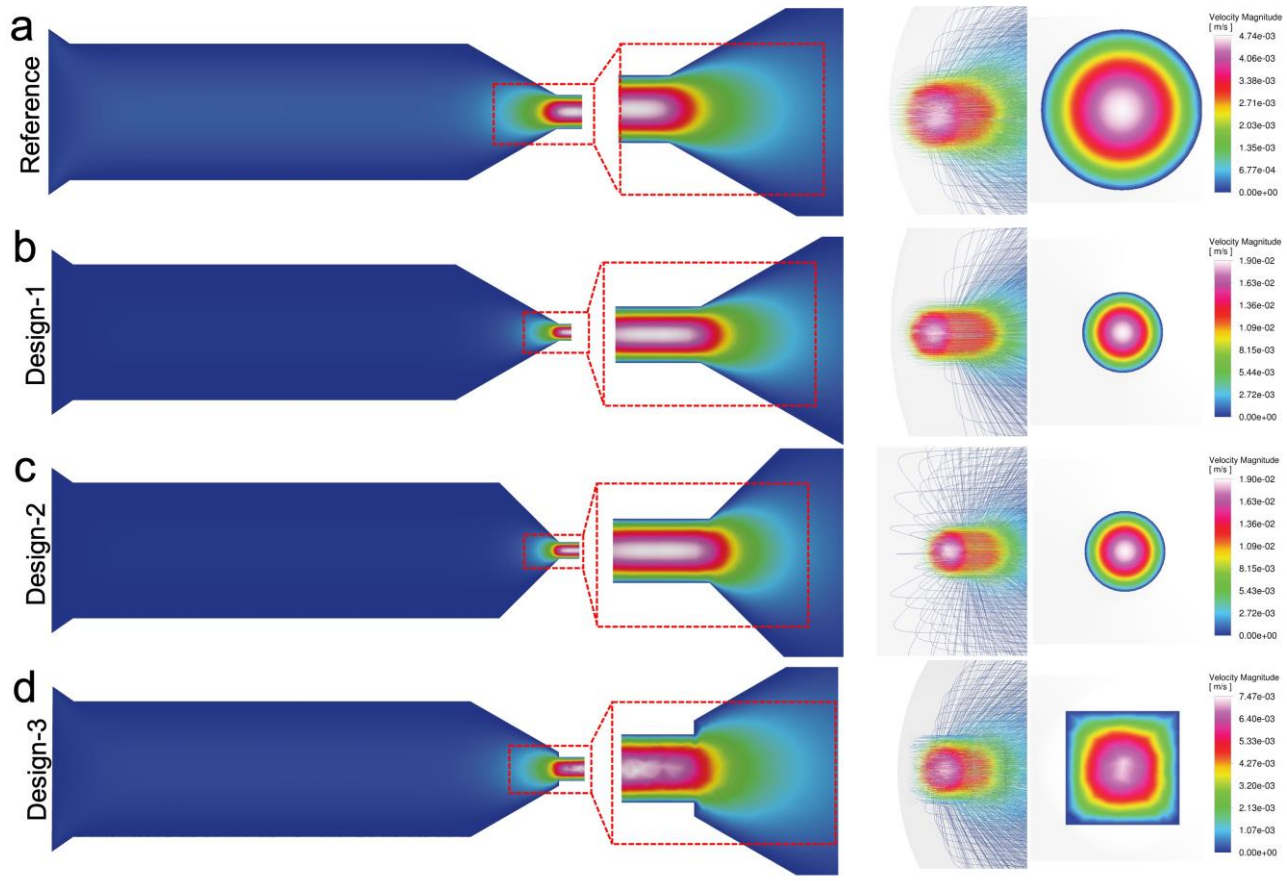


Figure 7. Cross-sectional magnitude, streamline, and frontal view of velocity contour for all nozzle designs. a) Reference, b) Design-1, c) Design-2, d) Design-3.

Table 2. Reynolds number, shear stress, velocity, and pressure drop for the four nozzle designs.

Nozzle	Reynolds Number (Inlet)	Shear Stress (Wall) (Min; Max), (Pa)	Velocity (Max), $\left(\frac{m}{s}\right)$	Pressure Drop (Max. difference), (Pa)
Reference	$4.618 * 10^{-5}$	13.2 ; 2316	0.00474	10345.8
Design-1	$9.256 * 10^{-5}$	12.4 ; 17519	0.0190	123906
Design-2	$9.256 * 10^{-5}$	4.6 ; 18008	0.0190	119859
Design-3	$7.278 * 10^{-5}$	10.0 ; 6438	0.00747	25122

Shear stress plays a crucial role in the performance of 3D printers, particularly in processes like FDM. Higher shear stress within the nozzle reduces the viscosity of the material, improving its flow, which leads to smoother extrusion and better print quality[3, 33].

Design-1 and 2 have almost similar and highest shear stress compared to the reference and Design-4 (Fig. 8). Hence, they can provide better printing quality compared to others. However, if their shear stress is higher than the adequate value, or they have excessive shear stress, it can cause overheating or degradation of the material, negatively affecting the strength and precision of the printing[41]. In the case of excessively low shear stress, the material may not flow properly, resulting in issues like uneven layers[33]. The reference nozzle has the lowest shear stress on the wall compared to the other nozzles (Table 2). This may be due to the reference design's larger outlet area compared to Designs-1 and 2; a smaller outlet area results in higher velocity, which in turn creates higher shear stress. As a result, the melted filament may not be extruded properly, leading to reduced print quality in the reference nozzle compared to the other nozzles. Although Design-3 has the largest outlet area, it experiences slightly

higher shear stress than the reference design due to its square shape. It may provide a similar print quality to the reference nozzle in terms of shear stress, but the square outlet geometry could offer better resolution because of its sharp edges. This might help the layers adhere perfectly without gaps, which may not occur with the curved structure of cylindrical extruded filament. To understand the effect of the die angle on shear stress, Designs-1 and 2 can be compared, as they share the same geometry and diameter but differ in die angles. Design-2 experiences higher shear stress than Design- 1, despite having a larger die angle of 90˚. This can be explained by the fact that the narrowing in Design-1 starts farther from the neck, causing the shear stress distribution to begin farther from the neck as well. In contrast, Design-2 has a less gradual narrowing, meaning the flow is more abruptly squeezed at the neck compared to Design-1. As a result, the larger die angle in Design-2 leads to higher shear stress in the narrowing cylindrical outlet region. Furthermore, the maximum shear stress occurs at the neck of the outlet region in all designs. This is due to the gradual decrease in flow area as the material moves through, reaching the minimum area at the neck. This pattern is also visible in the color spectrum, which ranges from blue (indicating the lowest shear) to white at the neck, where the shear stress is highest. Additionally, improper shear stress levels can increase the risk of nozzle clogging, as the material may not reach a low enough viscosity to maintain a smooth flow. Adequate shear stress is also essential for proper adhesion between layers, which directly influences the structural integrity of the printed object. Balancing shear stress is critical to ensure optimal material flow, layer bonding, and overall printing efficiency[33].

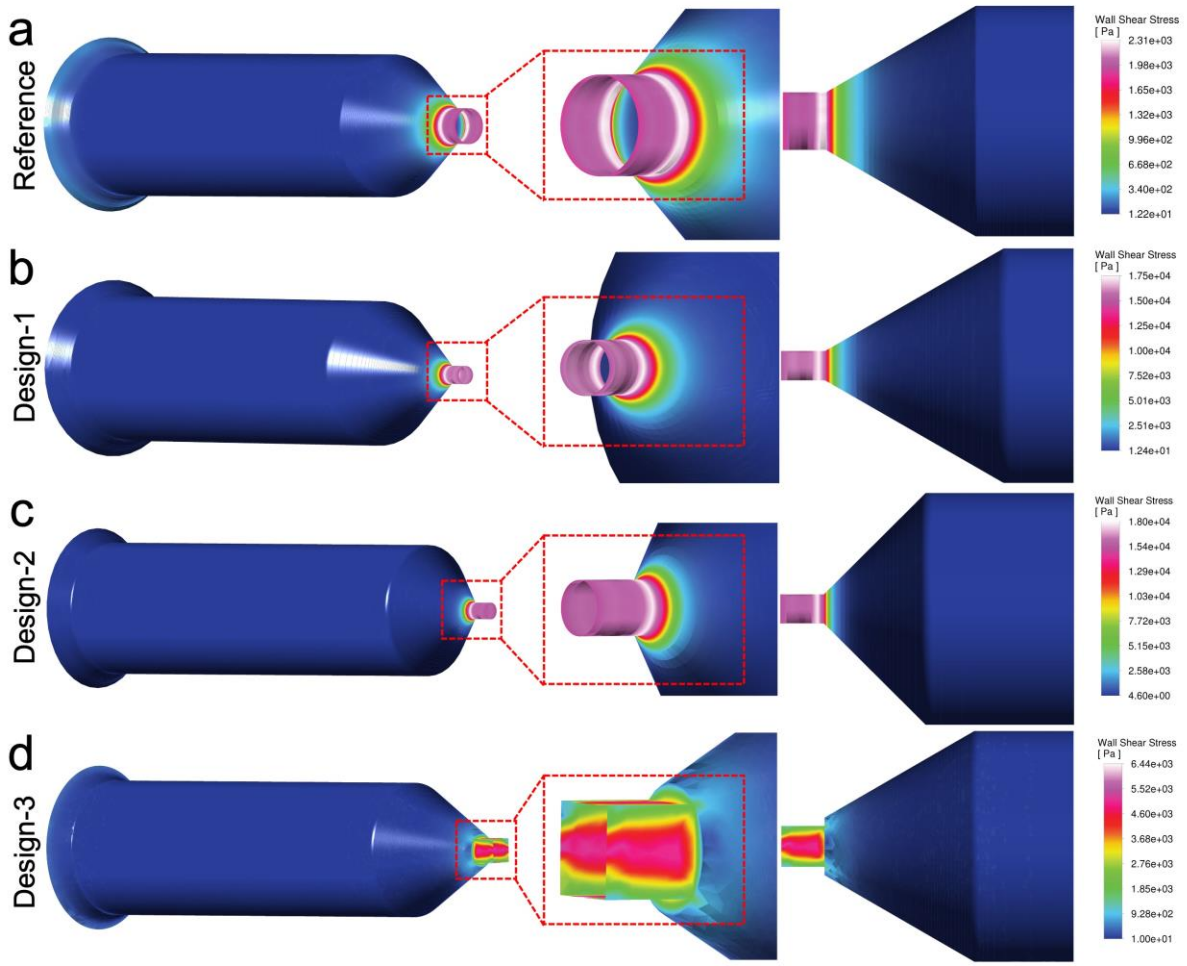


Figure 8. Wall shear stress magnitude throughout the nozzle designs. a) Reference, b) Design-1, c) Design-2, d) Design-3.

4. Conclusion

Consequently, under the specified assumptions and conditions, the CFD analysis of ABS filament in a liquid state flowing through different nozzle designs demonstrated the impact of various nozzle design parameters on system properties. These properties included the shear stress applied by the fluid on the nozzle's inner surface, the fluid velocity profile within the nozzle, and the total pressure variations

throughout the flow. A reduction in the outlet diameter of the reference nozzle led to a significant increase in maximum shear stress, maximum velocity, and the maximum magnitude of pressure drop. On the other hand, an increase in the die angle had a minimal effect on maximum shear stress and no impact on velocity, though it did slightly influence the pressure drop. Modifying the outlet geometry from circular to rectangular caused a moderate increase in shear stress and total pressure drop, with a partial increase in maximum velocity.

Although the manufacturing of the designed nozzles is not a highly challenging task, determining the pressure distribution, velocity profiles, and wall shear stress throughout the nozzle is not experimentally feasible. However, future experimental investigation and validation of the study can be conducted by manufacturing the designed nozzles and operating them in the FDM-based 3-D printer. Necessary sensors and techniques such as high-speed imaging for manufacturing can be employed to validate the CFD results. Alternatively, a feasible validation of this study can be achieved by setting the printing quality and nozzle performance as benchmarks and observing the effects of the study's parameters on these outputs. Thus, the conducted CFD study concerning the effect of nozzle design parameters on essential properties can be validated implicitly in a cost-effective manner. Moreover, future work with experimental analyses can be conducted to address the possible limitations of this study. For instance, rheological analysis can be performed to determine the flow properties of the used molten filament at various temperatures and ensure an accurate Computational Fluid Dynamics (CFD) study.

High shear stress can lead to surface wear. In this case, nozzle material with a high wear resistance should be prioritized. Higher maximum velocity could reduce production time during extrusion. However, the printing quality and cooling time of the extruded material must be inspected carefully when nozzles with higher outlet velocities are used. Moreover, an increase in maximum pressure drops, and correspondingly, an increase in velocity, may contribute to the die-swell in the nozzle. Findings in this study suggest that CFD is an efficient tool for scenarios where experimental research is costly or impractical, to balance nozzle design parameters while meeting printing quality requirements during the use of FDM-based 3D printers.

Acknowledgment

The authors would like to acknowledge Abdullah Gül University for giving an opportunity to work together. We also extend our thanks to Abdurrahman Sivesoğlu for his valuable assistance in the nozzle design.

Contributions of the authors

Conceptualization, M.T.; methodology, visualization, formal analysis, M.T. and R.H.; validation, data curation, writing original draft preparation, M.T. and R.H.; writing review and editing, M.T.; project administration, supervision, M.T. All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

There is no conflict of interest between the authors

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] V. K. Wong and A. Hernandez, "A Review of Additive Manufacturing," *ISRN Mechanical Engineering*, vol. 2012, pp. 1-10, 2012, doi: 10.5402/2012/208760.

[2] N. Shahrubudin, T. C. Lee, and R. Ramlan, "An Overview on 3D Printing Technology: Technological, Materials, and Applications," *Procedia Manufacturing*, vol. 35, pp. 1286-1296, 2019, doi: 10.1016/j.promfg.2019.06.089.

[3] N. Turner, B., Strong, R., & A. Gold, S. (2014). A review of melt extrusion additive manufacturing processes: I. Process design and modeling. Rapid prototyping journal, 20(3), 192-204

[4] A. Waldbaur, H. Rapp, K. Länge, and B. E. Rapp, "Let there be chip—towards rapid prototyping of microfluidic devices: one-step manufacturing processes," (in en), 2011/10/10 2011, doi: 10.1039/C1AY05253E.

[5] M. Lay, N. L. N. Thajudin, A. A. Z. Hamid, A. Rusli, K. M. Abdullah, and K. R. Shuib, "Comparison of physical and mechanical properties of PLA, ABS and nylon 6 fabricated using fused deposition modeling and injection molding," *Composites Part B: Engineering*, vol. 176, p. 107341, 2019, doi: 10.1016/j.compositesb.2019.107341.

[6] E. I. Basri, A. A. Basri, V. N. Riazuddin, S. F. Shahwir, Z. Mohammad, and K. A. Ahmad, "Computational fluid dynamics study in biomedical applications: a review," *International Journal of Fluids and Heat Transfer*, vol. 1, no. 2, pp. 2-14, 2016.

[7] M. H. Zawawi *et al.*, "A review: Fundamentals of computational fluid dynamics (CFD)," *AIP Conference Proceedings*, vol. 2030, no. 1, 2024, doi: 10.1063/1.5066893.

[8] M. M. Aslam Bhutta, N. Hayat, M. H. Bashir, A. R. Khan, K. N. Ahmad, and S. Khan, "CFD applications in various heat exchangers design: A review," *Applied Thermal Engineering*, vol. 32, pp. 1-12, 2012, doi: 10.1016/j.applthermaleng.2011.09.001.

[9] H. H. Hu, "Chapter 10 - Computational Fluid Dynamics," in *Fluid Mechanics (Fifth Edition)*, P. K. Kundu, I. M. Cohen, and D. R. Dowling Eds. Boston: Academic Press, 2012, pp. 421-472.

[10] R. K. Raman, Y. Dewang, and J. Raghuwanshi, "A review on applications of computational fluid dynamics," *International Journal of LNCT*, vol. 2, no. 6, pp. 137-143, 2018.

[11] P. R. Spalart and V. Venkatakrishnan, "On the role and challenges of CFD in the aerospace industry," *The Aeronautical Journal*, vol. 120, no. 1223, pp. 209-232, 2016.

[12] T. Oyinloye and W. Yoon, "Application of Computational Fluid Dynamics (CFD) Simulation for the Effective Design of Food 3D Printing (A Review)," *Processes*, vol. 9, no. 11, p. 1867, 2021, doi: 10.3390/pr9111867.

[13] C.-F. Guo, M. Zhang, and B. Bhandari, "A comparative study between syringe-based and screw-based 3D food printers by computational simulation," *Computers and Electronics in Agriculture*, vol. 162, pp. 397-404, 2019, doi: 10.1016/j.compag.2019.04.032.

[14] M. Temirel, B. Yenilmez, S. Knowlton, J. Walker, A. Joshi, and S. Tasoglu, "Three-Dimensional-Printed Carnivorous Plant with Snap Trap," *3D Printing and Additive Manufacturing*, vol. 3, no. 4, pp. 244-251, 2016, doi: 10.1089/3dp.2016.0036.

[15] T. Glatzel *et al.*, "Computational fluid dynamics (CFD) software tools for microfluidic applications – A case study," *Computers & Fluids*, vol. 37, no. 3, pp. 218-235, 2008, doi: 10.1016/j.compfluid.2007.07.014.

[16] R. Raj, S. V. V. Krishna, A. Desai, C. Sachin, and A. R. Dixit, "Print fidelity evaluation of PVA hydrogel using computational fluid dynamics for extrusion dependent 3D printing - IOPscience," (in en), Text 2022-02-01 2022, doi: doi:10.1088/1757-899X/1225/1/012009.

[17] N. H. Panchal and N. Patel, "ANSYS CFD and experimental comparison of various parameters of a solar still," *International Journal of Ambient Energy*, vol. 39, no. 6, pp. 551-557, 2018, doi: 10.1080/01430750.2017.1318785.

[18] S. O. Alamu, M. J. L. Caballes, Y. Yang, O. Mballa, and G. Chen, "3D Design and Manufacturing Analysis of Liquid Propellant Rocket Engine (LPRE) Nozzle," in *Proceedings of the Future Technologies Conference (FTC) 2019: Volume 2*, 2020: Springer, pp. 968-980.

[19] B. Apacoglu, A. Paksoy, and S. Aradag, "CFD Analysis and Reduced Order Modeling of Uncontrolled and Controlled Laminar Flow Over a Circular Cylinder," *Engineering Applications of Computational Fluid Mechanics*, vol. 5, no. 1, pp. 67-82, 2014, doi: 10.1080/19942060.2011.11015353.

[20] T. Al-Hassan, C. Habchi, T. Lemenand, and F. Azizi, "CFD simulation of creeping flows in a novel split-and-recombine multifunctional reactor," *Chemical Engineering and Processing - Process Intensification*, vol. 162, 2021, doi: 10.1016/j.cep.2021.108353.

[21] A. Tamburini, G. Gagliano, G. Micale, A. Brucato, F. Scargiali, and M. Ciofalo, "Direct numerical simulations of creeping to early turbulent flow in unbaffled and baffled stirred tanks," *Chemical Engineering Science*, vol. 192, pp. 161-175, 2018, doi: 10.1016/j.ces.2018.07.023.

[22] B. Wu and S. Chen, "CFD simulation of non-Newtonian fluid flow in anaerobic digesters," *Biotechnol Bioeng*, vol. 99, no. 3, pp. 700-11, Feb 15 2008, doi: 10.1002/bit.21613.

[23] M. Temirel, S. R. Dabbagh, and S. Tasoglu, "Shape Fidelity Evaluation of Alginate-Based Hydrogels through Extrusion-Based Bioprinting," (in en), *Journal of Functional Biomaterials*, Article vol. 13, no. 4, p. 225, 2022-11-07 2022, doi: 10.3390/jfb13040225.

[24] B. Yenilmez, M. Temirel, S. Knowlton, E. Lepowsky, and S. Tasoglu, "Development and characterization of a low-cost 3D bioprinter," *Bioprinting*, vol. 13, pp. e00044-e00044, 2019.

[25] M. Temirel, C. Hawxhurst, and S. Tasoglu, "Shape Fidelity of 3D-Bioprinted Biodegradable Patches," *Micromachines (Basel)*, vol. 12, no. 2, Feb 13 2021, doi: 10.3390/mi12020195.

[26] A. Malekpour and X. Chen, "Printability and cell viability in extrusion-based bioprinting from experimental, computational, and machine learning views," *Journal of Functional Biomaterials*, vol. 13, no. 2, pp. 40-40, 2022.

[27] M. N. U. Fareez, A. A. S. Naqvi, M. Mahmud, and M. Temirel, "Computational Fluid Dynamics (CFD) Analysis of Bioprinting," *Advanced Healthcare Materials*, 2024, doi: 10.1002/adhm.202400643.

[28] N. Zhang and J. Sanjayan, "Extrusion nozzle design and print parameter selections for 3D concrete printing," *Cement and Concrete Composites*, vol. 137, 2023, doi: 10.1016/j.cemconcomp.2023.104939.

[29] F. Kartal and A. Kaptan, "Investigating the Effect of Nozzle Diameter on Tensile Strength in 3D-Printed Printed Polylactic Acid Parts," (in en), *Black Sea Journal of Engineering and Science*, vol. 6, no. 3, pp. 276-287, July 2023, doi: 10.34248/bsengineering.1287141.

[30] D. Kaplan, S. Rorberg, B. M. Chen, and Y. Sterman, "NozMod: Nozzle Modification for Efficient FDM 3D Printing," in *Proceedings of the 7th Annual ACM Symposium on Computational Fabrication*, ed, 2022.

[31] T. Xu, "Various 3D printing materials with different nozzle design in biomedical area," in *International Conference on Automation Control, Algorithm, and Intelligent Bionics (ACAIB 2022)*, 2022, vol. 12253: SPIE, pp. 445-453.

[32] "3D Printer Nozzle Design and Its Parameters: A Systematic Review | SpringerLink," 2024, doi: 10.1007/978-981-15-2647-3_73.

[33] D. R. Ian Gibson , Brent Stucker, *Additive Manufacturing Technologies: 3D Printing, Rapid Prototyping, and Direct Digital Manufacturing | SpringerLink*. Springer, 2015.

[34] H. Zhang, L. Zhang, H. Zhang, J. Wu, X. An, and D. Yang, "Fibre bridging and nozzle clogging in 3D printing of discontinuous carbon fibre-reinforced polymer composites: coupled CFD-DEM modelling," *The International Journal of Advanced Manufacturing Technology*, vol. 117, no. 11-12, pp. 3549-3562, 2021, doi: 10.1007/s00170-021-07913-7.

[35] H. Zong *et al.*, "Simulation of printer nozzle for 3D printing TNT/HMX based melt-cast explosive," (in En), *The International Journal of Advanced Manufacturing Technology*, OriginalPaper vol. 119, no. 5, pp. 3105-3117, 2022-01-05 2022, doi: doi:10.1007/s00170-021-08593-z.

[36] B. Alphonse, R. Basavaraj, H. Koten, R. Balasubramanian, and S. Umrao, "Comparative design and CFD analysis of 3-D printed acrylonitrile butadiene styrene nozzle aerator for discharge reduction," *Thermal Science*, vol. 26, no. 2 Part A, pp. 857-869, 2022, doi: 10.2298/TSCI201114155A.

[37] O. Hıra, S. Yücedağ, S. Samankan, Y. Ö. Çiçek, and A. Altınkaynak, "Numerical and experimental analysis of optimal nozzle dimensions for FDM printers," *Progress in Additive Manufacturing*, 2022, doi: 10.1007/s40964-021-00241-y.

[38] B. Lizenboim, S. Kenig, and N. Naveh, "The Effect of Nozzle Geometry on the Structure and Properties of 3D Printed Carbon Polyamide Composites," *Applied Composite Materials*, vol. 31, no. 1, pp. 83-99, 2024, doi: 10.1007/s10443-023-10166-0.

[39] S. Han, Y. Xiao, T. Qi, Z. Li, and Q. Zeng, "Design and Analysis of Fused Deposition Modeling 3D Printer Nozzle for Color Mixing," *Advances in Materials Science and Engineering*, vol. 2017, pp. 1-12, 2017, doi: 10.1155/2017/2095137.

[40] Y. Cengel and J. Cimbala, *Ebook: Fluid mechanics fundamentals and applications (si units)*. McGraw Hill, 2013.

[41] Q. Sun, G. M. Rizvi, C. T. Bellehumeur, and P. Gu, "Effect of processing conditions on the bonding quality of FDM polymer filaments," *Rapid Prototyping Journal*, vol. 14, no. 2, pp. 72-80, 2008, doi: 10.1108/13552540810862028.

[42] A. L. Rutz, K. E. Hyland, A. E. Jakus, W. R. Burghardt, and R. N. Shah, "A multimaterial bioink method for 3D printing tunable, cell-compatible hydrogels," *Advanced Materials*, vol. 27, no. 9, pp. 1607-1614, 2015.

[43] N. Hasan and Y. B. Mathur, "Optimizing the Design of Earth Air Tunnel Heat Exchanger for Cooling in Summer Season," *International Multidisciplinary Mulitilingual E-Journal*, vol. 5, no. 1, pp. 19-29.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1247-1259 YEAR: 2024
DOI: 10.17798/bitlisfen.1545337

On The Construction of the Laplace Transform via Gamma Function

Ufuk KAYA[1*], Şeyda ERMİŞ[1]

[1]*Bitlis Eren University, Faculty of Science and Arts, Department of Mathematics, Bitlis, Türkiye (ORCID: 0000-0003-1278-997X) (ORCID: 0009-0007-1187-0588)*

Keywords: Differential equations, Laplace Transform, Operators, Gamma function, Series.

Abstract

The Laplace transform can be applied to integrable and exponential-type functions on the half-line $[0,\infty)$ by the formula $L\{f\}=\int_0^\infty f(x)e^{-sx}dx$. This transform reduces differential equations to algebraic equations and solves many non-homogeneous differential equations. However, the Laplace transform cannot be applied to some functions such as $x^{-\frac{9}{4}}$, because the given integral is divergent. So, the Laplace transform can not solve some differential equations with some terms such as $x^{-\frac{9}{4}}$. This transform requires revision to accommodate such functions and solve a wider class of differential equations. In this study, we defined the Ω-Laplace transform, which eliminates such insufficiency of the Laplace transform and is a generalization of it. We applied this new operator to previously unsolved differential equations and obtained solutions. Ω-Laplace transform given with the help of series:

$$f(x)=\sum_{n=0}^{\infty}c_n x^{r_n}\Rightarrow \Omega\{f\}=\sum_{n=0}^{\infty}\frac{c_n\Gamma(r_n+1)}{s^{r_n+1}}$$

Moreover, we compare the similarities and differences of this transform with the Laplace transform.

1. Introduction

The Laplace operator is a method for solving differential equations. It is an integral operator that transforms a function of a single variable. Laplace wrote extensive books on the use of functions in 1814, and the integral form of the Laplace transform was developed as a result [1]. Beginning in 1744, Leonhard Euler studied the integral transform of the form known as [2]:

$$z=\int X(x)e^{ax}\,dx,\quad z=\int X(x)x^{A}\,dx.$$

Lagrange was an admirer of Euler and in his work on the integration of probability density functions he investigated expressions of the form [3]:

$$\int X(x)e^{-ax}a^{x}\,dx.$$

Integrals of this type attracted the attention of Laplace, and in 1782 he began to use integral operations to solve equations, following Euler. In 1785, instead of simply looking for a solution in integral form, he used an integral of the form [4]:

$$\int x^{s}\varphi(x)dx,$$

where the last expression is called Mellin transform. Laplace developed this transform and finally gave the transform as [5]

$$L\{f\}=\int_0^\infty f(x)e^{-sx}dx. \qquad (1)$$

In sources M. Çağlayan et al. [6], A. Mısır [7], R. N. Bracewell [8], W. Feller [9], G. A. Korn [10], D. V. Widder [11] and J. William [12] Laplace's transformation is given in great detail.

*Corresponding author: mat-ufuk@hotmail.com

Received: 08.09.2024, Accepted: 19.11.2024

The form (1) is still in use today. In sources of differential equations or operator theory, there is usually a table of Laplace transform and inverse Laplace transform. This table is used when solving differential equations. In this table,

$$L\{f\} = \frac{\Gamma(r+1)}{s^{r+1}}$$

is the Laplace transform for $r > -1$of the function

$$f(x) = x^r.$$

When $r < -1$, this function does not have a Laplace transform, since the integral (1) for is divergent for the function $f(x) = x^r$. In this work, we generalize the Laplace transform to eliminate such problems. We will call this new operator Omega-Laplace transform and denote it by Ω-Laplace. The Ω-Laplace transform can only be applied for the functions of the form

$$f(x) = \sum_{n=1}^{\infty} c_n x^{r_n},$$

as

$$\Omega\{f\} = \sum_{n=1}^{\infty} \frac{c_n \Gamma(r_n+1)}{s^{r_n+1}} \qquad (2)$$

We assume that these two series converge on at least one subinterval of $\mathbb{R}$.

In this study, we will show that the Ω transform given by the equality (2) is a generalization of the Laplace transform, we will give the similar and different properties of this operator with the Laplace transform, and we will solve some differential equations that have not been solved before.

Note that There are other operators that have been constructed to solve differential equations that the Laplace transform cannot solve, such as Mellin transform [5]:

$$\mathcal{M}\{f\}(s) = \int_0^{\infty} x^{s-1} f(x) dx,$$

fractional Laplace transform [13]:

$$L_\alpha\{f\}(s) = \int_0^{\infty} f(x) E_\alpha(-sx) dx,$$

where E_α is the Mittag-Leffler function, and Heaviside calculus [14]. The Mellin transform is used to solve some differential equations. However, it turns some terms like $x^k y$ into $\mathcal{M}\{f\}(s+k)$. This makes it difficult to solve the kind of differential equations we solve. Fractional Laplace transform is useful for fractional differential equation. It is also not useful for the kind of differential equations we solve. Heaviside calculus is a method to turn differential expression into algebraic expression like the Laplace transform. However, it is insufficient to solve the differential equations we solved in our examples.
There are recent studies on series representations. In [15-17], series representations of mathematical expressions, such as constants, functions, operators, etc., are obtained and their applications to differential equations or different fields are given. However, in these studies, the definition of the Laplace transform with the help of series and its application to unsolved differential equations are not available.

Schrödinger equation is very important for mathematics, physics and engineering. Particularly, it gives the mathematical formulation of quantum mechanics [18]:

$$i\hbar \frac{\partial}{\partial t}\Psi(x,t) = \left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x,t)\right)\Psi(x,t)$$

in nuclear physics. The function $V(x,t)$ is called potential and it determinates the behavior of particles. In some cases, the potential takes a value such as x^q $(q < -1)$. Then, the Laplace transform or other transform cannot give a solution for the Schrödinger equation. For such equations the Omega-Laplace transform can be used.

In recent years, the generalized Mellin transform, the generalized Fourier transform and the fractional Laplace transform have been investigated in [19-24]. However, the generalizations here are given in the integral form, as in the classical definitions. The series approach is present in our work.

2. Results and Discussion

Classical Laplace transform is defined by

$$L\{f\}(s) = \int_0^{\infty} f(x) e^{-sx}\, dx$$

and is frequently used in differential equations, physics and engineering problems. Although it is very effective in solving problems, we noticed some shortcomings in the definition given by Laplace above. For this reason, we have defined a new Laplace transform using infinite series. So as

not to be confused with the L operator, we will use this new transform in the work of Ω. As we will show in the examples that it can solve some differential equations that the classical Laplace transform cannot solve with the Ω transform. Let's now examine this new definition, its properties and examples.

Definition 1. Let $x \in \mathbb{R}, r_n \in \mathbb{R} \setminus \mathbb{Z}^-, \{c_n\}, \{r_n\}$ be sequences of real numbers and the series

$$f(x) = \sum_{n=0}^{\infty} c_n x^{r_n} \tag{3}$$

converges on at least one subinterval of $\mathbb{R}$. We define the transform of f as follows:

$$\Omega\{f\} = \sum_{n=0}^{\infty} \frac{c_n \Gamma(r_n + 1)}{s^{r_n+1}}. \tag{4}$$

In this study, we will work on functions of type (3) that are expanded into series. When we try to solve differential equations, we try to find solutions in the form of (3). The transform of (3) is given in (4). The transform in (4) forms the basis of the present study and is an alternative to the classical Laplace transform. There is no image of functions like $x^{-\frac{3}{2}}$ under the classical Laplace transform. However, this function has an Ω transform. The Ω transform of this function is $2\sqrt{\pi s}$. Now, let's give some theorems of this new transform and use them in our differential equations.

Theorem 1. We split the series (3) into positive and negative exponents as follows.

$$f(x) = \sum_{n=0}^{\infty} c_n x^{r_n} + \sum_{n=0}^{\infty} d_n x^{-q_n}, \qquad r_n, q_n > 0,$$

Assume that the limits

$$R = \lim_{n\to\infty} |c_n|^{\frac{1}{r_n}}, \qquad Q = \lim_{n\to\infty} |d_n|^{\frac{1}{q_n}}$$

exist and $Q < R$. Then, the series (3) converges on the interval (Q, R). Also, the series converges on the set $(-R, -Q) \cup (Q, R)$ in the case that all the exponents of series f are integers.

Proof. Consider the series

$$\sum_{n=0}^{\infty} c_n x^{r_n} \tag{5}$$

and

$$\sum_{n=0}^{\infty} d_n x^{-q_n}. \tag{6}$$

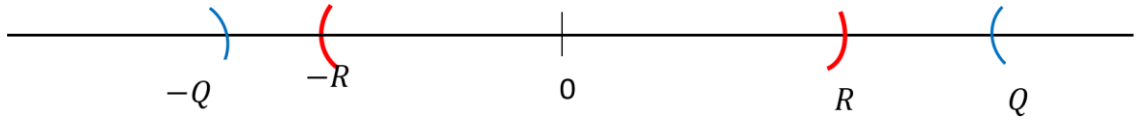
We investigate the convergence of the series (5) and (6) separately. We apply the Cauchy root test for the convergence of the series (5):

$$\begin{aligned}
\lim_{n\to\infty} \sqrt[n]{|c_n x^{r_n}|} &= \lim_{n\to\infty} \sqrt{|c_n|}. \lim_{n\to\infty} |x|^{r_n/n} \\
&= \lim_{n\to\infty} \sqrt[n]{|c_n|}. |x|^{\lim_{n\to\infty} \frac{r_n}{n}} < 1 \\
&\Leftrightarrow |x|^{\lim_{n\to\infty} \frac{r_n}{n}} < \frac{1}{\lim_{n\to\infty} \sqrt[n]{|c_n|}} \\
&\Leftrightarrow |x| < \frac{1}{\left(\lim_{n\to\infty} \sqrt[n]{|c_n|}\right)^{1/\lim_{n\to\infty} \frac{r_n}{n}}} \\
&\Leftrightarrow |x| < \frac{1}{\lim_{n\to\infty} \left(|c_n|^{\frac{1}{n}}\right)^{\frac{n}{r_n}}} \\
&\Leftrightarrow |x| < \frac{1}{\lim_{n\to\infty} |c_n|^{\frac{1}{r_n}}} \\
&\Leftrightarrow |x| < R = \lim_{n\to\infty} |c_n|^{-\frac{1}{r_n}} \\
&-R < x < R, x \in (-R, R),
\end{aligned}$$

We assume that R is positive. Otherwise, the series (2) cannot converge anywhere. If at least one of the exponents is not an integer, the interval takes the form $(0, R)$. Now, we investigate the series (6):

$$\begin{aligned}
\lim_{n\to\infty} \sqrt{|d_n \cdot x^{-q_n}|} &= \lim_{n\to\infty} \sqrt[n]{|d_n|} \cdot \lim_{n\to\infty} |x|^{-\frac{q_n}{n}} \\
&= \lim_{n\to\infty} \sqrt[n]{|d_n|} \cdot |x|^{-\lim_{n\to\infty} \frac{q_n}{n}} < 1 \\
&\Leftrightarrow |x|^{-\lim_{n\to\infty} \frac{q_n}{n}} < \frac{1}{\lim_{n\to\infty} \sqrt[n]{|d_n|}} \\
&\Leftrightarrow |x| > \frac{1}{\left(\lim_{n\to\infty} \sqrt[n]{|dn|}\right)^{-\frac{1}{\lim_{n\to\infty} \frac{q_n}{n}}}} \\
&\Leftrightarrow |x| > \lim_{n\to\infty} \left(|d_n|^{\frac{1}{n}}\right)^{\frac{n}{q_n}} \\
&\Leftrightarrow |x| > \lim_{n\to\infty} |d_n|^{\frac{1}{q_n}} = Q \\
&x \epsilon (-\infty, Q) \cup (Q, +\infty).
\end{aligned}$$

If $R \leq Q$, the set of convergence of the series (3) is null. See the following



So, we assume that $Q < R$. When we take the intersection of the convergence sets, we obtain $(-R,-Q) \cup (Q,R)$. If at least one of the exponents is not an integer, the set of convergence is (Q,R).

Theorem 2. For a function f that has the series expansion (3), this expansion is unique.

Proof. We sketch the proof in the case where all the exponents r_n are ordered in the form $r_0 < r_1 < r_2 < \cdots$. The function f has 2 series expansion as

$$f(x) = \sum_{n=0}^{\infty} c_n x^{r_n} = \sum_{n=0}^{\infty} d_n x^{q_n}, r_n, q_n \in \mathbb{R} \setminus \mathbb{Z}^-.$$

Assume that $r_0 = q_0, r_1 = q_1, r_2 = q_2, \ldots$. Even if the exponents are not equal, we can assume the equality as true by writing 0 to the coefficient of different exponents. The last equality can be written by

$$c_0 x^{r_0} + c_1 x^{r_1} + c_2 x^{r_2} + \cdots = d_0 x^{r_0} + d_1 x^{r_1} + d_2 x^{r_2} + \cdots.$$

We product both sides of the equality by x^{-r_0}:

$$c_0 + c_1 x^{r_1 - r_0} + c_2 x^{r_2 - r_0} + \cdots = d_0 + d_1 x^{r_1 - r_0} + d_2 x^{r_2 - r_0} + \cdots$$

The equality takes the form $c_0 = d_0$ for $x = 0$. Similarly, we have

$$\begin{aligned} c_1 &= d_1 \\ c_2 &= d_2 \\ &\vdots \\ c_n &= d_n \\ &\vdots \end{aligned}$$

Theorem 3. (Linearity Property). Let

$$f(x) = \sum_{n=0}^{\infty} c_n x^{r_n}, \qquad g(x) = \sum_{n=0}^{\infty} d_n x^{q_n}$$

and $a, b \in \mathbb{R}$. Then, the equality

$$\Omega\{af(x) + bg(x)\} = a\Omega\{f(x)\} + b\Omega\{g(x)\}$$

holds.

Proof. Consider two functions f and g in the form

$$f(x) = \sum_{n=0}^{\infty} c_n x^{r_n}, \qquad g(x) = \sum_{n=0}^{\infty} d_n x^{q_n}$$

and apply the operator Ω:

$$\begin{aligned} &\Omega\{af(x) + bg(x)\} \\ &= \Omega\left\{a \sum_{n=0}^{\infty} c_n x^{r_n} + b \sum_{n=0}^{\infty} d_n x^{q_n}\right\} \\ &= \Omega\left\{\sum_{n=0}^{\infty} a c_n x^{r_n} + \sum_{n=0}^{\infty} b d_n x^{q_n}\right\} \\ &= \sum_{n=0}^{\infty} \frac{a c_n \Gamma(r_n + 1)}{s^{r_n+1}} + \sum_{n=0}^{\infty} \frac{b d_n \Gamma(q_n + 1)}{s^{q_n+1}} \\ &= a \sum_{n=0}^{\infty} \frac{c_n \Gamma(r_n + 1)}{s^{r_n+1}} + b \sum_{n=0}^{\infty} \frac{d_n \Gamma(q_n + 1)}{s^{q_n+1}} \\ &= a\Omega\{f(x)\} + b\Omega\{g(x)\}. \end{aligned}$$

This completes the proof.

Theorem 4. Let f be a function of the series (3). Then the image of the function fe^{ax} under the transform of Ω is equal to the value of the image of f under the transform Ω at the point $s - a$. That is,

$$\Omega\{f(x)e^{ax}\}(s) = \Omega\{f\}(s - a).$$

Proof: We apply the transform Ω to the multiplication

$$f(x)e^{ax} = \left(\sum_{n=0}^{\infty} c_n x^{r_n}\right)\left(\sum_{n=0}^{\infty} \frac{a^n x^n}{n!}\right)$$

$$= (c_0 x^{r_0} + c_1 x^{r_1} + \cdots + c_n x^{r_n} + \cdots)\left(1 + ax + \frac{a^2x^2}{2!} + \frac{a^3x^3}{3!} + \cdots \frac{+a^n x^n}{n!} + \cdots\right)$$

$$= c_0 x^{r_0} + c_1 x^{r_1} + \cdots + c_n x^{r_n} + \cdots + a(c_0 x^{r_0+1} + c_1 x^{r_1+1} + \cdots + c_n x^{r_n+1} + \cdots)$$

$$+ \frac{a^2}{2!}(c_0 x^{r_0+2} + c_1 x^{r_1+2} + \cdots + c_n x^{r_n+2} + \cdots) + \cdots$$

we have

$$\frac{c_0\Gamma(r_0+1)}{s^{r_0+1}} + \frac{c_1\Gamma(r_1+1)}{s^{r_1+1}} + \cdots + \mathrm{a}\left(\frac{c_0\Gamma(r_0+2)}{s^{r_0+2}} + \frac{c_1\Gamma(r_1+2)}{s^{r_1+1}} + \cdots\right)$$

$$+ \frac{a^2}{2!}\left(\frac{c_0\Gamma(r_0+3)}{s^{r_0+3}} + \frac{c_1\Gamma(r_1+3)}{s^{r_1+3}} + \cdots\right) + \cdots.$$

For $t > -1$ and $\alpha \in \mathbb{R}$, we obtain

$$(1+t)^{\alpha} = \sum_{n=0}^{\infty} \binom{\alpha}{n} t^n$$

The last gives us the relation

$$\frac{1}{(s-a)^{\alpha}} = (s-a)^{-\alpha} = s^{-\alpha}\left(1 + \left(\frac{-a}{s}\right)\right)^{-\alpha}$$

$$= s^{-\alpha} \sum_{n=0}^{\infty} \binom{-\alpha}{n} (-1)^n \left(\frac{a^n}{s^n}\right),$$

where

$$\binom{-\alpha}{n} = \frac{-\alpha(-\alpha-1)\ldots(-\alpha-n+1)}{n!}$$

(see, [25]). Then, we have

$$\frac{1}{(s-a)^{\alpha}} = s^{-\alpha} \sum_{n=0}^{\infty} \frac{-\alpha(-\alpha-1)\ldots(-\alpha-n+1)}{n!}(-1)^n \frac{a^n}{s^n}$$

$$= \sum_{n=0}^{\infty} \frac{\alpha(\alpha+1)\ldots(\alpha+n-1)}{n!} \cdot \frac{a^n}{s^{n+\alpha}}.$$

and,

$$\frac{1}{(s-\alpha)^{r_k+1}} = \sum_{n=0}^{\infty} \frac{(r_k+1)\ldots(r_k+n)}{n!} \cdot \frac{a^n}{s^{n+r_k+1}}.$$

The image of the function f under the transform Ω at $s-a$ is

$$\Omega\{f\}(s-a) = \sum_{n=0}^{\infty} \frac{c_n\Gamma(r_n+1)}{(s-a)^{r_n+1}} = \frac{c_0\Gamma(r_0+1)}{(s-a)^{r_0+1}} + \frac{c_1\Gamma(r_1+1)}{(s-a)^{r_1+1}} + \cdots$$

$$c_0\Gamma(r_0+1)\left(\frac{1}{s^{r_0}+1} + a \cdot \frac{r_0+1}{s^{r_0+2}} + \frac{a^2}{2!}\frac{(r_0+1)(r_0+2)}{s^{r_0+3}} + \cdots\right)$$

$$+c_1\Gamma(r_1+1) + \left(\frac{1}{s^{r_1}+1} + a \cdot \frac{r_1+1}{s^{r_1+2}} + \frac{a^2}{2!}\frac{(r_1+1)(r_1+2)}{s^{r_1+3}} + \cdots\right).$$

$$+\cdots$$

This completes the proof.

For $f(x) = 1$, we have

$$\Omega\{e^{ax}\} = \frac{1}{s-a}.$$

Theorem 5. Let the function

$$f(x)=\sum_{n=0}^{\infty}c_n x^{r_n}, r_{k_0}=0, r_n\notin\mathbb{Z}^-$$

be a differentiable function and the series are uniform convergence on an interval of the reals. Then,

$$\Omega\{f'(x)\}=s\Omega\{f(x)\}-c_{k_0},$$

where, c_{k_0} is the coefficient of the term x^0 in the series. If, the function f is a differentiable function of order n, we have

Now, we apply the operator Ω:

$$\Omega\{f'\}=\sum_{n=0}^{\infty}\frac{c_n r_n\Gamma(r_n)}{s^{r_n}}=s\left(\sum_{n=0}^{\infty}\frac{c_n\Gamma(r_n+1)}{s^{r_n+1}}+\frac{c_{k_0}}{s}-\frac{c_{k_0}}{s}\right)=s\left(\Omega\{f\}-\frac{c_{k_0}}{s}\right)$$
$$=s\Omega\{f\}-c_{k_0}.$$

This completes the proof for first order derivative. If we apply the formula obtained for the first order derivative repeatedly for higher order derivatives, we have

$$\Omega\{f^{(n)}\}=s^n\Omega\{f\}-s^{n-1}c_{k_0}-\cdots-sc_{k_{n-2}}-c_{k_{n-1}}.$$

Theorem 6. If the series (3) and (4) are convergent, then

$$\frac{d}{ds}\Omega\{f(x)\}(s)=\Omega\{-xf(x)\}(s).$$

Proof: If we take the derivative of Ω of f in the form (4) with respect to s, we have

$$\frac{d}{ds}\Omega\{f(x)\}(s)=\sum_{n=0}^{\infty}-c_n(r_n+1)\cdot\Gamma(r_n+1)s^{-(r_n+2)}$$
$$=\sum_{n=0}^{\infty}-\frac{c_n\Gamma(r_n+2)}{s^{r_n+2}}$$
$$=\Omega\left\{\sum_{n=0}^{\infty}-c_n x^{r_n+1}\right\}$$
$$=\Omega\left\{(-x)\sum_{n=0}^{\infty}c_n x^{r_n}\right\}$$
$$=\Omega\{-xf(x)\}.$$

$$\Omega\{f^{(n)}\}=s^n\Omega\{f\}-s^{n-1}c_{k_0}-\cdots-sc_{k_{n-2}}-c_{k_{n-1}}$$

where, $c_{k_m}\ (m=\overline{0,n-1})$ is the coefficient of the term in the series x^m.

Proof: Consider a function f in the form

$$f(x)=\sum_{n=0}^{\infty}c_n x^{r_n}, r_{k_0}=0, r_n\notin\mathbb{Z}^-,$$

then we have

$$f'(x)=\sum_{n=0}^{\infty}c_n r_n x^{r_n-1}.$$

Similarly, we have the following for all the positive integers n

$$\frac{d^n s}{ds^n}\Omega\{f(x)\}=\Omega\{(-1)^n x^n f(x)\}.$$

Theorem 7. If the series (3) are uniformly convergent and the integral

$$\int_0^x f(u)\,du$$

is convergent on an interval $(0,R), R>0$, we have

$$\Omega\left\{\int_0^x f(u)\,du\right\}=\frac{1}{s}\Omega\{f(x)\}(s).$$

Proof. We write the series (3) as

$$f(u) = \sum_{n=0}^{\infty} c_n u^{r_n}.$$

$$\int_0^x \left(\sum_{n=0}^{\infty} c_n u^{r_n} \right) du = \sum_{n=0}^{\infty} \left(c_n \int_0^x u^{r_n}\, du \right) = \sum_{n=0}^{\infty} c_n \left(\frac{u^{r_n+1}}{r_n+1} \Big|_0^x \right) = \sum_{n=0}^{\infty} c_n \frac{x^{r_n+1}}{r_n+1}.$$

We apply the Ω transform:

$$\Omega\left\{ \int_0^x f(u)\, du \right\} = \sum_{n=0}^{\infty} \frac{c_n \Gamma(r_n+2)}{(r_n+1)s^{r_n+2}} = \sum_{n=0}^{\infty} \frac{c_n (r_n+1)\Gamma(r_n+1)}{(r_n+1)s^{r_n+1}s}$$

$$= \frac{1}{s} \sum_{n=0}^{\infty} \frac{c_n \Gamma(r_n+1)}{s^{r_n+1}} = \frac{1}{s} \Omega\{f(x)\}(s).$$

This completes the proof.

Theorem 8. If the series (3) and (4) are uniformly convergent on intervals I and J respectively and the exponents r_n in (3) are not 0, then we have

$$\Omega\left\{ \frac{f(x)}{x} \right\}(s) = \int_s^{\infty} \Omega\{f\}(u)\, du.$$

Proof: Since the following equality holds

$$\frac{f(x)}{x} = \sum_{n=0}^{\infty} c_n x^{r_n - 1}$$

Then we can write

$$\Omega\left\{ \frac{f(x)}{x} \right\} = \sum_{n=0}^{\infty} \frac{c_n \Gamma(r_n)}{s^{r_n}}.$$

If we both multiply and divide the series by r_n, we have

$$\Omega\left\{ \frac{f(x)}{x} \right\} = \sum_{n=0}^{\infty} \frac{c_n \Gamma(r_n) r_n}{s^{r_n} r_n} = \sum_{n=0}^{\infty} \frac{c_n \Gamma(r_n+1)}{s^{r_n} r_n} = \sum_{n=0}^{\infty} c_n \Gamma(r_n+1) \int_s^{\infty} \frac{du}{u^{r_n+1}} = \int_s^{\infty} \sum_{n=0}^{\infty} \frac{c_n \Gamma(r_n+1)}{u^{r_n+1}} du$$

$$= \int_s^{\infty} \Omega\{f\}(u) du.$$

Theorem 9 (Convolution). If f and g are functions given in the form of series (3), and those series both uniforms converge on an interval $(0, R), R > 0$, then the equality holds

$$\Omega\{(f * g)(x)\} = \Omega\{f(x)\} \cdot \Omega\{g(x)\}$$

where

$$f(x) = \sum_{n=0}^{\infty} c_n x^{r_n}, g(x) = \sum_{m=0}^{\infty} d_m x^{q_m}$$

and

$$(f * g)(x) = \int_0^x f(t) \cdot g(x-t)\, dt.$$

The last integral is known as the convolution of the functions f and g.

Proof: First, we find the series expansion of the function $f(t) \cdot g(x-t)$:

$$f(t) \cdot g(x-t) = \left(\sum_{n=0}^{\infty} c_n t^{r_n} \right) \cdot \left(\sum_{m=0}^{\infty} d_m (x-t)^{q_m} \right)$$

$$= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} c_n d_m t^{r_n} (x-t)^{q_m}.$$

Second, we integrate this function on the interval $[0, x]$. By the uniform convergence, we have

$$(f * g)(x) = \int_0^x \left(\sum_{n=0}^{\infty} \sum_{m=0}^{\infty} c_n d_m t^{r_n} (x - t)^{q_m} \right) dt$$
$$= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} c_n d_m \int_0^x t^{r_n} (x-t)^{q_m}\, dt.$$

Third, by change of variable $(t = xu)$, we have

$$(f * g)(x) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} c_n d_m \int_0^1 x^{r_n} u^{r_n} (x - xu)^{q_m}\, du$$
$$= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} c_n d_m x^{r_n+q_m+1} \int_0^1 u^{r_n} (1-u)^{q_m}\, du$$
$$= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} c_n d_m x^{r_n+q_m+1}\, \beta(r_n+1, q_m+1)$$
$$= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} c_n d_m x^{r_n+q_m+1} \frac{\Gamma(r_n+1)\Gamma(q_m+1)}{\Gamma(r_n+q_m+2)}.$$

Finally, we apply the operator Ω:

$$\Omega\{(f * g)\} = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} c_n d_m \frac{\Gamma(r_n+1)\Gamma(q_m+1)\Gamma(r_n+q_m+2)}{\Gamma(r_n+q_m+2) s^{r_n+q_m+2}}$$
$$= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{c_n \Gamma(r_n+1)}{s^{r_n+1}} \frac{d_m \Gamma(q_m+1)}{s^{q_m+1}}$$
$$= \sum_{n=0}^{\infty} \frac{c_n \Gamma(r_n+1)}{s^{r_n+1}} \sum_{m=0}^{\infty} \frac{d_m \Gamma(q_m+1)}{s^{q_m+1}}$$
$$\Omega\{f(x)\} \cdot \Omega\{g(x)\},$$

where β is the beta function, see [8/10].

3. Application

Example 1. Let's calculate the Ω transform of the trigonometric function $\sin bx$ and $\cos bx, b > 0$ by using the Taylor expansion:

$$\sin bx = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1} b^{2n+1}}{(2n+1)!}.$$

Apply the operator Ω to above series:

$$\Omega\{\sin bx\} = \sum_{n=0}^{\infty} \frac{(-1)^n b^{2n+1} (2n+1)!}{(2n+1)!\, s^{2n+2}} = \sum_{n=0}^{\infty} \frac{(-1)^n b^{2n+1}}{s^{2n+2}} = \frac{b}{s^2} \sum_{n=0}^{\infty} (-1)^n \left(\frac{b^2}{s^2}\right)^n$$
$$= \frac{b}{s^2} \frac{1}{1 + \frac{b^2}{s^2}} = \frac{b}{s^2 + b^2}.$$

Similarly, consider the following series and apply the same process:

$$\cos bx = \sum_{n=0}^{\infty} \frac{(-1)^n b^{2n} x^{2n}}{(2n)!}$$
$$\Omega\{\cos bx\} = \sum_{n=0}^{\infty} \frac{(-1)^n b^{2n} (2n)!}{(2n)!\, s^{2n+1}} = \frac{1}{s} \sum_{n=0}^{\infty} \frac{(-1)^n b^{2n}}{s^{2n}} = \frac{1}{s} \sum_{n=0}^{\infty} (-1)^n \left(\frac{b^2}{s^2}\right)^n$$
$$= \frac{1}{s} \frac{1}{1 + \frac{b^2}{s^2}} = \frac{s}{s^2 + b^2}.$$

We see that the images of cos and sin functions under Laplace transform and Ω transform are the same.

Example 2. Solve the first order equation $2xy' = -3y$ with the ; Ω transform.

Solution: Apply the Ω transform of both sides of the equation:

$$2\Omega\{(-x)y'\} = 3\Omega\{y\}.$$

By Theorem 6, we have

$$2\frac{d}{ds}[\Omega\{y'\}] = 3\Omega\{y\}.$$

By Theorem 5, we have

$$2\frac{d}{ds}\left[s\Omega\{y\} - c_{k_0}\right] = 3\Omega\{y\}.$$

By derivative of multiplication, we have

$$2\Omega\{y\} + 2s\frac{d\Omega\{y\}}{ds} = 3\Omega\{y\}$$
$$\Rightarrow 2s\frac{d\Omega\{y\}}{ds} = \Omega\{y\}.$$

Now, we separate the variables:

$$\frac{2\,d\Omega\{y\}}{\Omega\{y\}} = \frac{ds}{s},$$

then

$$2\ln\Omega\{y\} = \ln s + c_1.$$

Apply the exponential function to both sides of equation and reorganize the constant:

$$(\Omega\{y\})^2 = c_2^2 s \quad \left(c_2 = e^{\frac{c_1}{2}}\right)$$
$$\Rightarrow \Omega\{y\} = c_2\sqrt{s}$$
$$\Rightarrow \Omega\{y\} = \frac{c_2\sqrt{s}\Gamma(-\frac{1}{2})}{\Gamma(-\frac{1}{2})}.$$

By the definition of Ω, we find the solution:

$$y = \frac{c}{x^{\frac{3}{2}}} \quad \left(c = \frac{c_2}{\Gamma(-\frac{1}{2})}\right).$$

The general solution of the equation is $y = \frac{c}{x^{\frac{3}{2}}}$. Since the equation is separable, it could be solved even without Ω or Laplace transform. Now, we consider a second-order nonhomogeneous linear differential equation that cannot be solved by classical methods. Note that the analytical solution of the following equation cannot be obtained by classical methods.

Example 4. Solve the equation $4y'' + 2xy' + 5y = 35x^{-9/2}$ with the transform Ω. Let's transform both sides of the equation by Ω.

$$\Omega\{4y'' + 2xy' + 5y\} = \Omega\{35x^{-9/2}\}.$$

By linearity and definition of Ω, we have

$$4\Omega\{y''\} + 2\Omega\{xy'\} + 5\Omega\{y\} = 35\Gamma\left(-\frac{7}{2}\right)s^{\frac{7}{2}}.$$

By Theorem 5 and 6, we have

$$4\left(s^2\Omega\{y\} - sc_{k_0} - c_{k_1}\right) - 2\frac{d}{ds}\Omega\{y'\} + 5\Omega\{y\} = \frac{35.16\sqrt{\pi}}{105}s^{\frac{7}{2}}$$
$$\Rightarrow 4s^2\Omega\{y\} - 4sc_{k_0} - 4c_{k_1} - 2\frac{d}{ds}\left(s\Omega\{y\} - c_{k_0}\right) + 5\Omega\{y\} = \frac{16\sqrt{\pi}s^{\frac{7}{2}}}{3}$$
$$\Rightarrow 4s^2\Omega\{y\} - 4sc_{k_0} - 4c_{k_1} - 2\Omega\{y\} - 2s\frac{d}{ds}\Omega\{y\} + 5\Omega\{y\} = \frac{16\sqrt{\pi}s^{\frac{7}{2}}}{3}$$
$$\Rightarrow -2s\frac{d}{ds}\Omega\{y\} + (4s^2 + 3)\Omega\{y\} = \frac{16\sqrt{\pi}s^{\frac{7}{2}}}{3} + 4sc_{k_0} + 4c_{k_1}$$
$$\Rightarrow \frac{d}{ds}\Omega\{y\} - \left(2s + \frac{3}{2s}\right)\Omega\{y\} = -\frac{8\sqrt{\pi}s^{\frac{5}{2}}}{3} - 2c_{k_0} - \frac{2c_{k_1}}{s}.$$

To solve this differential equation, we multiply each side by $s^{-\frac{3}{2}}e^{-s^2}$ the integrating factor used for linear equations of order 1. To find a particular solution of the differential equation, we choose the coefficients $c_{k_0} = c_{k_1} = 0$ of the terms x^0 and x^1.

$$s^{\frac{-3}{2}}e^{-s^2}\frac{d}{ds}\Omega\{y\} - s^{\frac{-3}{2}}\left(2s+\frac{3}{2s}\right)e^{-s^2}\Omega\{y\} = -\frac{8\sqrt{\pi}se^{-s^2}}{3}$$

$$\Rightarrow \Omega\{y\} = s^{\frac{3}{2}}e^{s^2}\int\left(-\frac{8\sqrt{\pi}se^{-s^2}}{3}ds\right) = s^{\frac{3}{2}}e^{s^2}\left(\frac{4\sqrt{\pi}}{3}e^{-s^2}+c\right) = \frac{4\sqrt{\pi}s^{\frac{3}{2}}}{3} + cs^{\frac{3}{2}}e^{s^2}$$

We choose again $c = 0$. Since $\Omega\{y\} = \frac{4\sqrt{\pi}}{3s^{-3/2}}$, by the definition of the transform Ω, we have

$$y = x^{-\frac{5}{2}}$$

It is an analytical, particular, solution of the differential equation. The general solution of the equation can be obtained by classical methods of differential equations theory using that particular solution.

Ω operator is an operator like Laplace. It has similar and different properties with the Laplace operator. Below is an Ω transform table for the operator like Laplace transform table.

Table 1. Ω Transform table

$f(x)$	$\Omega\{f\}$
1	$\frac{1}{s}$
$x^r e^{ax}$	$\frac{\Gamma(r+1)}{(s-a)^{r+1}}$
$e^{ax}\sin bx$	$\frac{b}{(s-a)^2+b^2}$
$e^{ax}\cos bx$	$\frac{s-a}{(s-a)^2+b^2}$
$(-1)^n x^n f(x)$	$\frac{d^n s}{ds^n}\Omega\{f(x)\}$
$\sin bx$	$\frac{b}{s^2+b^2}$
$\cos bx$	$\frac{s}{s^2+b^2}$
e^{ax}	$\frac{1}{s-a}$
$f(x)e^{ax}$	$\Omega\{f\}(s-a)$
$\int_0^x f(u)\,du$	$\frac{1}{s}\Omega\{f(x)\}(s)$
$f^{(n)}$	$s^n\Omega\{f\} - s^{n-1}c_{k_0} - \cdots - sc_{k_{n-2}} - c_{k_{n-1}}$
$\frac{f(x)}{x}$	$\int_s^\infty \Omega\{f\}(u)\,du$
$\Omega\{(f*g)(x)\}$	$\Omega\{f(x)\}\cdot\Omega\{g(x)\}$

Now, we explain the similarities and differences between Laplace and Ω-Laplace operators. First, we give the similarities

1. Both operators turn a differential equation into an algebraic equation.
2. Both operators turn the functions like $x^r\sin bx$, $\cos bx$ into $\frac{\Gamma(r+1)}{s^{r+1}}, \frac{b}{s^2+b^2}, \frac{s}{s^2+b^2}$ respectively.
3. Both operators have shifting property (see Theorem 4)
4. Both operators have derivative property (see Theorem 6).
5. Both operators have two integration properties (see Theorem 7 and 8).
6. Both operators have convolution property (see Theorem 9).

Second, we give a differences table

Table 2. Differences between Laplace and Ω-Laplace Transform

Laplace transform	Ω-Laplace transform
can be applied exponential order functions integrable on $[0,\infty)$	can be applied functions of the form $f(x)=\sum_{n=0}^{\infty} c_n x^{r_n}$
can sometimes lead to difficult integrals	to take the transformation of a function, it is enough to know that it transforms x^r into $\frac{\Gamma(r+1)}{s^{r+1}}$
can not be applied to functions of the form x^r, $(r<-1)$	can be applied to functions of the form x^r, $(r<-1)$
to be applied to the derivative of a function, the function must be defined at 0	to be applied to the derivative of a function, the function need not be defined at 0
the derivative formula is $L\{f'\}=sL\{f\}-f(0)$	the derivative formula is $L\{f'\}=sL\{f\}-c_{k_0}$
to be applied to the n^{th} derivative of a function, the function and derivatives must be defined at 0	to be applied to the n^{th} derivative of a function, the function and derivatives need not be defined at 0
the higher derivative formula is $L\{f^{(n)}\}=s^nL\{f\}-s^{n-1}f(0)-s^{n-2}f'(0)-\cdots-f^{(n-1)}(0)$	the higher derivative formula is $L\{f^{(n)}\}=s^nL\{f\}-s^{n-1}c_{k_0}-s^{n-2}c_{k_1}-\cdots-c_{k_{n-1}}$
can not solve all differential equations that Ω-Laplace transform can solve	can solve all differential equations that Laplace transform can solve, and it can also solve differential equations other than this

4. Conclusion

In this study, we introduced the Ω-Laplace transform, a significant generalization of the classical Laplace transform. The classical Laplace transform, while powerful in reducing complex differential equations to algebraic forms, faces limitations when applied to functions with terms like x^r where $r<-1$, as the integral involved becomes divergent. The Ω-Laplace transform overcomes this limitation by incorporating a series-based approach, extending its applicability to a broader class of functions. Through several examples, we demonstrated that this new transform is capable of solving differential equations that were previously unsolvable with traditional methods. In addition, the similarities and differences between the Ω-Laplace and classical Laplace transforms were thoroughly analyzed, leading to the development of a comprehensive transformation table for Ω-Laplace.

The practical utility of the Ω-Laplace transform lies in its ability to handle functions that arise in fields like quantum mechanics, where potential functions can take values outside the scope of the classical Laplace transform. By providing an alternative approach to solving complex differential equations, the Ω-Laplace transform opens new avenues for mathematical analysis in applied contexts.

The introduction of the Ω-Laplace transform suggests several promising directions for future research. First, the exploration of further generalizations of this operator may yield even more versatile tools for solving a wider variety of differential equations. Additionally, applying the Ω-Laplace transform to higher-dimensional problems, such as partial differential equations in physics and engineering, could significantly enhance its practical applications. Moreover, studying the connections between Ω-Laplace and other integral transforms, such as the Mellin or Fourier transforms, might provide new insights and powerful hybrid techniques for advanced mathematical and physical models.

Furthermore, implementing the Ω-Laplace transform in computational software could make it accessible to a broader audience of researchers in applied sciences. This work sets the foundation for a deeper exploration of series-based transformations and their potential to revolutionize the solution of intricate mathematical problems.

Contributions of the authors

This study is the master's thesis of Ş.E. U.K. is the thesis supervisor. The authors' contribution to the study is equal.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] E. T. Jaynes, *Probability Theory: The Logic of Science*, vol. 727, Cambridge, UK: Cambridge University Press, 2003.

[2] L. Euler, "De constructione aequationum" [The Construction of Equations], *Opera Omnia*, 1st series (in Latin), vol. 22, pp. 150–161, 1744.

[3] J. L. Lagrange, "Mémoire sur l'utilité de la méthode," *Œuvres de Lagrange*, vol. 2, pp. 171–234, 1773.

[4] I. Grattan-Guinness, "Laplace's integral solutions to partial differential equations," in *Pierre Simon Laplace 1749-1827: A Life in Exact Science*, C. C. Gillispie, Ed., Princeton, USA: Princeton University Press, 1997.

[5] L. Debnath and D. Bhatta, *Integral Transforms and Their Application*, Boca Raton, USA: CRC Press, 2014.

[6] M. Çağlayan, N. Çelik, and S. Doğan, *Adi Diferansiyel Denklemler*, 5th ed., Bursa, Türkiye: Dora Yayıncılık, 2013.

[7] A. Mısır, *Teori Teknik ve Uygulamalı Diferensiyel Denklemler*, Ankara, Türkiye: Gazi Kitabevi, 2016.

[8] R. N. Bracewell, *The Fourier Transform and its Applications*, 3rd ed., Boston, USA: McGraw-Hill, 2000.

[9] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. II, 2nd ed., New York, USA: Wiley, 1971.

[10] G. A. Korn, *Mathematical Handbook for Scientists and Engineers*, 2nd ed., Boston, USA: McGraw-Hill, 1967.

[11] D. V. Widder, *The Laplace Transform*, Princeton Mathematical Series, vol. 6, Princeton, USA: Princeton University Press, 1941.

[12] J. William, *Laplace Transforms*, Problem Solvers, UK: George Allen & Unwin, 1973.

[13] B. J. West, M. Bologna, and P. Grigolini, *Physics of Fractal Operators*, New York, USA: Springer-Verlag, 2003.

[14] J. Mikusinski, *Operational Calculus*, 2nd ed., Oxford, UK: Pergamon Press, 1983.

[15] M. A. Chaudhry and A. Qadir, "Fourier transform and distributional representation of Gamma function leading to some new identities," *International Journal of Mathematics and Mathematical Sciences*, vol. 39, pp. 2091–2096, 2004.

[16] A. Tassaddig, R. Safdar, and T. A. Kanwal, "Distributional representation of gamma function with generalized complex domain," *Advances in Pure Mathematics*, vol. 7, pp. 441–449, 2017.

[17] H. Alzer and K. C. Richards, "Series representations for special functions and mathematical constants," *The Ramanujan Journal*, vol. 40, pp. 291–310, Mar. 2015.

[18] E. Schrödinger, "An undulatory theory of the mechanics of atoms and molecules," *Physical Review*, vol. 28, no. 6, pp. 1049–1070, 1926.

[19] N. Chittora and D. Babel, "A brief study on Fourier transform and its applications," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, no. 12, pp. 1127–1130, 2018.

[20] P. Singh, A. Gupta, and S. D. Joshi, "General parameterized Fourier transform: A unified framework for the Fourier, Laplace, Mellin, and Z Transforms," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1295–1309, 2022.

[21] E. Ata and I. O. Kıymaz, "New generalized Mellin transform and applications to partial and fractional differential equations," *International Journal of Mathematics and Computer in Engineering*, vol. 1, no. 1, pp. 45–66, 2023.

[22] S. Al-Omari, "Estimates and properties of certain q-Mellin transform on generalized q-calculus theory," *Advances in Difference Equations*, vol. 2021, no. 1, p. 233, 2021.

[23] M. Vivas-Cortez, J. N. Valdés, J. E. H. Hernández, J. V. Velasco, and O. Larreal, "On non-conformable fractional Laplace transform," *Applied Mathematics and Information Sciences*, vol. 15, no. 4, pp. 403–409, 2021.

[24] N. A. Khan, O. A. Razzaq, and M. Ayaz, "Some properties and applications of conformable fractional Laplace transform (CFLT)," *Journal of Fractional Calculus and Application*, vol. 9, no. 1, pp. 72–81, 2018.

[25] B. Musayev, M. Alp, N. Mustafayev, and İ. Ekincioğlu, *Analiz I*, Ankara, Türkiye: Tekağaç Eylül Yayıncılık, 2003.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1260-1270 YEAR: 2024
DOI: 10.17798/bitlisfen.1549483

Optimisation of Warehouse Location and Inventory Management for an Industrial Textile Manufacturer Company in Türkiye

Tutku TUTKUN[1], İrem Nur NERGIZ[1], Rukiye KAYA[1], Uğur SATIÇ[1*]

[1]*Abdullah Gül University, Industrial Engineering Department, Kayseri, Türkiye*
(ORCID: 0009-0008-5176-2722) (ORCID: 0009-0005-1602-7130) (ORCID: 0009-0003-5881-0305) (ORCID: 0000-0002-9160-0006)

Keywords: Facility Location, Demand Forecasting, Inventory Management, ABC Analysis, Genetic Algorithms.

Abstract

In this study, we consider the demand forecasting, facility location, and inventory management problems of an industrial textile manufacturer company in Türkiye. First, we begin with the demand forecasting problem for thirty-two different products and employ ABC analysis to categorise the products. Then we test multiple forecasting methods and find out that Exponential Smoothing and Croston's TSB methods perform better in our categories. Using the demand forecast results in the facility location problem, we search for a location in Europe for a warehouse. For the facility location problem, we use a mixed-integer nonlinear mathematical model to minimise the transportation cost, and warehouse rental cost. We solve the model by using GAMS Solver. Then, we handle the inventory management problem and determine the quantity of the products that are sent from the factory and the warehouse to the customer. We propose a genetic algorithm approach that generates reorder quantities and reorder points for both the factory and the warehouse to minimise the total logistics costs, including holding, ordering and stockout costs. We use simulation models to calculate the logistics costs then we use these costs as fitness values to choose the best reorder quantities and reorder points. The proposed approach offers improvement in demand forecasting, inventory management, and facility location problems and brings up a 26% reduction in total logistic costs.

1. Introduction

In this research, we considered an industrial textile manufacturer company which has a wide product range from yarn production and carpets to various kinds of fabrics. Approximately 20% of the company's production capacity serves one main customer, which is an international furniture retailer. This furniture retailer provides a forty-eight-week order forecast to the company. However, only the first five weeks of this forecast represent actual orders, while the remaining forty-three weeks can be altered by the retailer at any time. This forecasting structure creates significant challenges for the company, as fluctuations such as changed or cancelled orders disrupt the production process. Consequently, the company experiences instability in both its production planning and its inventory management, leading to inefficiencies and increased costs. The first stage of this research focuses on addressing the demand forecasting problem experienced in the supply process while dealing with orders from the furniture retailer. This stage employs a variety of forecasting techniques tailored to the specific needs of the company's product lines. Methods such as ABC analysis, Moving Average, Exponential Smoothing, Linear Regression, and Croston's TSB method are utilised to improve forecasting accuracy. The literature has highlighted the importance of these techniques in managing demand variability and enhancing production efficiency. For instance, Tadayonrad and Ndiaye [1] emphasise the importance

*Corresponding author: ugur.satic@agu.edu.tr
Received: 13.09.2024, Accepted: 22.11.2024

of accurate demand forecasting on the efficiency of the inventory management. Abolghasemi et al. [2] state that demand variability is a key challenge for supply chain management since it causes significant forecast errors, disruption in operations and additional costs. The authors investigate 843 real demand time series with different values of coefficient of variations which indicates the volatility in series. They emphasise the challenge of forecasting with this variability and offer a hybrid model to forecast demand. Song and Lu [3] also indicate that uncertain demand makes it difficult to match demand and supply and manage inventory without shortage. Moayedi and Sadeghian [4] present that demand uncertainty and changes in production processes affect supply chain management in a variety of aspects including environmental effects such as excess carbon emission. The authors consider demand, supplies, processing, transportation, shortage and capacity expansion costs as uncertain parameters and offer multi-objective stochastic programming for green supply chain management under uncertainty.

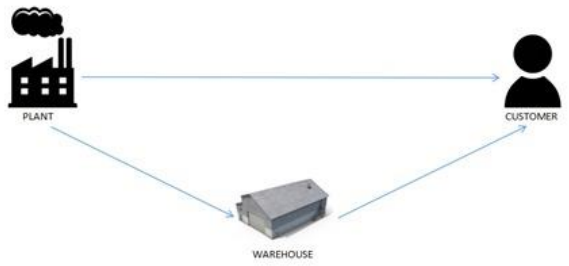


Figure 1. Suggested Operational Flow of the Company's Supply Chain

The second stage of this research addresses the facility location problem. The company requires a warehouse strategically located in Europe to respond more quickly and effectively to orders from the continent. Figure 1 represents the suggested system where some customers will be served from the plant and some customers will be served from the warehouse. The importance of strategic warehouse location is well-documented in the literature. Santosa and Kresna [5] explore the Single Stage Capacitated Warehouse Location Problem (SSCWLP) and emphasise the significance of optimising warehouse locations to reduce logistical costs and improve service levels. Szczepański et al. [6] also discuss the impact of optimal warehouse placement on transportation time and inventory management, noting that proper location selection can lead to substantial operational improvements and cost reductions. In this study, we use a mixed-integer nonlinear mathematical model (MINLP) to optimise the warehouse location, with the objective of minimising total costs, including transportation and warehouse rental costs. Due to the problem structure of the facility location problem, integer programming is extensively adopted in the literature as a solution methodology. Basciftci et al. [7] emphasise the significance of the determination of facility location for decision-dependent demand and offer MILP for the problem. Kchaou Boujelben et al. [8] also uses mixed integer linear programming for multi-period facility location problem. Aboolian et al. [9] handle both facility location and design optimisation in the context of competition with existing facilities. The authors incorporate the uncertainty of customer demands and different design strategies into the approach and offer a generalised facility location and design problem (GFLDP). They also utilise mixed integer nonlinear programming (MINLP) as a part of the solution methodology.

Finally, the third stage of this study is to handle the inventory management problem and generate an effective stock management strategy that sorts out stock-related problems caused by disruptions in the order forecast. One of the significant factors influencing high service levels is the effective management of stocks. By effective stock control company can meet customer demand on time, minimise stockouts, and avoid excess inventory. However, due to the lack of reliable and consistent order forecasts the company is not able to manage their stocks effectively. The company need a robust stock management system to align its stock levels with actual demand and keep the service level high. In this study, we propose an approach based on genetic algorithm to generate optimal reorder quantities and reorder points for the products.

Genetic algorithm (GA) is a stochastic search technique inspired by the natural selection theory of Charles Darwin. The advantages of GA, such as versatility, flexibility, simplicity, and capability to search effectively in large and poorly understood search space with little information, permit it to solve NP-hard optimisation problems across many disciplines [10]. GA offers robustness in navigating complex, nonlinear problem spaces with multiple constraints, making them particularly advantageous over traditional optimisation methods in inventory management [11].

The integration of genetic algorithms in inventory management has been extensively studied. Hernandez and Süer [12] suggested a GA approach to obtain the reorder quantity for an incapacitated, no shortages allowed, single item, single-level situation lot sizing problem by minimising the sum of holding

and ordering costs. Lo [10] considered a production-inventory management model and suggested a GA approach to obtain reorder quantities and reorder points by minimising the average total cost. Pasandideh et al. [13] designed the multi-product, single-supplier inventory management problem as an economic order quantity model and proposed a GA approach which generates order quantities and order levels by minimising the total inventory cost of the supply chain. GA combined with the economic order quantity (EOQ) models, provides a robust solution for improving inventory management, ensuring that the company can maintain optimal stock levels while minimising costs. Babai et al. [11] demonstrates the effectiveness of genetic algorithms in minimising total logistics costs by optimising inventory levels and order intervals in a real-world case study. This approach is particularly relevant for handling complex, nonlinear inventory problems like those encountered by the textile manufacturer in this research. Mahjoob et al. [14] handled a multi-product multi-period inventory routing problem and used a modified adaptive genetic algorithm for the problem. The results present the efficiency of the proposed algorithm. Vidal et al. [15] also emphasise the importance of effective inventory management and propose a decision support system for effective inventory management. They adopt fuzzy MCDM methods to determine and rank SKUs according to importance. Then they develop a ML model which combines genetic algorithm (GA) and artificial neural network (ANN) to forecast demand for these SKUs. The forecast results indicate a significant enhancement in the accuracy of the demand forecast for SKUs compared to the previous forecast of the company. The proposed approach enhances both the responsiveness and the effectiveness of the company's decision-making process in inventory management.

In conclusion, this study integrates advanced forecasting methods, strategic facility location modelling, and genetic algorithms in inventory management to improve the overall efficiency of the company's supply chain operations by improving the production planning of the company and reducing the inventory and transportation costs.

2. Material and Method

Our three-staged study requires different approaches in each stage. The first stage, the demand forecasting problem requires data analysis consisting of data cleaning and categorising the products with ABC analysis. Then different forecasting methods will be selected for each product class (e.g. A, B and C classes). In the second stage which is the facility location problem, a mixed integer nonlinear programming approach using GAMS will be used. In the third stage which is the inventory management problem, we will suggest a genetic algorithm approach.

2.1. The Demand Forecasting Problem

In this stage of the study, we handle the demand forecasting problem. We acquired sales data of 32 products for the years between 2019-2023 from the sales and marketing department of the company. Given the significant impact of the COVID-19 pandemic on sales for the years between 2019-2021, the number of products sold before the pandemic was compared with the number of products sold after the pandemic. There were significant differences between the sales during the COVID-19 pandemic period and after the pandemic period. Thus, we removed the data of the COVID-19 pandemic period from our data set and we only used the data from 2022 and 2023 in this research.

Firstly, we investigated the demand type of the 32 products with ABC analysis. ABC analysis is a critical component of inventory management, categorising inventory items into three categories (A, B, and C) based on their importance and frequency of demand. Alfawaer [16] uses ABC analysis for the design of a forecasting inventory classification model. We used it for the same purpose. We applied ABC analysis to categorize the products based on their order frequency. Firstly, we segmented the sales data into weekly intervals. Then we calculated the frequency of orders by recording the number of weeks in which the product was ordered. Then we calculated the relative percentages of order frequencies. We ranked the products in descending order according to their order frequency percentages. Then we calculated the cumulative values of these relative percentages. Based on the relative percentages, we classified the products into three categories: the products from 0% to 40% are categorised as A, the products from 41% to 80% are categorised as B, and the remaining products are categorised as C. Category A represents products with the highest order frequency, category B includes products with moderate order frequency, and category C consists of products with the lowest order frequency. This method allows for the systematic prioritization of inventory based on demand patterns.

Dutta et al. [17] suggest that the choice of forecasting method can be determined based on the categories determined by ABC classification. Fattah et al. [18] emphasise the importance of selecting appropriate forecasting methods based on demand categorization. They recommend the use of moving average and exponential smoothing for products with

stable demand patterns, highlighting their effectiveness in smoothing out fluctuations and capturing underlying trends. Similarly, Chau [19] advocates for the application of linear regression and exponential moving average, particularly in scenarios where demand trends are linear and consistent over time. These methods provide a robust foundation for forecasting in environments characterized by predictable and steady demand, ensuring accurate predictions and efficient inventory management. In contrast, for products with sporadic and intermittent demand, the use of methods such as Croston's TSB is strongly advocated by Rožanec et al. [20] who highlight their effectiveness in reducing forecasting errors and improving accuracy in industries where demand patterns are volatile.

According to our ABC analysis, 8 products fall into category A, representing the highest priority items. 12 products were categorised as category B. Although ABC classification does not indicate which forecasting method to use for each category, it guides the selection of forecasting techniques for each category according to the specifications and demand patterns of them. Categories of A and B are characterized by high and consistent demand, and we used moving average, exponential smoothing and linear regression methods as demand forecasting methods. The moving average method is a fundamental technique used to analyse time series data. It predicts the next data point in a time series by averaging the data over a given time interval. This method is particularly useful in smoothing out short-term fluctuations and highlighting longer-term trends [21]. Exponential smoothing was utilised to respond to small trend changes, giving more weight to recent observations. This method is effective for time series without significant trends or seasonality, as it adapts quickly to short-term fluctuations, making recent data more relevant [18]. The exponential moving average method is a frequently used method in statistics and financial time series. This method tries to make a forecast by giving more weight to the latest data [22]. On the other hand, the linear regression method is implemented to identify the relationship between time and demand, aiming to improve trend forecasts. This method models demand changes over time, making it particularly effective for datasets with trends offering more accurate predictions of future demand based on historical data [20]. We used these methods for 10-week demand forecasting for each product. Since the exponential smoothing method provided more reliable forecast results among these methods, we adopted this method for classes A and B. The remaining 12 products were categorised as category C by ABC classification. This category exhibits irregular and intermittent demand patterns, and we used Croston, Croston's TSB, bootstrap, and Syntetos-Boylan methods for this category. Croston's method decomposes demand into two separate processes: the demand size and the interval between demand occurrences, making it particularly effective for slow-moving items where traditional methods struggle to provide accurate forecasts [23]. The bootstrap method which is a non-parametric resampling technique estimates statistical measures and their variability without assuming a specific population distribution. It helps create robust forecasts and confidence intervals by resampling the original data multiple times [19]. On the other hand, the Syntetos-Boylan method refines Croston's method by incorporating a correction factor to reduce bias in forecasts [15]. It categorizes demand into intermittent, regular, and quiet periods, improving accuracy for low-volume, sporadic demand. Among the methods for category C, Croston TSB and bootstrap methods provided more reliable predictions for irregular demand. The results are given in detail in section 3.

2.2. The Facility Location Problem

After the demand forecasting problem, we handled the facility location problem. Facility location decision is a strategic decision due to its high cost and long-term effects [16]. We adopted mixed integer nonlinear programming for the facility location problem in this study. In the literature, application of integer programming is extensively observed. Boujelben et. al. [8] used MILP for the facility location problem. They propose a two-phase solution approach. In the first phase, they evaluate the average distances and transportation costs from the distribution centres to customers using an exact clustering procedure based on a set-partitioning formulation. These evaluated costs are then used as input for the second phase, where the facility location problem is formulated as a mixed-integer nonlinear program and solved with a state-of-the-art commercial solver. On the other hand, Branco et al. [24] also utilised MINLP for the location of sugarcane mills.

In this study, firstly we carried out the sales forecast of 32 products, then we continued our research with the facility location problem that the company faced since they want to respond to the orders from Europe more quickly and less costly. 12 possible warehouse locations were considered by the company. The suggested warehouse locations are shown in Table 1. Only one warehouse can be opened.

Table 1. Possible Warehouse Location

County	City
Poland	Warsaw
Bosnia and Herzengoniva	Sarejevo
Romania	Bucharest
Lithuania	Vilnius
Netherlands	Amsterdam
Italy	Rome
Macedonia	Skopje
Bulgaria	Sofia
Hungary	Budapest
Serbia	Belgrade
Germany	Berlin
Portugal	Lisbon

Products can be delivered to customers in two different ways which are from the factory or the warehouse. The company has five customers in Poland, and one customer each in Romania, Bosnia, Portugal, and Lithuania. The demands of the customers are different from each other. A mixed integer nonlinear approach was used for this problem to minimise the total cost which consists of the transportation costs for all alternative routes, the fixed costs of opening a warehouse, and the holding costs for safety stock. A Nomenclature table for this model is shown in Table 2.

2.2.1. Objective function

$$Min\ Z = \sum_i^I \sum_k^K \left(c_{i,k}^{fw} \sum_j^J \left(d_{j,k}\beta_{i,j}\right)\right) + \sum_j^J \sum_k^K \left(c_{j,k}^{fc} d_{j,k}\alpha_j\right) + \sum_i^I \sum_j^J \sum_k^K \left(\left(c_{i,j,k}^{wc} + h_i\right) d_{j,k}\beta_{i,j}\right) + \sum_i^I (\gamma_i g_i) + \sum_i^I \sum_k^K h_i s \sqrt{\sum_j^J g_i \beta_{i,j} \sigma_{j,k}^2} \quad (1)$$

Equation (1) shows the objection function of this model. The objective function calculates the total cost by adding up the transportation costs (composed of factory to warehouse $\left(\sum_i^I \sum_k^K \left(c_{i,k}^{fw} \sum_j^J (d_{j,k}\beta_{i,j})\right)\right)$, factory to customer $\left(\sum_j^J \sum_k^K \left(c_{j,k}^{fc} d_{j,k}\alpha_j\right)\right)$ and warehouse to customer $\left(\sum_i^I \sum_j^J \sum_k^K ((c_{i,j,k}^{wc} + h_i) d_{j,k}\beta_{i,j})\right)$ costs), the rent of a warehouse $\left(\sum_i^I (\gamma_i g_i)\right)$, and the holding costs for safety stock $\left(\sum_i^I \sum_k^K h_i s \sqrt{\sum_j^J g_i \beta_{i,j} \sigma_{j,k}^2}\right)$. The purpose of the function is to minimise the total cost.

The transportation costs are created from many parameters such as customs fees, the fee paid to the vehicles and the driver, and gasoline prices. Additionally, holding costs for only taken into account for the safety stocks. Holding costs amount is obtained from the company.

Table 2. Nomenclature of the MINLP model

Nomenclature	Description
I, J, K	Set of warehouse locations i, set of customer j, set of products k
$c_{i,k}^{fw}$	Transportation cost of one unit k type product from the factory to warehouse i
$c_{j,k}^{fc}$	Transportation cost of one unit k type product from the factory to customer j
$c_{i,j,k}^{wc}$	Transportation cost of one unit k type product warehouse i to customer j
h_i	Holding cost at warehouse locations i
g_i	Rent cost of warehouse locations i
$d_{\mathrm{j,k}}$	(Annual mean) demand of k type product at customer j
$\sigma_{j,k}$	Standard deviation of the demand of k type product at customer j
t_i^{fw}	Transportation duration from the factory to warehouse i
t_i^{fc}	Transportation duration from the factory to customer j
$t_{i,j}^{wc}$	Transportation duration from warehouse i to customer j
p_k	Annual capacity of producing k type product
x_i	The capacity of the warehouse i
s	Desired level of safety factor
m	Threshold for the weighted lead time
α_j	1 if customer j is served by factory, 0 otherwise
$\beta_{i,j}$	1 if customer j is served by warehouse i, 0 otherwise
γ_i	1 if warehouse i is opened, 0 otherwise

2.2.2. Constraints

$$\sum_i^I \sum_j^J (d_{j,k}\beta_{i,j}) + \sum_j^J (d_{j,k}\alpha_j) \leq p_k, \quad \forall k \in K \tag{2}$$

$$\sum_k^K \sum_j^J (d_{j,k}\beta_{i,j}) + \sum_k^K s\sqrt{\sum_j^J (\sigma_{j,k}^2 t_i^{fw}\beta_{i,j})} \leq \gamma_i x_i, \ \forall i \in I \tag{3}$$

$$\sum_j^J \sum_k^K \left(d_{j,k}\left(\sum_i^I (t_{i,j}^{wc}\beta_{i,j}) + t_i^{fc}\alpha_j\right)\right) / \sum_j^J \sum_k^K d_{j,k} \leq m \tag{4}$$

$$\beta_{i,j} \leq \gamma_i, \quad \forall i \in I \ \& \ \forall j \in J \tag{5}$$

$$\sum_k^K s\sqrt{\sum_j^J (\sigma_{j,k}^2 t_i^{fw}\beta_{i,j})} \geq 0, \quad \forall i \in I \tag{6}$$

$$\sum_j^J (d_{j,k}\beta_{i,j}) \geq 0, \quad \forall i \in I \ \& \ \forall k \in K \tag{7}$$

$$\alpha_j + \sum_i^I \beta_{i,j} = 1, \quad \forall j \in J \tag{8}$$

$$\alpha_j, \beta_{i,j}, \gamma_i \in (1,0), \quad \forall i \in I \ \& \ \forall j \in J \tag{9}$$

Equation (2) ensures that the total amount of product that leaves the factory to the warehouse $\left(\sum_i^I \sum_j^J (d_{j,k}\beta_{i,j})\right)$ or directly to customers $\left(\sum_j^J (d_{j,k}\alpha_j)\right)$ cannot exceed the production capacity (p_k) for each product type k. Equation (3) ensures that the total amount of product $\left(\sum_k^K \sum_j^J (d_{j,k}\beta_{i,j})\right)$ in the warehouse j including the safety stock $\left(\sum_k^K s\sqrt{\sum_j^J (\sigma_{j,k}^2 t_i^{fw}\beta_{i,j})}\right)$ cannot exceed the capacity of the warehouse j $(\gamma_i x_i)$. Equation (4) ensures that the weighted average lead time (composed of the lead time from warehouse to customers $\left(\sum_i^I (t_{i,j}^{wc}\beta_{i,j})\right)$, from factory to customers $\left(t_i^{fc}\alpha_j\right)$ and a weighting function of demand $\left(d_{j,k}\right)$) is under the desired threshold (m). Equation (5) ensures that a customer can only be served from an existing warehouse. Equation (6) ensures that the safety stock cannot be negative. Equation (7) ensures that the amount of product sent to the warehouse j cannot be negative. Equation (8) ensures that the customers cannot be served by the warehouse $\left(\beta_{i,j}\right)$ and factory $\left(\alpha_j\right)$ at the same time. Equation (9) represents that these variables are binary.

2.3. The Inventory Management Problem

In this stage of the study, we consider the inventory management problem of the warehouse. Since this problem includes the inventory management of a future warehouse, we assumed that the demand is uncertain. Thus, the weekly shipment quantities from 2021 to 2023 have been examined and a probabilistic distribution of 32 products was found.

This inventory management problem where the demand is stochastic is addressed using a genetic algorithm (GA) to optimise the reorder quantity (Q) and reorder point (R) for 32 products. EOQ equations of the reorder quantity (Q) (i.e., Equation (10)) and reorder point (R) (i.e., Equation (11)) are used to determine holding, ordering and penalty costs.

$$Q_k = \sqrt{\frac{2D_k O_k}{h_i}} \tag{10}$$

$$R_k = D_k L + S \tag{11}$$

Here, D_k is the demand of the product k which is a random variable. O_k is ordering cost per order of product k. h_i is the holding cost of a single unit product in the selected warehouse. L is the lead time. S is the safety stock amount.

GA is a class of optimisation algorithms inspired by the process of natural selection [23]. The GA was implemented using Python [25] and optimisation libraries such as SciPy for efficient computations [26]. The goal of the GA was the minimisation of the total costs (Equation (12)) which includes holding costs (Equation (13)), ordering costs (Equation (14)), and penalty costs.

$$Min\ (total\ cost) = holding\ costs + ordering\ costs + penalty\ costs \tag{12}$$

$$holding\ costs = Average\ inventory\ during\ the\ period * h_i \tag{13}$$

$$ordering\ costs = Number\ of\ orders\ placed\ in\ the\ period * O_k \tag{14}$$

The key steps of GA are as follows: initialisation and iteration. The initialization step consists of the generation of the first population by generation combination of random Q and R values, and evaluation of the fitness values of the first population. We used the total cost as the fitness value. The fitness values of each individual are calculated with a simulation process, where each individual is used in a scenario and based on its Q and R values the sum of the holding cost, penalty cost, and ordering cost are calculated.

The iteration step of GA consists of elitist selection, crossover, mutation and evaluation. The elitist selection is the selection of some individuals with the highest fitness value from the population and transferring them to the next generation without any change. This approach guarantees that our next generation will be worse than its predecessor. The remaining population of the next generation is generated by the crossover phase. In this approach, two individuals are randomly selected from the current population which are called father and mother individuals. A new individual is generated by mixing genes (Q and R values) of the father and mother individuals. We used the one exchange point method where a random exchange point is decided. Then genes are transferred from the mother to the new individual till the exchange point and the remaining genes are transferred from the father after the exchange point. The mutation represents a random change in one gene. All individuals who are generated with the crossover phase may enter the mutation phase by random chance. In this phase, a random gene is selected, and its Q and R values are replaced with a random value. After that in the evaluation phase, the fitness values of the newly generated generation are calculated. The initialisation step only occurs once at the beginning of the algorithm then the iteration step repeats until the termination condition is met.

3. Results and Discussion

3.1. Results of the Demand Forecasting Problem

We investigated the demand type of these 32 products with ABC analysis. The ABC analysis result is shown in Table 3. The analysis revealed that 8 products fell into category A, representing the highest priority items. 12 products were categorised as category B. The remaining 12 products were categorised as category C.

We assessed various forecasting methods such as moving average (3 and 6 periods), exponential smoothing, and linear regression methods for A and B class products, and Croston, Croston TSB, bootstrap and Syntetos-Boylan methods for C class products to determine the best fit for the product classes.

Table 3. ABC Analysis Results

Item	# of week	%	Cumulative	ABC Analysis
A1	86	5.71%	5.71%	A
A2	85	5.64%	11.35%	A
A3	83	5.51%	16.85%	A
A4	74	4.91%	21.77%	A
A5	68	4.51%	26.28%	A
A6	67	4.45%	30.72%	A
A7	65	4.31%	35.04%	A
A8	63	4.18%	39.22%	A
B1	62	4.11%	43.33%	B
B2	61	4.05%	47.38%	B
B3	54	3.58%	50.96%	B
B4	52	3.45%	54.41%	B
B5	49	3.25%	57.66%	B
B6	48	3.19%	60.85%	B
B7	47	3.12%	63.97%	B
B8	44	2.92%	66.89%	B
B9	44	2.92%	69.81%	B
B10	44	2.92%	72.73%	B
B11	39	2.59%	75.32%	B
B12	39	2.59%	77.90%	B
C1	38	2.52%	80.42%	C
C2	35	2.32%	82.75%	C
C3	35	2.32%	85.07%	C
C4	35	2.32%	87.39%	C
C5	29	1.92%	89.32%	C
C6	28	1.86%	91.17%	C
C7	27	1.79%	92.97%	C
C8	24	1.59%	94.56%	C
C9	22	1.46%	96.02%	C
C10	22	1.46%	97.48%	C
C11	19	1.26%	98.74%	C
C12	19	1.26%	100.00%	C

Using the prediction results generated by these methods and the actual outcomes, we calculated the mean absolute percentage error (MAPE) values to compare the performance of the mentioned forecasting methods. MAPE is the average absolute percentage difference between the predicted values and the actual values. MAPE values are calculated as below:

$$\frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t-F_t}{A_t}\right| x100 \qquad (13)$$

Here, A_t is the actual value at time t, F_t is the predicted value at time t, and n is the number of observations. Low MAPE values mean that the prediction method is performing well, and its prediction is close to the actual values. High MAPE values mean that the prediction method performs poorly, and its prediction deviates from the actual values.

Table 4. Comparison of MAPE Values on A and B Types Products

Product Code*	Moving Average 3P	Moving Average 6P	Exponential Smoothing	Linear Regression
A-1	0,34	0,55	**0,27**	1,09
A-2	0,33	0,32	**0,21**	0,54
A-3	**0,57**	1,24	0,61	0,83
A-4	0,36	0,45	**0,25**	0,36
A-5	2,42	2,15	**1,85**	3,43
A-6	0,64	0,95	**0,47**	0,73
A-7	0,78	0,90	0,59	**0,58**
A-8	0,08	0,13	**0,04**	0,58
B-1	0,58	1,03	**0,49**	1,81
B-2	0,84	1,26	**0,54**	1,65
B-3	0,10	0,08	**0,07**	0,73
B-4	0,24	0,45	**0,18**	0,20
B-5	0,31	0,39	**0,21**	0,39
B-6	0,91	1,98	**0,77**	2,17
B-7	0,12	0,15	**0,08**	0,11
B-8	0,11	0,11	**0,07**	0,10
B-9	0,35	0,51	**0,26**	0,49
B-10	16,23	13,19	**9,53**	12,85
B-11	0,42	0,56	**0,29**	0,78
B-12	0,67	0,91	**0,50**	1,45

*Product codes are replaced for anonymity

Performance comparison of method that applied to A and B class products are shown in Table 4. According to Table 4, exponential smoothing generated the lowest MAPE values for most products except for one product where the moving average with 3 periods generated the lowest result and another product where linear regression periods generated the lowest result. Performance comparison of methods that applied to C class products are shown in Table 5. Croston's TSB method generated the lowest MAPE values for all products. Based on the results presented in the tables, we suggest that exponential smoothing should be used for demand forecasting of A and B class products and Croston's TSB method for C class products.

Table 5. Comparison of Mape Values on C Type Products

Product Code*	Croston TSB	Bootstrap	Croston	TSB
C-1	**0,13**	0,42	0,72	0,46
C-2	**0,09**	1,16	0,69	0,76
C-3	**0,11**	1,56	0,77	0,72
C-4	**0,13**	0,76	0,73	0,73
C-5	**0,12**	0,70	0,80	0,43
C-6	**0,03**	2,20	0,63	1,43
C-7	**0,10**	0,73	0,72	0,65
C-8	**0,13**	1,34	0,84	0,74
C-9	**0,06**	1,19	0,76	0,53
C-10	**0,11**	0,41	0,84	0,57
C-11	**0,11**	4,00	0,84	1,52
C-12	**0,02**	2,42	0,78	0,75

*Product codes are replaced for anonymity

3.2. Results of the Facility Location Problem

The facility location problem described in Section 2.2 is solved by a MINLP approach on GAMS solver in 9.22 seconds on a laptop computer with an Intel i5-1155G7 CPU with 2.50 GHz clock speed and 8 GB of RAM.

Our MINLP model, identified Warsaw, Poland as the optimal warehouse location with €197,004.78 profit where both absolute and relative gaps are at 0%. The model specified which customers are to be served directly by the factory and which are to be served via warehouses. The model also calculated the lead times from locations to customers which are shown in Table 6. Additionally, the model calculated the annual quantities of products to be sent to the warehouse and customers which will be used as an input in the inventory management problem.

Establishing a warehouse in Poland not only reduces logistics costs but also improves lead times and service levels for customers in European markets. A strategically placed facility can enhance operational efficiency and responsiveness, thus potentially improving the company's competitive position within the industry.

Table 6. Customer Served by locations

From Factory		From Warehouse	
Customer location	Lead Time (hour)	Customer location	Lead Time (hour)
Poland-City 1	73	Romania-City 1	12
Poland-City 2	61	Poland-City 3	2
Bosnia-City 1	39	Poland-City 4	3
Portugal-City 1	104	Poland-City 5	3
Lithuania-City 1	77		

*City names are replaced for anonymity

3.3. Results of the Inventory Management Problem

Multiple GA parameters are tested to find the best settings to minimise the total costs and it is observed that the minimum cost can be achieved with a population size of 100, a maximum generation number of 1000, an elitist rate of 0.1, the mutation rate of 0.4 and the stopping criteria are either reaching the maximum number of generations or if the best solution remains unchanged for 50 generations.

By implementing the genetic algorithm-based inventory management approach, the company has not only minimised holding and ordering costs but also enhanced stock availability, thus improving customer satisfaction.

3.4. Cost and Savings Analysis

Finally, we compared the existing system with the suggested system where a warehouse is opened in Poland-Warsaw and the ideal stock quantity (Q) and reorder point (R) determined by GA.

The direct shipment costs in the existing system were compared in detail with the shipment costs in the optimised system, which uses a warehouse-based approach. This comparison examined all logistical components, including product costs, direct and warehouse shipment transportation costs, delay costs associated with shipment delays, warehouse rent, and operational expenses. In this comparison, we used recent cost elements, such as 2024 fuel prices, driver services, warehouse rent, and other operational expenses, allowing for a comparison between shipments made from the new warehouse location and those previously made directly from the factory. In this process, customer locations and order quantities were evaluated as critical factors for determining the optimal warehouse location.

Through the optimisation model and genetic algorithm, the ideal stock quantity (Q) and reorder point (R) for the designated warehouse location abroad were calculated based on historical data and cost analysis. These values help optimise stock costs while supporting an on-time delivery process that enhances customer satisfaction.

As a result of these comparisons, the suggested warehouse-based system was found to provide a 26% cost saving compared to the existing non-warehouse-based system. This outcome clearly demonstrates the cost advantages of the suggested system that uses optimisation methods such as MINLP and GA approaches. Our study offers a strategic solution that reduces logistics costs and enhances operational efficiency.

3.5. Long Term Effects

The suggested system in this research contributes to the company's long-term strategic goals and long-term benefits towards customer satisfaction and supply chain management. Our cost saving analysis shows that the suggested system promises a significant 26% cost reduction opportunity compared to the current system on the logistic costs involving storage and transportation costs. However, the cost reduction is not the only benefit of the suggested system. The system suggested warehouse with the optimal location and stock management strategy found by GA also promise reduction in delivery times that enhances customer satisfaction. The increased capacity to respond to customer demands more quickly and efficiently improve both customer loyalty and service quality. These long-term benefits will increase the firm's competitive advantage and improve its position in the market.

4. Conclusion and Suggestions

This study addresses critical issues in the supply chain of a textile manufacturer company in Türkiye. Advanced forecasting techniques improved production planning and stock management accuracy. Methods like ABC analysis, Exponential Smoothing, and Croston's TSB method were used.

A mathematical model for facility location and inventory management optimised the Poland warehouse, reduced logistics costs, and improved delivery times. Combining EOQ with Genetic Algorithms minimised inventory costs and ensured optimal stock levels. Overall, the study enhanced efficiency, reduced costs, improved customer satisfaction, and resulted in a 26% profit increase.

Although the suggested system is designed for the considered textile company, its adaptability extends far beyond the textile sector and can be applied to other sectors or scenarios that involves supply chain and inventory management. For example, in the food and retail sectors, where consumption cycles and product lifespans are short, our suggested system may help to optimize the inventory levels and reduces waste. Especially in the electronics industry, where stock holding costs are high and demand chance quickly, our suggested system may lead significant cost savings. In the pharmaceutical and healthcare industries, our system helps to select optimal warehouse locations to ensures consistent availability of essential supplies. Through these diverse applications, our suggested system also increases the service levels and enhances the customer satisfaction.

Future research should explore the integration of emerging technologies such as machine learning and artificial intelligence to further enhance demand forecasting and inventory management. Additionally, continued evaluation of warehouse operations and logistics strategies will be essential to maintain and improve the firm's competitive edge.

Contributions of the authors

T.Tutkun: literature review, designing the study, methodology and performing the Experiments, writing, editing

İ. N. Nergiz: literature review, designing the study, methodology and performing the Experiments, writing, editing
R. Kaya: literature review, designing the study, methodology, writing and editing
U. Satıç: literature review, designing the study, methodology (only GA), writing and editing

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] Y. Tadayonrad and A. B. Ndiaye, "A new key performance indicator model for demand forecasting in inventory management considering supply chain reliability and seasonality," *Supply Chain Analytics*, vol. 3, Sep. 2023, doi: 10.1016/j.sca.2023.100026.

[2] M. Abolghasemi, E. Beh, G. Tarr, and R. Gerlach, "Demand forecasting in the supply chain: The impact of demand volatility in the presence of promotion," *Comput Ind Eng*, vol. 142, Apr. 2020, doi: 10.1016/j.cie.2020.106380.

[3] Y. Y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," *Shanghai Arch Psychiatry*, vol. 27, no. 2, pp. 130–135, Apr. 2015, doi: 10.11919/j.issn.1002-0829.215044.

[4] M. Moayedi and R. Sadeghian, "A multi-objective stochastic programming approach with untrusted suppliers for green supply chain design by uncertain demand, shortage, and transportation costs," *Journal of Cleaner Production*, vol. 408, Jul. 2023, doi: 10.1016/j.jclepro.2023.137007.

[5] B. Santosa and I. G. N. A. Kresna, "Simulated Annealing to Solve Single Stage Capacitated Warehouse Location Problem," *Procedia Manufacturing*, Elsevier B.V., 2015, pp. 62–70. doi: 10.1016/j.promfg.2015.11.015.

[6] E. Szczepański, R. Jachimowski, M. Izdebski, and I. Jacyna-Gołda, "Warehouse location problem in supply chain designing: A simulation analysis," *Archives of Transport*, vol. 50, no. 2, pp. 101–110, 2019, doi: 10.5604/01.3001.0013.5752.

[7] B. Basciftci, S. Ahmed, and S. Shen, "Distributionally robust facility location problem under decision-dependent stochastic demand," *European Journal of Operational Research*, vol. 292, no. 2, pp. 548–561, Jul. 2021, doi: 10.1016/j.ejor.2020.11.002.

[8] M. Kchaou Boujelben, C. Gicquel, and M. Minoux, "A MILP model and heuristic approach for facility location under multiple operational constraints," *Computers & Industrial Engineering*, vol. 98, pp. 446–461, Aug. 2016, doi: 10.1016/j.cie.2016.06.022.

[9] R. Aboolian, O. Berman, and D. Krass, "Optimizing facility location and design," *European Journal of Operational Research*, vol. 289, no. 1, pp. 31–43, Feb. 2021, doi: 10.1016/j.ejor.2020.06.044.

[10] C. Y. Lo, "Advance of Dynamic Production-Inventory Strategy for Multiple Policies Using Genetic Algorithm" *Information Technology Journal*, vol. 7, pp. 647-653, 2008, doi: 10.3923/itj.2008.647.653

[11] M. Z. Babai, A. Syntetos, and R. Teunter, "Intermittent demand forecasting: An empirical study on accuracy and the risk of obsolescence," *International Journal of Production Economics*, Elsevier B.V., 2014, pp. 212–219. doi: 10.1016/j.ijpe.2014.08.019.

[12] W. Hernandez and G. A. Suer, "Genetic algorithms in lot sizing decisions" *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99* (Cat. No. 99TH8406), Washington, DC, USA, 1999, vol. 3, pp. 2280-2286 doi: 10.1109/CEC.1999.785558.

[13] S. H. R. Pasandideh, S. T. A. Niaki, and A.R. Nia, "A genetic algorithm for vendor managed inventory control system of multi-product multi-constraint economic order quantity model", *Expert Systems with Applications*, vol. 38, no. 3, pp. 2708-2716, 2011, doi: 10.1016/j.eswa.2010.08.060.

[14] M. Mahjoob, S. S. Fazeli, S. Milanlouei, L. S. Tavassoli, and M. Mirmozaffari, "A modified adaptive genetic algorithm for multi-product multi-period inventory routing problem," *Sustainable Operations and Computers*, vol. 3, pp. 1–9, Jan. 2022, doi: 10.1016/j.susoc.2021.08.002.

[15] G. H. de Paula Vidal, R. G. G. Caiado, L. F. Scavarda, P. Ivson, and J. A. Garza-Reyes, "Decision support framework for inventory management combining fuzzy multicriteria methods, genetic algorithm, and artificial neural network," *Computers & Industrial Engineering*, vol. 174, Dec. 2022, doi: 10.1016/j.cie.2022.108777.

[16] Alfawaer Zeyad M., *Proceedings, 2020 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA): Epoka University, Albania: partially held online as a live interactive virtual conference, 9th-10th December, 2020*. IEEE, 2020.

[17] S. Dutta, H. Shah, and P. A. Dasari, K. Singal, N. Y. Harikeerthi, Y. R. Talakola, "Optimizing Inventory Through Abc Classification and Demand Forecasting." *Proceedings of the International Annual Conference of the American Society for Engineering Management.* American Society for Engineering Management (ASEM), 2017.

[18] J. Fattah, L. Ezzine, Z. Aman, H. El Moussami, and A. Lachhab, "Forecasting of demand using ARIMA model," *International Journal of Engineering Business Management*, vol. 10, Oct. 2018, doi: 10.1177/1847979018808673.

[19] N. N. Chau, "Intermittent Demand Forecasting for Inventory Control: The Impact of Temporal and Cross-sectional Aggregation," May 2020.

[20] J. M. Rožanec, B. Fortuna, and D. Mladenić, "Reframing Demand Forecasting: A Two-Fold Approach for Lumpy and Intermittent Demand," *Sustainability (Switzerland)*, vol. 14, no. 15, Aug. 2022, doi: 10.3390/su14159295.

[21] K. C. So and X. Zheng, "Impact of supplier's lead time and forecast demand updating on retailer's order quantity variability in a two-level supply chain," *Int J Prod Econ*, vol. 86, no. 2, pp. 169–179, Nov. 2003, doi: 10.1016/S0925-5273(03)00050-1.

[22] I. Samuel, T. Ojewola, A. Awelewa, and P. Amaize, "Short-Term Load Forecasting Using The Time Series And Artificial Neural Network Methods," *IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE)*, vol. 11, no. 1, pp. 72–81, doi: 10.9790/1676-11137281.

[23] D. SAATÇIOĞLU, "Yapay Sinir Ağlari Yöntemi ile Aralikli Talep Tahmini," *Beykoz Akademi Dergisi*, vol. 4, no. 1, pp. 1–32, Apr. 2016, doi: 10.14514/byk.m.21478082.2016.4/1.1-32.

[24] J. E. H. Branco, D. H. Branco, E. M. de Aguiar, J. V. Caixeta Filho, and L. Rodrigues, "Study of optimal locations for new sugarcane mills in Brazil: Application of a MINLP network equilibrium model," *Biomass Bioenergy*, vol. 127, Aug. 2019, doi: 10.1016/j.biombioe.2019.05.018.

[25] A. C. Türkmen, T. Januschowski, Y. Wang, and A. T. Cemgil, "Intermittent Demand Forecasting with Renewal Processes" *ArXiv*, 2020. Accessed November 10, 2024. https://arxiv.org/abs/2010.01550.

[26] J. Feizabadi. "Machine learning demand forecasting and supply chain performance" *International Journal of Logistics Research and Applications*, vol. 25, no. 2, pp. 119–142, April 2020, doi: 10.1080/13675567.2020.1803246

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1271-1281 YEAR: 2024
DOI: 10.17798/bitlisfen.1552110

Assessing Exposure Levels of Different Groups to Respirable Dust from Building Collapses During the 2023 Kahramanmaraş Earthquakes

Tuba RASTGELDİ DOGAN[1*], Ali SUZERGOZ[1]

[1] *Harran University, Faculty of Engineering, Department of Environmental Engineering, Şanlıurfa, Türkiye*
(ORCID: 0000-0002-8246-388X) (ORCID: 0009-0005-6928-7438)

Keywords: Kahramanmaraş earthquakes, Damaged building, Demolition, Dust, Particulate matter, Environmental exposure.

Abstract

In addition to the immediate destruction caused by earthquakes, significant long-term issues arise, including health problems resulting from dust produced during building demolitions. This study aimed to analyze the perceptual effects of this dust on individuals. A survey was conducted to assess the extent of dust exposure among three groups: a) Demolition officers (DO), b) Environmental safety officers (SO), and c. Individuals present during demolition (IP). Participants provided demographic data and responded to 28 questions about their dust exposure levels. The SO group reported the highest impact from dust, both physiologically and psychologically, while the DO and IP groups were less affected. Health issues, particularly respiratory problems, were notably prevalent. The SO group demonstrated higher sensitivity to dust exposure and the need for protective equipment, highlighting the inadequacy of current protective measures. Our findings also revealed that DO and IP groups had insufficient knowledge about the health risks associated with dust exposure and displayed limited interest in using protective measures. Conversely, although the SO group, with higher education levels, demonstrated greater awareness, they also failed to adequately prioritize protective measures. Overall, the results emphasize the need for improved awareness and more effective protective practices for all individuals involved in demolition activities.

1. Introduction

Earthquakes are among the most devastating natural disasters, causing significant damage to the environment and affecting millions of people worldwide. They occur because of sudden movements in the earth's crust. They seriously affect not only human life but also settlements and infrastructures. Due to its location on active fault lines such as the North Anatolian Fault, East Anatolian Fault, and West Anatolian Fault, Türkiye frequently experiences catastrophic earthquakes [1]. Recent major earthquakes provide notable examples both in terms of destruction, demolition, and their scientific aspects. The 1999 Gölcük Earthquake, with a magnitude of 7.4 MW, and the 7.2 MW Düzce Earthquake are among the most devastating natural disasters in modern Türkiye history [2]. These earthquakes deeply affected the Marmara Region, causing thousands of fatalities and significant damage to hundreds of thousands of buildings. The 7.2 MW Van Earthquake in 2011 resulted in hundreds of deaths and severe infrastructure losses in the Eastern Anatolia Region [3]. In one of the studies, it was stated that the Elazığ-Sivrice earthquake of 2020, with a magnitude of 6.8 MW, caused significant damage to or destruction of reinforced concrete buildings in the city center of Elazığ. It was particularly noted that buildings constructed before the year 2000 suffered more damage, raising concerns about the potential release of large amounts of toxic dust during debris removal activities [4]. In the same year, the 6.6 MW İzmir-Samos Earthquake [5] revealed that the collapsed and heavily damaged

*Corresponding author: trastgeldi@harran.edu.tr
Received: 18.09.2024, Accepted: 27.11.2024

buildings were generally over 30 years old and had inadequate design and structural details. This brings up the issue of debris and its consequences [6]. In our case on February 6, 2023, two major earthquakes centered in the Kahramanmaraş province of Türkiye occurred, the first was recorded at 04:17 with a magnitude of 7.8 Mw, and the second was recorded at 13:27 with a magnitude of 7.6 Mw [7]. These earthquakes also caused severe destruction in the surrounding provinces, such as Hatay, Adıyaman, Gaziantep, Adana, Osmaniye, Kilis, Malatya, Şanlıurfa, Elazığ, and Diyarbakır which were affected to different levels depending on their distance from the earthquake epicenter [8]-[12]. The degree of building damage is shown in table 1.

Table 1. The extent of destruction caused by earthquakes is categorized, showing the levels of damage to buildings and the amount of debris removed as of January 2024 [13]

Province	Demolished	To be demolished urgently	heavily damaged	Moderately damaged	Slightly damaged and undamaged	Removed debris
Adana	38	41	3.330	4.087	358.645	517
Adıyaman	6.187	2.327	21.027	4.215	82.775	17.068
Diyarbakır	44	59	5.491	2.783	183.712	4.120
Elazığ	58	44	10.671	300	23.646	9.074
Gaziantep	4.126	1.988	14.304	5.513	265.262	11.109
Hatay	13.889	9.041	56.214	13.006	257.403	51.974
Kahramanmaraş	7.490	4.436	35.721	6.040	181.459	29.326
Kilis	457	151	1.926	488	32.335	1.769
Malatya	5.651	1.841	36.369	2.520	119.157	27.979
Osmaniye	702	531	9.167	1.074	122.840	8.009
Şanlıurfa	719	732	8.351	2.818	324.921	7.244
Total	39.361	21.191	202.571	43.344	1.952.155	168.189

In addition to the immediate destructive effects of earthquakes, the safe demolition of damaged buildings is also of particular importance. After the completion of search and rescue operations after an earthquake, the removal of debris from collapsed buildings and the controlled demolition of severely damaged buildings require careful planning. During the demolition and debris removal operations, large amounts of dust and hazardous materials of different compositions, especially those originating from construction materials, carry the risk of being released into the atmosphere [14]. The possibility that respirable particles generated during building demolition and debris removal may contain toxic substances [15], heavy metals [16], and particulate matter (PM) can pose serious risks to both the environment and health [17]-[19]. This is a significant concern.

The World Health Organization (WHO) has established strict limits for PM exposure. According to WHO guidelines, the daily limit of 50 μg/m³ for PM10 should not be exceeded more than three times per year, with an annual average limit of 20 μg/m³. For PM2.5, WHO sets a more rigorous annual limit of 10 μg/m³ [20]. The presence of PM2.5 and PM10 in the atmosphere can be attributed to natural events like wildfires and dust storms [21], as well as anthropogenic sources such as the increasing number of vehicles, industrial activities, and other human activities [22]. Additionally, activities like building demolition and the disposal of demolition waste also contribute to PM emissions (Figure 1).

Figure 1. Demolition of a building generating significant PM emissions, illustrating the environmental impact of demolition activities on air quality (Şanlıurfa 2024)

It is well established that respirable dust particles (particulate matter with a diameter of <10 μm, PM10) also known as fine particles, have significant adverse effects on heart and lung health. Inhalation of these particles is strongly associated with an increased risk of cardiovascular and respiratory diseases. As the size of particles decreases, they are able to reach the smallest structures of the lungs, particularly the alveoli, where they can enter the bloodstream, thereby amplifying their harmful effects [23]-[26]. Inhalation of such dust has been linked to serious long-term health risks, including respiratory issues, cardiovascular diseases, and cancer [27]. It is stated that PM and other hazardous components emitted from damaged buildings, especially during earthquakes, can increase these health effects [28]. Elderly people, children, and people with chronic respiratory and cardiovascular diseases may be more sensitive to the adverse health effects of these toxic substances [29]. The demolition of buildings attracts the attention of people around, especially children (Figure 2). While the spectacle of collapsing structures can be fascinating, it's critical to acknowledge the potential hazards associated with this process. For vulnerable groups like children, exposure is even more concerning, as their respiratory systems are still developing. Therefore, implementing stringent dust control measures and ensuring proper safety protocols are essential to mitigate these risks and protect public health during demolition activities.

Figure 2. Children observing a building demolition, a source of PM pollution (Kahramanmaraş 2024)

Within the scope of this study, a survey was conducted to examine the opinions and suggestions of demolition workers (DO), environmental safety officers (SO), and individuals in the demolition area (IP) who were exposed to dust during the demolition of damaged buildings, regarding being affected by dust, taking protective measures and measures related to demolition, and the perceptual responses of all individuals in this category were evaluated.

The devastation caused by the earthquakes severely damaged infrastructure and affected local life, highlighting the urgent need to reconsider Türkiye strategies for coping with seismic risks. It also underscored the necessity of enhancing measures to protect air quality and public health during construction and demolition processes following earthquakes.

Exposure to respirable PM and other hazardous substances after natural disasters plays a critical role in determining potential public health and environmental risks in future disasters. The survey aimed to evaluate public awareness regarding the health risks posed by dust and particles from collapsed buildings and the precautions that should be taken to mitigate these risks. The data obtained will guide the planning of preventive measures for similar disaster situations in the future and contribute to raising public awareness on this issue. Studies of this nature are important for protecting public health and minimizing environmental risks after disasters. This research conducted in such cities as Şanlıurfa, Adıyaman, and Kahramanmaraş is also considered an important step in this direction.2. Material and Method

2.1. Study Area

The survey was conducted during the demolition of severely damaged buildings that the Ministry of Environment, Urbanization, and Climate Change decided to demolish in the provinces of Şanlıurfa, Adıyaman, and Kahramanmaraş which were affected by the February 6 earthquakes (Figure 3). In selecting the demolition area for the survey, priority was given to buildings located in residential areas and densely populated neighborhoods to better assess the potential impact of the demolition dust on a larger number of nearby residents.

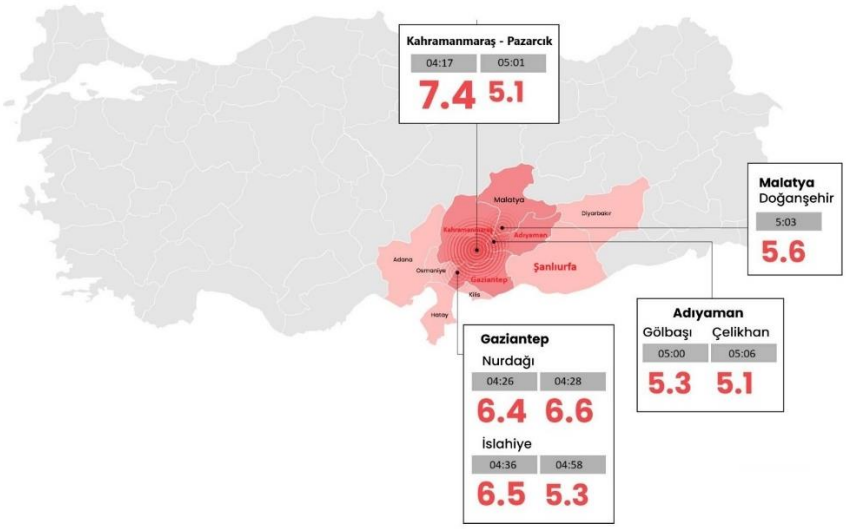


Figure 3. Showing the provinces affected by the Kahramanmaraş earthquakes on the map

2.2. Target Group

The survey targeted three groups: crane operators and other demolition workers at the demolition site (DO, 27 individuals), security personnel who ensured environmental security (SO, 28 individuals), and individuals who were not on duty at the time of the demolition but were in the vicinity (IP, 55 individuals).

2.3. Survey Application

Questions were designed to evaluate the perceptual effects of dust exposure during the demolition process on the target groups (DO, SO, and IP) in the region where the severely damaged buildings were demolished (Table 2). Participation in the survey was voluntary (Figure 4).

Figure 4. Conducting a survey in the severely damaged building demolition area.

2.4. Statistical Analysis

The ages of the participants were reported as mean ± SD (minimum-maximum) for each group. The average duration of stay in the demolition area was analyzed using ANOVA from the parametric tests, and group comparisons were conducted with the Bonferroni method in the post-hoc tests. The working hours in the sector and the duty hours in demolition activities of the DO and SO groups were compared with the t-test. The non-parametric Wilcoxon-Mann-Whitney U test was used to compare the responses given to the survey questions between the groups. The level of "$p<0.05$" was accepted as statistically significant.

3. Results and Discussion

The mean age of the participants in the survey was 40.0±9.9 (24-60 years old) in the DO group, 34.5±5.6 (25-43 years old) in the SO group, and 42.2±15.7 (18-73 years old) in the SO group. The genders of the participants were determined as 26 ♂ (man) and 1 ♀ (woman) in the DO group, 25 ♂ and 3 ♀ in the SO group, and 29 ♂ and 26 ♀ in the CG group. The education levels of the individuals in the groups participating in the survey were divided into primary, secondary school, high school, and university and the distribution of the education levels of each group is given in Figure 5.

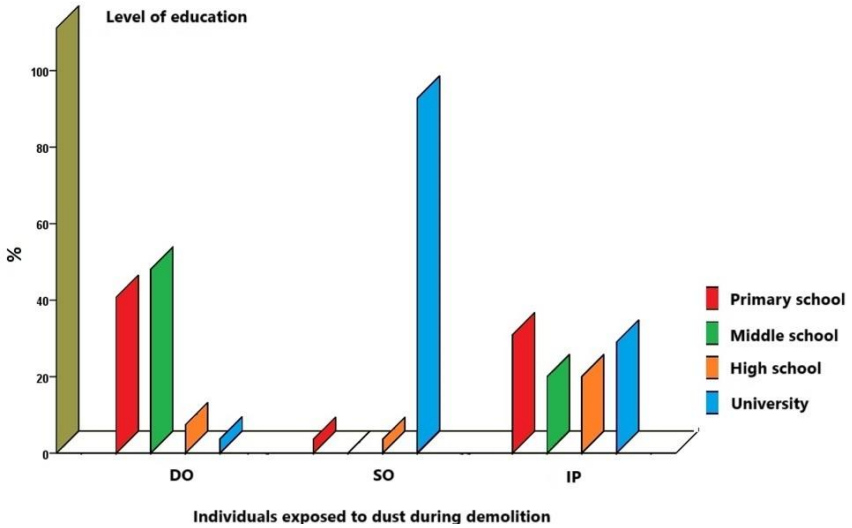


Figure 5. Distribution of education levels of individuals from the DO and SO groups and the IP group who were on duty during the demolition.

The average values obtained from the data regarding the working hours in the business line and the demolition activities of the DO and SO groups of the survey participants are shown in Figures 6 and 7.

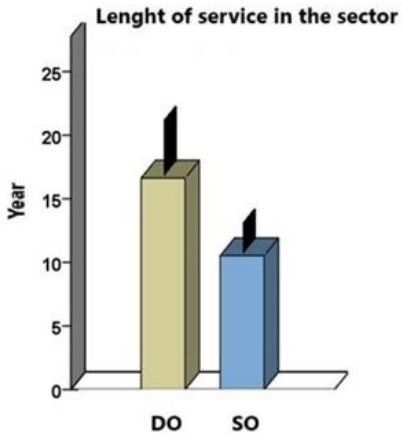


Figure 6. Working time of the DO and SO group individuals working in demolition in their job branches.

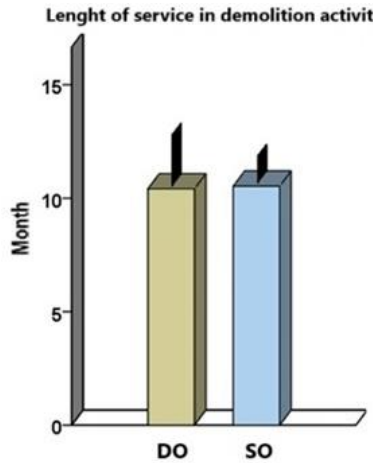


Figure 7. Working hours of the DO and SO group individuals working in the demolition activities.

The average values of the data indicating the average daily duration of stay in the demolition area of all survey participants are shown in Figure 8.

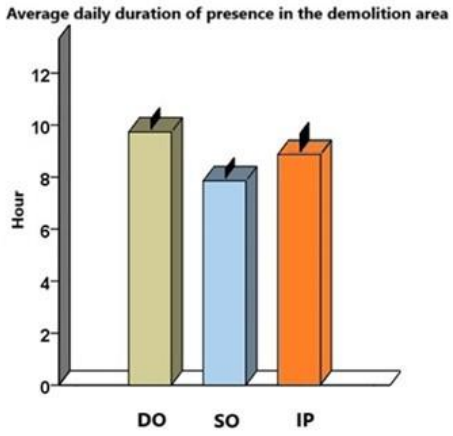


Figure 8. The average duration of stay in the destruction area of individuals in the DO, SO, and IP groups.

Participants' responses to the question regarding the perceptual determination of dust exposure levels during demolition were determined as low, medium, and high, and the obtained data are presented in Figure 9.

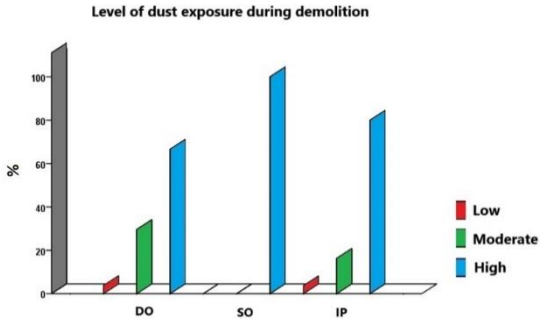


Figure 9. Proportional distribution of responses of survey participants showing the perceptual determination of dust exposure levels.

The responses given by the participants in the survey and the statistical analysis results of the differences between the groups are given in (Table 2). While preparing the table, only the "yes" responses of the participants were reflected, and all responses were used together in the statistical analyses.

Table 2. Proportional evaluation of the responses to the questions asked to the participants in the survey between the groups

Survey questions	Percentage of those who answered "Yes" (%)		
	DO	SO	IP
Does the dust generated during the demolition of the building cause difficulties in doing your job?	74,1	100,0 [a]*	89,1
Do you have any health problems related to your respiratory system?	33,3	35,7	38,2
Do you have any health problems due to the dust generated during the demolition of the building?	48,1	78,6 [a]*	70,9
Were the respiratory tract (nose, throat, etc.) affected by the dust generated during the demolition?	77,8	100,0 [a]*	90,9
Did you have a cough due to the dust generated during the demolition?	63,0	89,3 [a]*	80,0
Did the dust generated during the demolition affect your eyes?	85,2	100,0[a]*	85,5[c]*
Did you feel any difference in your skin due to the dust generated during the demolition?	70,4	92,9 [a]*	58,2[c]*
Did the building demolition affect your psychological state?	70,4	89,3	74,5
Do you smoke?	74,1	39,3 [a]*	47,3[b]*
Do you feel any difference in the level of dust affecting you if you smoke?	75,0	75,0	76,9
Did you feel the need to reduce the amount of cigarettes you smoked during the demolition?	70,5	83,0	80,8
Do you have information about the content of the dust generated during the demolition of the building?	3,7	57,1[a]**	23,6[b]* [c]*
Have you been informed about the risks and safety measures related to exposure to dust during the demolition of damaged buildings?	22,2	10,7	5,5[b]*
Do you know what precautions should be taken during the demolition of a building?	22,2	28,6	30,9
Have you been provided with adequate equipment such as goggles or a filtered respiratory mask?	14,8	10,7	3,6
Do you take care to use the protective equipment provided to you during the demolition of a building?	14,8	92,9[a]**	56,4[b]** [c]*
Do you believe that the dust-related equipment provided to you is sufficiently protective?	14,8	71,4[a]**	49,1[b]*
Do you think the precautions taken during the demolition of damaged buildings are sufficient?	22,2	0,0[a]*	12,7[c]*
Have you provided any feedback to any official institutions regarding the precautions taken during the demolition of damaged buildings?	18,5	17,9	27,3
Have you made any observations that you think should be taken into consideration regarding the disposal of dust generated during the demolition of buildings?	81,5	96,4	80,0
Have you informed the official institutions about this observation?	7,4	21,4	30,9[b]*

"a" shows the comparison between DO and SO, "b" shows the comparison between DO and ÇH, and "c" shows the comparison between SO and IP. "*" indicates $p<0.05$, "**" indicates $p<0.001$.

The demolition of buildings that have been severely damaged in earthquakes involves work that requires the utmost attention in terms of environmental safety. During these works, every individual, from the individuals in the vicinity to the demolition workers and the security personnel who ensure environmental safety, is affected to varying degrees by the dust generated during the demolition process. There is always a risk that these effects will cause health problems for all those involved. This

research, conducted in such cities as Şanlıurfa, Adıyaman, and Kahramanmaraş is also considered an important step in this direction. Therefore, understanding the effects of dust exposure on individuals and developing appropriate solutions will help mitigate the potential problems that may arise in the future. It will also contribute significantly to the implementation of predictive and more conscious approaches when similar situations occur. In our study conducted for this purpose, questions were asked to group individuals who are likely to be exposed to dust during building demolitions and to obtain information about their status of being affected by this exposure and the level of precautions taken. It is reported that the dust generated during demolition can cause significant harm to human health for those living and working in places close to demolition activities, along with the relatively harmful properties of the dust content [30]-[31]. Our study found that the negative impact of the dust generated during the demolition of damaged buildings on work performance was significantly higher in the SO group compared to the DO group and higher than in the IP group. It was found to be particularly striking that the DO group was found to be least affected by the dust. We can assume that since these individuals perform their duties in the demolition vehicles, it is estimated that this may be a result of these individuals being accustomed to such working conditions (Figure 6). Although the work branches were different between the DO and SO groups, there was no significant difference in the duration of their involvement in demolition activities (Figure 7). This suggests that both groups experienced similar levels of exposure to the dust generated during demolition activities. However, the stronger reactions observed among the SO group, despite their comparable experience in demolition work, may be attributed to other factors, such as their higher level of education, which might have increased their awareness and sensitivity to dust exposure (Figure 5).

While no significant difference was found between the groups when the individuals participating in our study were evaluated in terms of ongoing health problems related to the respiratory system in general, it was revealed that the effects of the respiratory system organs (nose and throat) due to the dust formed during the demolition occurred mostly in the SO group and were significantly lower in the DO group than in the SO group. In the IP group, it was determined that the level of this impact was close to that of the SO group. While a significantly higher rate of cough, eye, and skin problems was observed in all individuals in the SO group, these issues were found to be at a lower level in the DO group. In individuals in the IP group, these symptoms were observed at a higher level than in the DO group (Table 2). All these findings indicate that personnel involved in tasks related to environmental safety are in a more sensitive position in terms of dust exposure. It draws our attention that merely wearing a protective mask during SO's demolition duties may not be sufficient and that protective goggles or medications may be needed to protect the eyes from dust. We believe that this situation requires consultation with an ophthalmologist and that preventive measures should be arranged according to their recommendations. The same should be considered for skin sensitivity, and we believe that consulting a dermatologist and implementing additional protective measures for SO (such as protective creams, dust-impermeable clothing, etc.) would bring positive results. Another important aspect in terms of human health is human psychology. Disasters disrupt the flow of life and sometimes result in significant loss of life within the community. Among natural disasters, earthquakes, which have the most devastating impact, are considered the most traumatic events causing the greatest harm to society [32]-[33]. After the occurrence of two severe and destructive earthquakes, the beginning of the demolition process of damaged buildings causes new traumatic effects with new associations on individuals whose psychological state is already devastated. In our study, we observed that the psychological effects of building demolitions were at the highest level among individuals in the SO group in the demolition area. Although the psychological responses of individuals in the DO and IP groups were found to be at a lower level compared to the SO group, no significant difference was detected among the three groups (Table 2). This situation suggests that compared to the DO group, who had previously participated in similar activities due to their duties, the lesser familiarity of individuals in the SO group with the event may have been an important factor in them being more psychologically affected by the dramatically disturbing sight of building demolitions. Examination of the psychological effects of the post-earthquake destruction processes on these three separate groups will be of critical importance to increase the effectiveness of post-earthquake support programs. The more intense effects observed in the SO group indicate the need to provide psychosocial support to these individuals as a priority. At the same time, understanding the less obvious but potentially long-term psychological effects in the DO and IP groups will contribute to the comprehensive planning of support strategies. Additionally, providing moral

support to individuals who are psychologically affected will play a significant role [34].

While both smoking and dust exposure can lead to clinically significant respiratory dysfunction [35], it is also known that smoking itself is a major problem in our society. The numerous harmful effects of smoking, along with the differences in levels of habit due to people's socioeconomic status, present a separate social problem. It is not surprising that the smoking rate is higher in the DO group due to their lower education levels and living standards compared to the SO group. However, it was observed that the responses of smokers regarding both the difference in the level of impact from dust and the need to reduce the amount of smoking during demolition were surprisingly similar across all three groups. It can be said that there are similarities in the behavior and characteristics of smokers.

Individuals in the DO group stand out as having the least awareness regarding the content of the dust released during building demolition. It has been observed that the SO group has a significantly higher level of awareness on this subject compared to other groups. The knowledge of individuals in the IP group about the content of the dust released during demolition is also observed to be behind that of the SO group, similar to the DO group. This situation parallels individuals' education levels and suggests that individuals in the SO group might have consulted various sources due to curiosity sparked by their involvement in demolition activities. Although all three groups share the belief that dust exposure during demolition can lead to serious health problems in the future, this approach was found to be significantly higher in the SO group. All SO personnel were concerned that this exposure could lead to serious and permanent health problems for them in the future.

Although the SO group appears more sensitive to the risks of exposure to dust during demolition, the level of awareness regarding the risks of dust exposure during the demolition of damaged buildings and related safety measures was quite low in all groups. A very small amount of individuals in the DO and SO groups received information on this matter, and this important issue has been seriously overlooked. Similarly, it was observed that individuals from all groups had low awareness of the precautions that should be taken during building demolition. Despite the low level of information among the public in the demolition area, they were more knowledgeable about the precautions to be taken during demolition compared to the DO and SO groups. We think this may be related to the public paying more attention to warnings given in visual and written media. Additionally, the lack of adequate information from institutions about the risks of dust exposure and safety measures significantly increases the risk of future health problems, especially in the DO and SO groups.

The fact that a very small proportion of individuals, especially in the DO and SO groups, as well as in the IP group, were provided with equipment such as goggles or respiratory masks with filters shows that institutional support for protecting the health of individuals has been quite weak. It is noteworthy that only the individuals in the DO group were provided with goggles and masks with filters by their institutions, while a large proportion of the individuals in the SO group had to obtain this equipment on their own. Although the IP group individuals were not as active as the SO group, they showed more effort in acquiring and using this equipment on their own. The fact that demolition workers receive little support in terms of equipment and show weakness in acquiring and using protective equipment due to their own negligence lays the groundwork for the risk of many serious diseases, especially respiratory system diseases, due to dust exposure in the future.

The DO personnel approached the protective equipment provided against dust, even if limited, with skepticism and did not have much faith in their protective effect. This situation might be related to the lower education level of the DO group because the belief in the adequacy of the protective effect of the equipment was quite high among the SO group, which had a higher level of education. In the IP group, positive and negative opinions on this matter were expressed equally. This could also be related to the education level in the IP group, similar to the DO group, since nearly half of the DO group had high school and university education, while the other half had only primary or secondary school education. It is not surprising that individuals with lower levels of education have a weaker approach to protective equipment against dust, as believing in the reliability of protective equipment requires a sufficient informational background.

When the opinions of individuals consulted about the measures taken during demolition were considered, it was observed that there was significant distrust, especially within the SO group regarding these precautions. None of the participants in the SO group reported feedback indicating that the measures taken during demolition were sufficient, whereas there was a slight indication in the DO group that the measures could be sufficient. Individuals in the IP group provided low levels of feedback on the sufficiency of the measures taken during demolition.

In this situation, it is paradoxical that the feedback from individuals in all three groups to official authorities regarding the measures taken during demolition remained quite limited. Although the IP group showed slightly more sensitivity in this regard, the level of attempting to provide feedback to official authorities remained low. However, it was observed that individuals in all three groups had quite a few of their own unique ideas regarding the disposal of the dust. Despite having many observations in their minds about dust disposal, the rate of conveying these ideas to official authorities remained low in all groups, with the lowest being in the DO group. Notably, the performance of the SO personnel, particularly those with higher education levels, was also disappointingly low. This lack of feedback from the DO group is likely related to lower education levels, which may contribute to a lack of confidence in conveying their ideas to official authorities.

4. Conclusion and Suggestions

In conclusion, the education levels of individuals across the three categories had a significant impact on their physical and psychological responses to dust exposure from the demolition of damaged buildings, as well as their approaches and suggestions regarding protective measures. However, a general lack of confidence hindered their ability to effectively communicate their concerns and demands to official authorities.

As a result of our study, even when filtered masks were provided to the DO group, their proper usage and adherence to guidelines regarding when to use the masks were insufficient. Therefore, protective equipment (filtered masks, protective goggles, dust-proof clothing, gloves) should be provided and accompanied by training programs aimed at ensuring their effective use. These programs should cover the correct use, maintenance, and importance of the equipment. In addition, continuous monitoring of protective equipment usage in work areas is essential to enforce compliance and address any gaps in implementation. The protective equipment provided should not be generic but tailored to the specific needs of each worker through a personalized Personal Protective Equipment (PPE) plan. This plan should account for the worker's role, responsibilities, and working conditions, ensuring that the equipment is fit for purpose. Furthermore, the equipment should be designed to protect not only the workers but also the surrounding public, and a system should be established that enforces the mandatory use of this equipment.

Acknowledgment

We gratefully acknowledge the Scientific and Technical Research Council of Türkiye (TÜBİTAK) for financially supporting this study with project 1919B012317689. We would also like to thank the Presidency of Environment, Urbanization, and Climate Change of the Republic of Türkiye and the governorships of Şanlıurfa, Adıyaman, and Kahramanmaraş for their contributions and support to our study.

Contributions of the authors

Tuba Rastgeldi Dogan: Supervision, Review, Writing original draft preparation, Data Collection Final Approval. Ali Süzergöz: Literature Review, Data Collection, Data Analysis.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study complies with research and publication ethics.

References

[1] E. Işık, E. Harirchian, A. Büyüksaraç, and Y. Levent Ekinci, “Seismic and structural analyses of the Eastern Anatolian Region (Turkey) using different probabilities of exceedance,” *Appl. Syst. Innov.*, vol. 4, no. 4, p. 89, 2021.

[2] M. Sahin, E. Tari, “The August 17 Kocaeli and the November 12 Duzce earthquakes in Turkey,” *jstage*, vol. 52, no. 10, pp. 753–757, 2000.

[3] M. Erdik, Y. Kamer, M. Demircioğlu, and K. Şeşetyan, “23 October 2011 Van (Turkey) earthquake,” *Nat. Hazards* (Dordr.), vol. 64, no. 1, pp. 651–665, 2012.

[4] G. Dogan, A. S. Ecemis, S. Z. Korkmaz, M. H. Arslan, and H. H. Korkmaz, "Buildings damages after elazığ, turkey earthquake on January 24, 2020," *Nat. Hazards* (Dordr.), vol. 109, no. 1, pp. 161–200, 2021.

[5] H. E. Demirci, M. Karaman, and S. Bhattacharya, "A survey of damage observed in Izmir due to 2020 Samos-Izmir earthquake," *Nat. Hazards* (Dordr.), 2021.

[6] J. Xiao, Q. Deng, M. Hou, J. Shen, and O. Gencel, "Where are demolition wastes going: reflection and analysis of the February 6, 2023 earthquake disaster in Turkey," *Low-carbon Mater. Green Constr.*, vol. 1, no. 1, 2023.

[7] KOERI, Boğaziçi University, *Kandilli Observatory and Earthquake Research Institute*, 2023. [Online]. Available: http://www.koeri.boun.edu.tr. [Accessed: 08-Oct-2024].

[8] E. Işık, F. Avcil, A. Büyüksaraç, R. İzol, M. H. Arslan, C. Aksoylu, E. Harirchian, O. Eyisüren, E. Arkan, M. Ş. Güngür, M. Günay, and H. Ulutaş, "Structural damages in masonry buildings in Adıyaman during the Kahramanmaraş (Turkiye) earthquakes (Mw 7.7 and Mw 7.6) on 06 February 2023," *Engineering Failure Analysis*, vol. 151, 2023, doi: 10.1016/j.engfailanal.2023.107405.

[9] E. Altunsu, O. Güneş, S. Öztürk, S. Sorosh, A. Sarı, and S. T. Beeson, "Investigating the structural damage in Hatay province after Kahramanmaraş-Türkiye earthquake sequences, " *Engineering Failure Analysis*, vol. 157, no. 107857, 2024.

[10] I. Gkougkoustamos, P. Krassakis, G. Kalogeropoulou, and I. Parcharidis, "Correlation of Ground Deformation Induced by the 6 February 2023 M7. 8 and M7. 5 Earthquakes in Turkey Inferred by Sentinel-2 and Critical Exposure in Gaziantep and Kahramanmaraş Cities," *GeoHazards*, vol. 4, no. 3, pp. 267–285, 2023.

[11] A. I. Turan, A. Celik, A. Kumbasaroglu, and H. Yalciner, "Assessment of reinforced concrete building damages following the Kahramanmaraş earthquakes in Malatya, Turkey (February 6, 2023)," *Engineering Science and Technology, an International Journal*, vol. 54, 2023.

[12] M. Yetkin, I. Ö. Dedeoğlu, and T. U. N. Ç. Gülen, "Kahramanmaraş twin earthquakes: Evaluation of ground motions and seismic performance of buildings for Elazığ, southeast of Türkiye," *Soil Dynamics and Earthquake Engineering*, 2023.

[13] Türkiye Cumhuriyeti Cumhurbaşkanlığı Strateji ve Bütçe Başkanlığı, "Kahramanmaraş ve Hatay Depremleri Yeniden İmar ve Gelişme Raporu," https://www.sbb.gov.tr, 2024. [Online]. Available: https://www.sbb.gov.tr/wp-content/uploads/2024/02/Kahramanmaras-ve-Hatay-Depremleri-Yeniden-Imar-ve-Gelisme-Raporu-1.pdf. [Accessed: 25-Nov-2024].

[14] F. Wang et al., "Respiratory diseases are positively associated with PM2.5 concentrations in different areas of Taiwan," *PLoS One*, vol. 16, no. 4, p. e0249694, 2021

[15] H. Bayram, T. Rastgeldi Dogan, Ü. A. Şahin, and C. A. Akdis, "Environmental and health hazards by massive earthquakes," *Allergy*, vol. 78, no. 8, pp. 2081–2084, 2023.

[16] L. Sahakyan, N. Maghakyan, O. Belyaeva, G. Tepanosyan, M. Kafyan, and A. Saghatelyan, "Heavy metals in urban dust: contamination and health risk assessment: a case study from Gyumri, Armenia," *Arab. J. Geosci.*, vol. 9, no. 2, 2016.

[17] F. Azarmi and P. Kumar, "Ambient exposure to coarse and fine particle emissions from building demolition," *Atmos. Environ.* (1994), vol. 137, pp. 62–79, 2016.

[18] M. Normohammadi, H. Kakooei, L. Omidi, S. Yari, and R. Alimi, "Risk assessment of exposure to silica dust in building demolition sites," *Saf. Health Work*, vol. 7, no. 3, pp. 251–255, 2016.

[19] S. Mavroulis, M. Mavrouli, E. Lekkas, and A. Tsakris, "Managing earthquake debris: Environmental issues, health impacts, and risk reduction measures," *Environments*, vol. 10, no. 11, p. 192, 2023.

[20] "WHO global air quality guidelines: particulate matter (PM2. 5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide," *World Health Organization*, 2021.

[21] J. Castagna, A. Senatore, M. Bencardino, and G. Mendicino, "Concurrent influence of different natural sources on the particulate matter in the central Mediterranean region during a wildfire season," *Atmosphere (Basel)*, vol. 12, no. 2, p. 144, 2021.

[22] X. Zhang, L. Shen, and L. Zhang, "Life cycle assessment of the air emissions during building construction process: A case study in Hong Kong," *Renew. Sustain. Energy Rev.*, vol. 17, pp. 160–169, 2013.

[23] C. A. Pope 3rd et al., "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution," *JAMA*, vol. 287, no. 9, pp. 1132–1141, 2002.

[24] R. W. Atkinson, I. M. Carey, A. J. Kent, T. P. van Staa, H. R. Anderson, and D. G. Cook, "Long-term exposure to outdoor air pollution and the incidence of chronic obstructive pulmonary disease in a national English cohort," *Occup. Environ. Med.*, vol. 72, no. 1, pp. 42–48, 2015.

[25] I.-S. Kim et al., "Long-term fine particulate matter exposure and cardiovascular mortality in the general population: a nationwide cohort study," *J. Cardiol.*, vol. 75, no. 5, pp. 549–558, 2020.

[26] W. R. Wan Mahiyuddin, R. Ismail, N. Mohammad Sham, N. I. Ahmad, and N. M. N. Nik Hassan, "Cardiovascular and respiratory health effects of fine particulate matters (PM2.5): A review on time series studies," *Atmosphere (Basel)*, vol. 14, no. 5, p. 856, 2023.

[27] M. Zieliński, M. Gąsior, D. Jastrzębski, A. Desperak, and D. Ziora, "Influence of particulate matter air pollution on exacerbation of chronic obstructive pulmonary disease depending on aerodynamic diameter and the time of exposure in the selected population with coexistent cardiovascular diseases," *Adv. Respir. Med.*, vol. 86, no. 5, pp. 227–233, 2018.

[28] M. D. Petersen et al., "The 2023 US 50-state National Seismic Hazard Model: Overview and implications," *Earthq. Spectra*, vol. 40, no. 1, pp. 5–88, 2024.

[29] C. Wu, F. Lan, and B. Chen et al. "Respiratory symptoms among search and rescue workers who responded to the 2016 Taiwan earthquake *"Occupational and Environmental Medicine*, vol. 75, pp. 639–646, 2018.

[30] E. Cook, C. A. Velis, and L. Black, "Construction and demolition waste management: A systematic scoping review of risks to occupational and public health," *Front. Sustain.*, vol. 3, 2022.

[31] İ. Demir, "The use of demolition waste in concrete production and its effect on physical and mechanical properties," *AKÜ*, vol. 9, no. 2, pp. 105–114, 2009.

[32] A. C. McFarlane, "Family functioning and overprotection following a natural disaster: the longitudinal effects of post-traumatic morbidity," *Aust. N. Z. J. Psychiatry*, vol. 21, no. 2, pp. 210–218, 1987.

[33] B. Ozkan and F. C. Kutun, "Disaster psychology," *J Health Academicians*, vol. 8, pp. 249–256, 2021.

[34] M. Kerkez and M. E. Şanli, "Mediating role of spirituality in the relationship of anxiety, stress and depression with resilience in individuals exposed to earthquakes in Türkiye," *Int. J. Disaster Risk Reduct.*, vol. 104, no. 104347, p. 104347, 2024.

[35] W. M. Marine, D. Gurr, and M. Jacobsen, "Clinically important respiratory effects of dust exposure and smoking in British coal miners," *Am. Rev. Respir. Dis.*, vol. 137, no. 1, pp. 106–112, 1988.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1282-1292 YEAR: 2024
DOI: 10.17798/bitlisfen.1554402

Exploring Antioxidant and Genotoxic Activities of Silver Nanoparticles Synthesized from Karaerik Grape Leaves: A Green Approach

Gürhan BAYĞU[1], Deniz ALTUN ÇOLAK[2 *]

[1]*Erzincan Binali Yıldırım University, Institue of Natural and Applied Science, Erzincan, Türkiye*
[2]*Erzincan Binali Yıldırım University, Faculty of Arts and Sciences, Department of Biology, Erzincan, Türkiye*
(ORCID: 0000-0002-2699-0722) (ORCID: 0000-0002-3576-0355)

Keywords: Green Synthesis, Silver Nanoparticles, *Drosophila melanogaster*, Grape, Oxidative stress, Wing spot test.

Abstract

Nowadays, plant extract-mediated biosynthesis of nanoparticles has gained prominence as a pivotal research domain. Silver nanoparticles are traditionally synthesized using highly toxic and ecologically hazardous chemical and physical methods. The emerging green synthesis approach offers more eco-friendly alternatives while reducing production costs. Hence, the present study opted for a nature-friendly green synthesis method to produce silver nanoparticles. Silver nanoparticles were characterized using UV-visible spectroscopy (UV-VIS), scanning electron microscopy (SEM), X-ray diffraction (XRD), and Fourier transform infrared spectroscopy (FT-IR). The reaction involved the treatment of $AgNO_3$ (5 mM) with an aqueous extract of Karaerik grape leaf. Subsequently, the potentially toxic, genotoxic, and antioxidant effects of purchased chemically produced silver nanoparticles ($AgNP_{(c)}$) and silver nanoparticles ($AgNP_{(b)}$) that we synthesized using the green method were investigated on *Drosophila melanogaster* transheterozygous larvae. The study employed 72 ± 4 hour-old larvae bearing the recessive *flr*^{3} and *mwh* determinant genes on their third chromosomes. Exposure scenarios included 1 mM Ethyl methanesulfonate (EMS), AgNPs (1.25, 2.5, and 5 mg/mL), and EMS+ AgNPs. Interestingly, the EMS+ AgNPs combination reduced total oxidant status while increasing total antioxidant status significantly compared to EMS alone. To assess genotoxic effects, mutant trichomes resulting from genetic changes in the development of wing imaginal discs were examined. Furthermore, the AgNP synthesized through green synthesis demonstrated antioxidant properties and displayed no genotoxicity. In conclusion, the research highlights the promising potential of green-synthesized silver nanoparticles, which provide an eco-friendly and safe method for various applications.

1. Introduction

The main reason nano-sized structures attract attention is that they show functionality with their unique properties, different from their diverse volumetric dimensions and structures [1]. Medical-oriented nanoparticles are recognized for their ability to facilitate disease diagnosis and subsequently develop effective treatment methods. The fact that nanoparticles can be used in the medical field in this way provides benefits in many areas, such as gene therapies with nanospheres, earlier diagnosis and treatment of cancer, and targeted drug delivery [2]. Nanosilver is also widely used in biomedical fields such as infusion ports, contraceptive devices, endovascular stents, peritoneal dialysis devices, urological stents, endotracheal tubes, contact lens coatings, surgical and dental instruments, endoscopes, electrodes, subcutaneous cuffs [3]. Traditional methods for synthesizing nanoparticles often utilize highly toxic chemicals, resulting in harmful side effects during their application. Therefore, there is a need for an

* Corresponding author: daltun@erzincan.edu.tr
Received: 23.09.2024, Accepted: 20.11.2024

alternative approach to reduce these hazardous consequences. Green synthesis ensures the production of safe and effective nanoparticles while also enabling affordable and non-toxic manufacturing processes [4]. Concerns regarding the genotoxic and cytotoxic effects of metal nanoparticles, utilized in various applications and synthesized through physical, chemical, and biological methods, are rising on. Research is being conducted on the applications of gold and silver nanoparticles, which are classified as metal nanoparticles [5]. It is also known that silver in nanoparticle form, which is antimicrobial, is used more frequently than other metals and has low toxicity in mammalian cells [6]. The advantages of the green synthesis method, which utilizes plant extracts, include the easy availability of plants, the absence of special conditions, the biocompatibility of waste produced, their suitability for large-scale production, and their low cost and environmental friendliness [7, 8]. A substantial portion of present nanotoxicology research relies on *in vitro* models that do not provide insights into the destiny of nanoparticles within host organisms, including aspects such as metabolism, accumulation, biodistribution, elimination, persistence, and more [9]. D. *melanogaster* is an ideal model organism for conducting mutation screens and assessing the biological activity of various chemical compounds, including nanoparticles [10, 11].

In this context, we have used *Vitis vinifera* L. cv. Karaerik, a subspecies of the black grape cultivar known as Cimin grape, grown endemic in the Üzümlü district of Erzincan province, to determine the antioxidant and genotoxic effects of silver nanoparticles obtained from the leaves of Karaerik using the green synthesis method on *D. melanogaster*. The somatic mutation and recombination test (SMART), a fast and convenient method with many advantages, was used.

2. Material and Method

2.1. Strains of *Drosophila*

In our experiments, two different *Drosophila* strains were used: *multiple-wing hair* males (*mwh/mwh*) and *flare* females (*flr*^{3}/*In*(*3LR*)TM3, *ri p*^{p} *sep I*(*3*)*89Aabx*^{34e}). These strains were acquired from the repositories at Erzincan Binali Yıldırım University's Basic Sciences Application and Research Center Laboratories. The *multiple-wing hair* marker is located on chromosome 3, and its unique phenotype is multiple trichomes per cell when homozygous. The *flare* marker is recessive and also located on chromosome 3. This produces abnormal epidermal point-like hairs on the wing. The genetic symbols and descriptions have been detailed extensively in previous publications [12, 13].

2.2. Chemicals

Ethyl methanesulfonate (EMS, 62-50-0) and silver nitrate ($AgNO_3$, 7761-88-8) were obtained from Sigma Aldrich. Silver nanopowder ($AgNP_{(c)}$, 28-48nm) was purchased from Nanografi company.

2.3. Fresh on plant extract

Karaerik grape leaf extract was chosen to synthesize silver nanoparticles (AgNPs) due to its cost-effectiveness, medicinal properties, and easy accessibility, making it a practical and advantageous option for the synthesis process.

Fresh leaves of *Vitis vinifera* L. cv. Karaerik (Vitaceae) were collected from grape farms in Üzümlü, Erzincan, Turkey, in September 2018. Dr. Ali Kandemir expertly performed the taxonomic identification of the plant species using "Flora of Turkey and the East Aegean Islands" [14- 16], and all plant specimens were subsequently deposited in the Herbarium of the Science and Art Faculty at Erzincan Binali Yıldırım University. The grape leaves were thoroughly washed to remove dust particles and then air-dried at room temperature for one week to eliminate residual moisture. To reduce silver ions (Ag^{+}) to silver nanoparticles (Ag^{0}), a grape leaf extract was prepared by placing 5 g of the washed and dried finely-cut leaves into 100 mL of sterile distilled water. This mixture was continually heated at 80°C at ambient temperature for 1 hour, with regular agitation. After approximately 10 minutes of heating, the aqueous solution's color transitioned from clear to light yellow. After being cooled to room temperature, the extract was filtered using Whatman No. 1 paper and centrifuged at 3500 rpm for 5 minutes to remove residues. The final extract was then stored at room temperature for future experiments.

2.4. Green synthesis methods of silver nanoparticles

A 1 mM aqueous $AgNO_3$ solution was prepared with distilled water. Then, 10 mL of Karaerik grape leaf extract was added to a 90 mL solution with 1 mM $AgNO_3$. The first indication of AgNP

synthesis is the visual color change in the reaction mixture. In this study, the initially colorless reaction mixture gradually turned brown. After reducing the $AgNO_3$ solution to silver nanoparticles, the synthesized AgNPs were first centrifuged at 10.000 g for 10 minutes. After the centrifugation, it was lyophilized for characterization and experimental studies.

2.5. Characterization of green synthesized silver nanoparticles

UV-visible spectroscopy (PerkinElmer Lambda 35) was used to scan spectra from 200-700 nm with 1 nm resolution to validate optical measurements. Structural and morphological characterization was performed using SEM (FEI-Quanta FEG 450) with 120,000X magnification and 30 kV voltage. Elemental analysis via XRD (Panalytical Empyrean) determined the crystal structure and purity of AgNPs. FTIR spectra (Thermo Scientific Nicolet 6700) in the 400-4000 cm^{-1} range identified oxide forms and biomaterials on NP surfaces.

2.6. Survival rate determination

To determine the LD_{50} dose of AgNPs obtained by the green synthesis method ($AgNP_{(b)}$) and commercially purchased in synthetic form ($AgNP_{(c)}$), concentrations were determined according to the LD_{50} doses determined after 24 h applications at various concentration ranges. Adult individuals were obtained from third-instar transheterozygote larvae (72±4 h) treated with AgNPs to obtain an adequate number of flies for the antioxidant assay and SMART. The survival rate was determined by calculating the number of adult individuals per 100 larvae. Larvae were treated with EMS (1 mM), $AgNP_{(b)}$, and $AgNP_{(c)}$ (1.25, 2.5, and 5 mg/mL) separately and together. In the EMS + AgNP groups, only the highest concentrations of $AgNP_{(b)}$ and $AgNP_{(c)}$ were used.

2.7. Antioxidant assay

Total oxidant status (TOS) and total antioxidant status (TAS) were determined using photometric methods with commercial kits provided by Rel Assay (Rel Assay Kit Diagnostics, Turkey), developed by O. Erel [17]. To standardize these data, Trolox, a water-soluble version of vitamin E, was used as a TAS calibrator, while hydrogen peroxide was used as a TOS calibrator. The oxidative stress index (OSI) was determined as a percentage of the TOS level relative to the TAS level.

2.8. Somatic mutation and recombination test

The *mwh* virgin males and *flr*^{3} females of mutant strains were crossbred, and after 8 hours, the eggs were collected. Third-instar transheterozygote larvae (72 ± 4 h) obtained from these eggs were subjected to chronic treatment with 1.25, 2.5, and 5 mg/mL AgNPs. For the SMART, the eggs were nourished with *Drosophila* Instant Medium (Carolina Biological Supply Company, Burlington, NC, USA), which was prepared by dissolving 1.5 g of powder in 7 mL of water, and AgNP was incorporated into this medium at the specified final concentrations. Distilled water was utilized for the negative control; ethyl methanesulfonate (1 mM) served as a positive control. EMS is an alkylating chemical that may cause point mutations, small deletions, and chromosomal breaks; it also has recombinogenic activity [18, 19]. All treatments were repeated three times. These treatment flasks were maintained at 25 ± 1°C and a relative humidity of approximately 65% until the adult subjects hatched. Following the treatment phase, newly emerged adults were collected and preserved in 70% ethanol. Their wings were removed, mounted on glass slides, and fixed with Faure's solution (30 g acacia gum, 20 mL glycerol, 50 g chloral hydrate, and 50 mL water). The slides were then examined under a light microscope at 400X magnification.

2.9. Data analysis

The results were reported as mean values along with their standard errors (SE) or percentages, where appropriate. Duncan's one-way range test was used to determine the statistical significance of the data. The data was analyzed using SPSS software (version 24 for WindowsTM). The genotoxic potential in the Somatic Mutation and Recombination test was assessed by comparing the spot frequency (small single, large single, and twin spots) per fly in each treated group to that of the negative control. For the frequencies of spots per wing, a multiple-decision procedure is used to decide whether a result is positive, weakly positive, inconclusive, or negative. Statistical analysis was performed using a two-tailed χ^2 test for proportions ($\alpha = \beta = 0.05$) following the method described by H. Frei, and F. E. Würgler [20].

3. Results and Discussion

3.1. Characterization of synthesized silver nanoparticles

$AgNP_{(b)}$ characterization was performed using SEM equipped with EDX, XRD, UV-vis, and FT-IR techniques. The UV-vis absorbance spectra of $AgNP_{(b)}$ are presented in Figure 1.

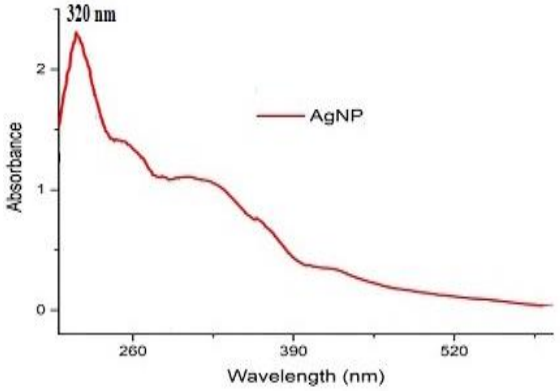


Figure 1. UV-visible absorption spectrum of $AgNP_{(b)}$

For the experiment, a 1 cm quartz cuvette was utilized. The study identified the maximum peak for $AgNP_{(b)}$ at 320 nm (Figure 1). Aromatic compounds from Karaerik grape leaf extract probably cause the peak at 320 nm.

The SEM images and the energy dispersive spectra of the AgNPs are shown in Figure 2. It is seen that AgNPs of different shapes were obtained in the case of various leaf extracts being used as reducing and capping agents. The extracts from *V. vinifera* gave rise to AgNPs that were mostly spherical and exhibited a polydisperse distribution. The size of the AgNPs ranged from 28.79 nm to 34.24 nm, with occasional instances of agglomeration being observed. This could be attributed to the quantity and nature of capping agents available in the various leaf extracts, as evidenced by the shifts and variations in the peak regions observed in the FTIR analysis.

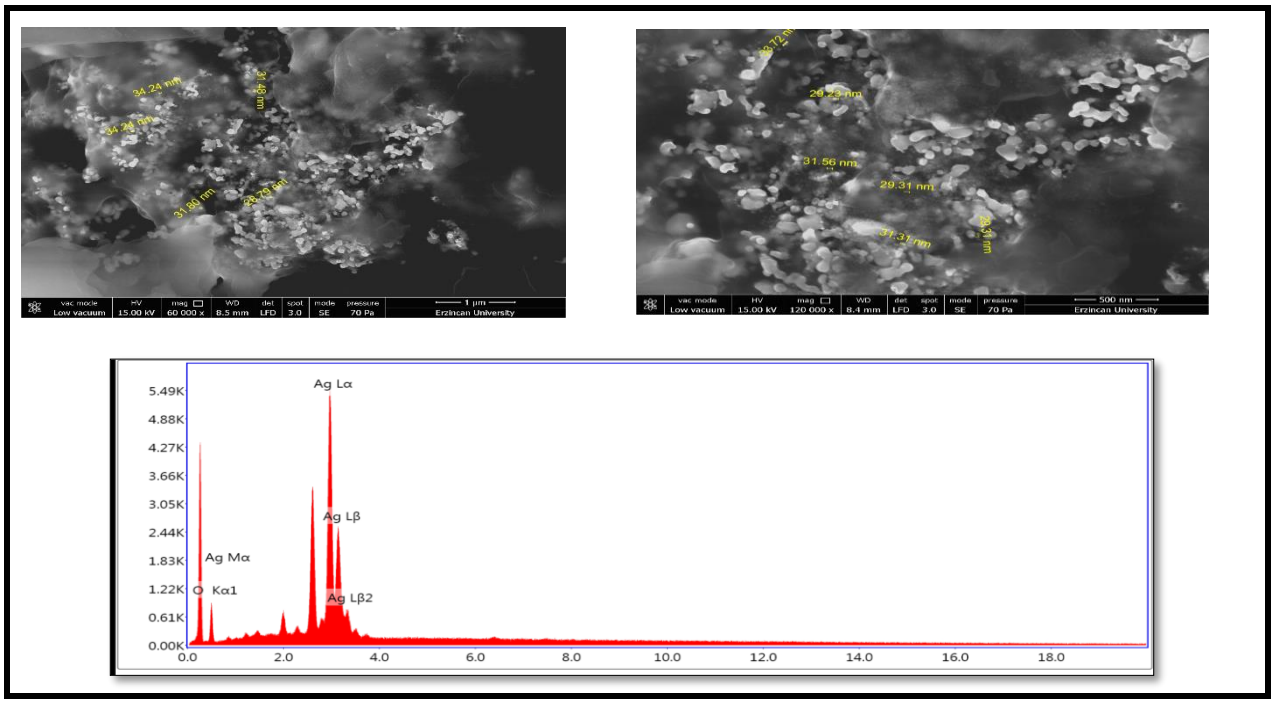


Figure 2. SEM and EDAX analysis of silver nanoparticles

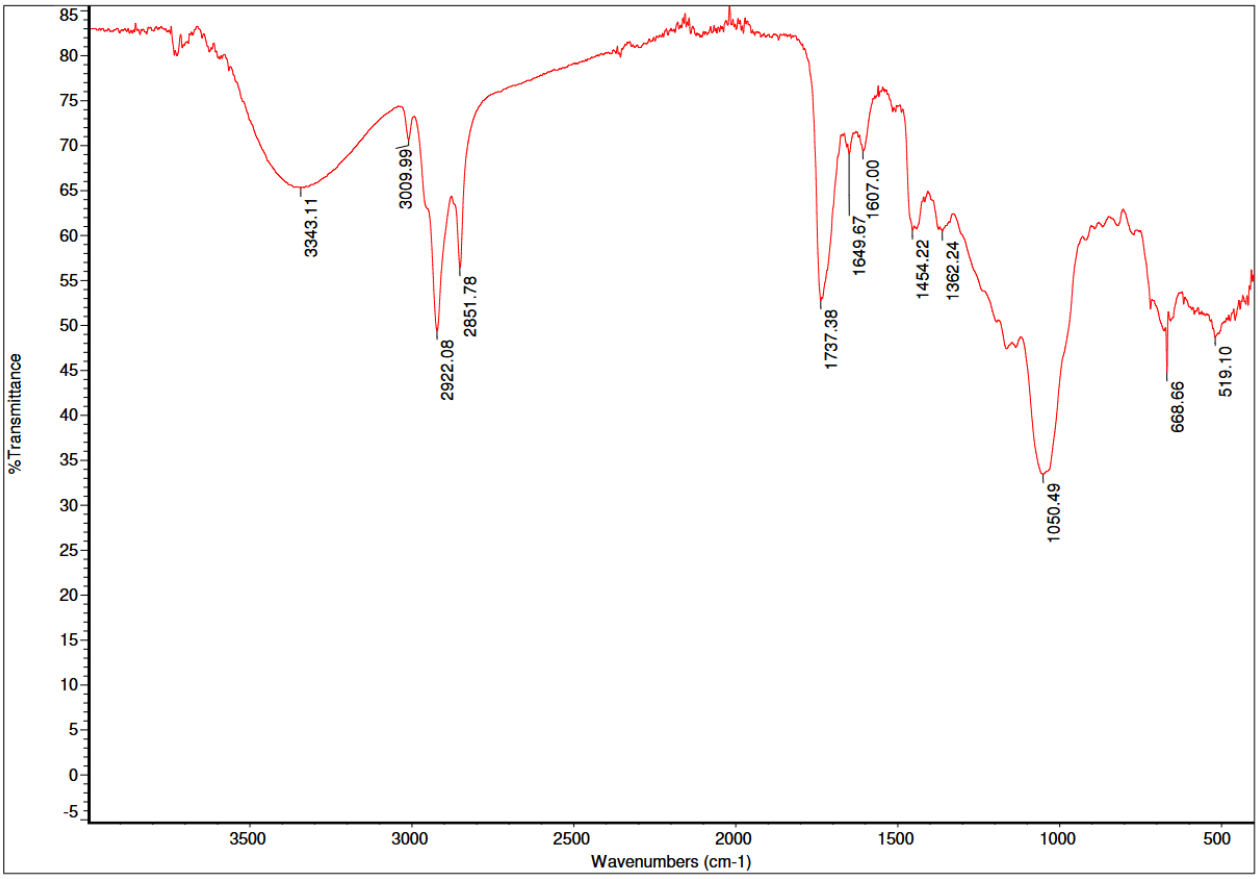


Figure 3. FTIR spectra of $AgNP_{(b)}$

FTIR analysis to characterize the AgNPs obtained from plant extract is shown in Figure 3. These distinctive bands suggest the presence of structures within the extract, primarily including polysaccharides/sugars, phenolic compounds, and proteins, with a particular emphasis on flavonoids.

Elemental analysis was conducted using energy-dispersive X-ray spectroscopy on the Panalytical, Empyrean. A powder X-ray diffraction (XRD) spectrum was used to identify the crystal shape and purity of AgNPs.The findings are depicted in Figure 4. No diffraction peaks associated with impurities were detected in the XRD patterns, indicating the product's high purity. Furthermore, the sharp and narrow nature of these peaks signifies that the crystallization of AgNPs is well-defined.

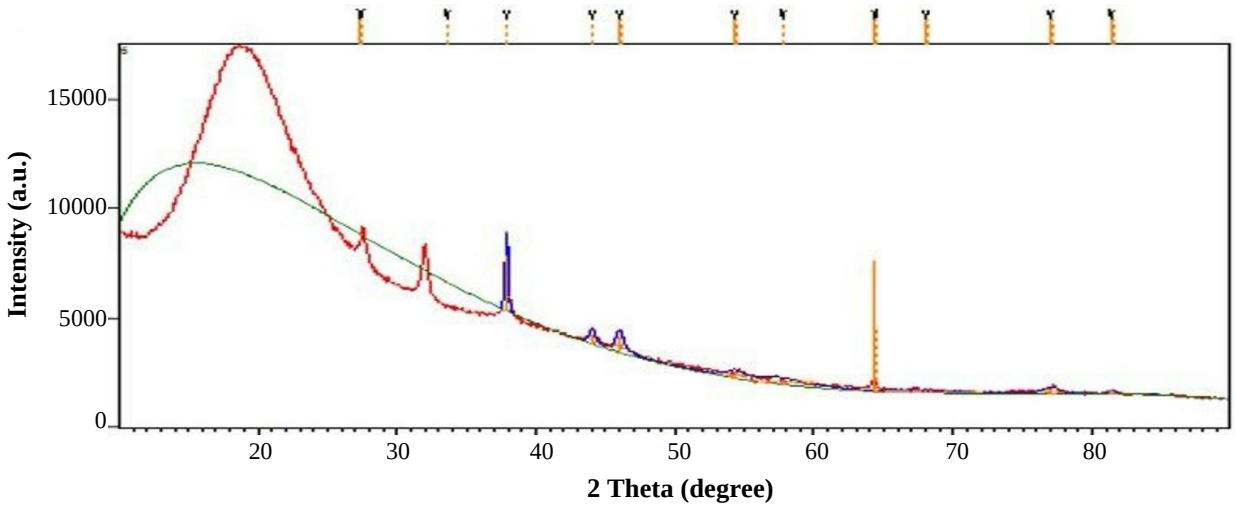


Figure 4. XRD patterns of $AgNP_{(b)}$

3.2. Survival rate determination

The data from the experiments, which aimed to investigate the impact of AgNP application on the survival rate, are presented in Figure 5. The findings showed a reduced survival rate in larvae treated with 1.25, 2.5, and 5 mg/mL AgNP extracts compared to the control group.

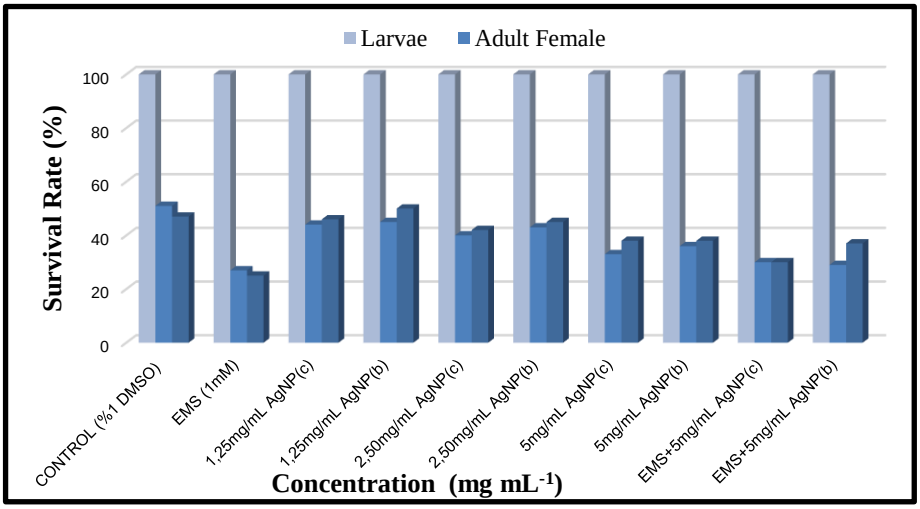


Figure 5. Percentage survival rates of $AgNP_{(c)}$ ve $AgNP_{(b)}$ in *D. melanogaster*

3.3. Antioxidant assay

The oxidative parameters total antioxidant status (TAS), total oxidant status (TOS), and oxidative stress index (OSI) were assessed in adult *D. melanogaster* to investigate the impact of AgNPs on transheterozygous larvae. All AgNP-treated groups showed higher TOS and lower TAS values when compared to the EMS group (Table 1). Results were statistically significant ($p<0.05$). In addition, a high OSI value (OSI=TOS/TAS) indicated adverse effects (Table 1).

Table 1. Oxidative stress parameter data were measured in *D. melanogaster* groups treated with $AgNP_{(c)}$ and $AgNP_{(b)}$

Experimental groups	TOS (μmol H_2O_2 Equiv. L^{-1})	TAS (mmol Trolox Eqvui. L^{-1})	OSI (AU)
Control (%1 DMSO)	11.15±3.51[a]	1.24±0.07[a]	8.95±2.11[a]
EMS (1 mM)	19.26±6.48[b]	0.07±0.01[b]	263.83±40.05[b]
1.25 mg/mL $AgNP_{(c)}$	11.08±4.12[a]	1.32±0.09[c]	8.33±3.42[a]
1.25 mg/mL $AgNP_{(b)}$	10.83±3.67[c]	1.41±0.10[d]	7.66±2.22[c]
2.5 mg/mL $AgNP_{(c)}$	9.28±1.16[d]	1.42±0.08[d]	6.49±1.54[d]
2.5 mg/mL $AgNP_{(b)}$	8.82±0.18[d]	1.50±0.02[e]	5.87±1.10[e]
5 mg/mL $AgNP_{(c)}$	7.75±0.21[d, e]	1.59±0.11[f]	4.84±0.23[f]
5 mg/mL $AgNP_{(b)}$	6.33±0.18[f]	1.67±0.07[g]	3.77±0.65[g]
EMS+5 mg/mL $AgNP_{(c)}$	13.14±2.15[g]	0.24±0.01[h]	54.29±7.17[h]
EMS+5 mg/mL $AgNP_{(b)}$	12.79±2.32[g]	0.29±0.02[i]	42.77±5.31[i]

[a-i]: Letters in the same column indicate significant differences at the 0.05 level.

3.4. Somatic mutation and recombination test

It was compared whether $AgNP_{(c)}$ and $AgNP_{(b)}$ had any genotoxic effects on *D. melanogaster* wings with *mwh/flr*^{3} and *mwh/TM3* genotypes at the determined doses. For this purpose, the results of $AgNP_{(c)}$ and $AgNP_{(b)}$ at different concentrations (1.25, 2.5, and 5 mg/mL) and the negative control group (distilled water) were statistically compared. From the data obtained, it was observed that the number of wing spots in the $AgNP_{(c)}$ and $AgNP_{(b)}$ application groups increased

in parallel with the increasing concentration compared to the negative control group (Table 2). When the results were analyzed statistically, the numerically observed increases were not significant and were evaluated as negative (-) ($p>0.05$). In addition, it was determined that $AgNP_{(b)}$ was not genotoxic, and the clone induction frequencies of $AgNP_{(c)}$ were higher than $AgNP_{(b)}$ at all concentrations.

Table 2. Genotoxicity of the $AgNP_{(c)}$ ve $AgNP_{(b)}$ in the *D. melanogaster* wing spot test results obtained with *mwh/flr*^{3} and *mwh/TM3* wings

Experimental groups		(N)	Small single spots (1–2 cells) (*m = 2*)			Large single spots (>2 cells) (*m = 5*)			Twin spots (*m = 5*)			Total *mwh* spots (*m = 2*)			Total spots (m = 2)			CIF
			No	Fr.	*D*	No	Fr.	*D*	No	Fr.	*D*	No	Fr.	*D*	No	Fr.	*D*	
mwh/flr^3	Distilled water	80	23	(0.28)		3	(0.03)		2	(0.03)		22	(0.27)		28	(0.35)		1.12
	EMS (1mM)	80	188	(2.35)	+	110	(1.37)	+	26	(0.32)	+	190	(2.37)	+	324	(4.05)	+	9.72
	$AgNP_{(c)}$ (1.25 mg/mL)	80	27	(0.33)	-	2	(0.03)	-	2	(0.03)	-	25	(0.31)	-	31	(0.38)	-	1.28
	$AgNP_{(k)}$ (2.5 mg/mL)	80	30	(0.37)	-	3	(0.03)	-	1	(0.01)	-	31	(0.38)	-	34	(0.42)	-	1.58
	$AgNP_{(c)}$ (5 mg/mL)	80	35	(0.43)	-	5	(0.06)	-	3	(0.03)	-	37	(0.46)	-	43	(0.53)	-	1.89
	$AgNP_{(b)}$ (1.25 mg/mL)	80	24	(0.30)	-	1	(0.01)	-	0	(0.00)	-	24	(0.30)	-	25	(0.31)	-	1.22
	$AgNP_{(c)}$ (2.5 mg/mL)	80	26	(0.32)	-	1	(0.01)	-	0	(0.00)	-	26	(0.32)	-	27	(0.33)	-	1.33
	$AgNP_{(b)}$ (5 mg/mL)	80	28	(0.35)	-	1	(0.01)	-	1	(0.01)	-	29	(0.36)	-	30	(0.37)	-	1.48
mvh/TM3	Distilled water	80	18	(0.22)		2	(0.03)					20	(0.25)		20	(0.25)		1.02
	EMS (1mM)	80	147	(1.83)	+	87	(1.08)	+				234	(2.92)	+	234	(2.92)	+	11.98
	$AgNP_{(c)}$ (1.25 mg/mL)	80	26	(0.32)	-	3	(0.03)	-				29	(0.36)	-	29	(0.36)	-	1.48
	$AgNP_{(b)}$ (2.5 mg/mL)	80	29	(0.36)	-	2	(0.03)	-	*			31	(0.38)	-	31	(0.38)	-	1.58
	$AgNP_{(c)}$ (5 mg/mL)	80	33	(0.41)	-	1	(0.01)	-				34	(0.42)	-	34	(0.42)	-	1.74
	$AgNP_{(b)}$ (1.25 mg/mL)	80	20	(0.25)	-	0	(0.00)	-				20	(0.25)	-	20	(0.25)	-	1.02
	$AgNP_{(c)}$ (2.5 mg/mL)	80	23	(0.28)	-	0	(0.00)	-				23	(0.28)	-	23	(0.28)	-	1.17
	$AgNP_{(b)}$ (5 mg/mL)	80	25	(0.31)	-	0	(0.00)	-				25	(0.31)	-	25	(0.31)	-	1.28

N: Number of wings, No: number of clones, Fr.: frequency, D: statistical diagnosis according to Frei and Würgler (1985), CIF: Frequency of clone formation per 10^5 cells, *: balancer chromosome *TM3* does not carry the *flr*^{3} mutation.+: positive, -: negative, i: inconclusive, m: multiplication factor, probability levels $\alpha=\beta=0.05$.

The results of our study suggest that AgNPs exhibit antigenotoxic effects in a dose-dependent manner, as evidenced by the oxidative stress parameters. Specifically, all AgNP-treated groups demonstrated higher TOS and lower TAS values than the EMS group, with statistically significant results ($p<0.05$). Our findings also indicate that biologically synthesized AgNPs produce more favorable results than chemically synthesized ones. This could be attributed to the fact that biologically synthesized AgNPs often exhibit better biocompatibility and reduced cytotoxicity, which might enhance their protective effects against genotoxicity. The presence of biomolecules in biologically synthesized AgNPs can enhance their stability and reduce potential adverse effects, leading to a more effective response in mitigating oxidative damage. This observation could be explained by the notion that low to moderate levels of oxidative stress may

induce adaptive responses in cells, leading to enhanced DNA repair mechanisms and increased antioxidant defenses. Therefore, while AgNPs increase oxidative stress markers, their presence may also stimulate cellular protective mechanisms that reduce the genotoxic impact compared to the control group.

The biological pathway for synthesizing nanoparticles has now been shown to save energy and produce less harmful waste [21]. By applying green principles in chemistry to nanotechnology, it is anticipated that nanotechnological processes can produce new products using environmentally friendly materials [22]. Plant metabolites and plant extracts are used in such processes, including peptides or biological macromolecules such as proteins, nucleic acids, carbohydrates, and lipids [23]. One of the objectives of this research is to emphasize the need for alternative methods to minimize the use of hazardous procedures in NP synthesis. The biological synthesis technique is one of the most often used ways of producing silver nanoparticles. This method employs traditional chemistry processes to reduce or eliminate the use and production of hazardous substances, benefiting both the environment and the economy. Chemical methods for synthesizing AgNPs utilize reagents that are non-toxic and environmentally friendly [24]. The green synthesis method is a new way to produce AgNPs [25]. Sahin and Gubbuk [26] analyzed the silver nanoparticles synthesized using the green synthesis method, employing turmeric, ginger, cinnamon, and carob extracts with antioxidant properties. That demonstrated that the obtained AgNPs serve as effective reducing agents. When silver interacts with the aromatic compounds in the extracts, it generates $AgNP_{(b).}$ This eco-friendly method not only minimizes the use of hazardous reagents but also showcases the potential to produce nanoparticles using sustainable materials [27]. Acay et al. [28] showed that AgNPs are effective against hospital pathogens by characterizing AgNPs produced by green synthesis from *Vitis vinifera* leaf extract by UV-vis spectrophotometer, SEM, EDX, FTIR, TGA-DTA, and XRD methods. According to Gnanajobitha et al. [29], fruit proteins and metabolites are integral to stabilizing silver ions throughout the synthesis of silver nanoparticles. Roy et al. [4] obtained spherical and crystalline AgNPs with an average size of 18-20 nm by green synthesis method using *Vitis vinifera* fruit extract in a study they conducted. It has been reported that these nanoparticles show effective antibacterial activity against *B. subtilis* and *E. coli*. Asaduzzaman et al. [30] investigated the effects of 17 nm silver nanoparticles synthesized by the green synthesis method using the Bangladesh grape plant on shrimp and *Ehrlich ascites* carcinoma cells. They found that although it had a toxic effect on shrimps, it showed strong antibacterial and anticancer activity in *E. ascites* carcinoma cells. Silver nanoparticles induce cytotoxicity and genotoxicity in a size- and coating-dependent manner, with smaller particles causing greater toxicity and genotoxicity than larger ones. These studies suggest that smaller silver nanoparticles (≤20 nm) are generally more cytotoxic than larger ones due to their higher cellular uptake and bioavailability [31, 32].

Drosophila melanogaster, employed as a living model organism, particularly in nanotoxicity, has recently been acknowledged as an appropriate organism for studying nanomaterial-induced toxicity [33]. Demir et al. [34] showed that when AgNPs were applied to 3rd instar larvae at concentrations that ranged from 0.1 to 10 nm, small yet notable increases in the frequency of total spots were observed in their genotoxic analysis studies using the *Drosophila* wing spot assay. In the present study, which showed that silver nanoparticles can induce genotoxic activity in the wing spot assay, it was also determined that it contrasted with silver nitrate, where negative findings were obtained. Gorth et al. [35] evaluated silver nanoparticles by examining their size and assessing the toxicity of both nanoscale and microscale silver particles in *Drosophila* egg development. While the pupalization rates of *Drosophila* eggs exposed to the 20-30 nm AgNPs did not show a statistically significant decrease, it was determined that this rate decreased in eggs exposed to the 500-1200 nm AgNPs. At 10 ppm AgNP exposure, pupae exposed to only the 20-30 nm AgNPs can mature very significantly, while it has been reported that they mature at a lower rate in the 500-1200 nm and 100 nm AgNP groups compared to the control. As a result, it has been noted that nanoscale AgNPs are less toxic to *Drosophila* eggs than micro-sized silver particles. In another *in vivo* study using the Somatic Mutation and Recombination Test (SMART), the genotoxicity of AgNPs on *D. melanogaster* was evaluated, with larvae exposed to 4.7 nm AgNPs at 25, 30, and 50 µg/mL and to 42 nm AgNPs at 250, 500, and 1000 µg/mL. The study found that speckle formation, which indicates mutagenic and recombinogenic effects in the phenotype, remained unchanged at

spontaneous frequencies for all doses, but AgNPs of both sizes caused pigmentation defects and reduced locomotor activity in adult flies [36].

4. Conclusion and Suggestions

In the present study, the comparative genotoxic evaluation of $AgNP_{(c)}$ and $AgNP_{(b)}$ was conducted on third-instar larvae that were exposed to NPs, and $AgNP_{(b)}$ did not exhibit any significant genotoxic activity in the wing spot assay. This study employed a straightforward, rapid, environmentally friendly, and cost-effective approach to synthesize silver nanoparticles using Karaerik grape leaves. The nanoparticles were characterized using FTIR, XRD, UV-vis spectroscopy, SEM, and EDAX. These Ag nanoparticles exhibited robust preventative properties, making them a valuable asset as potent antioxidants when functionalized with *V. vinifera*. These studies suggest that biologically synthesized silver nanoparticles (AgNPs) generally exhibit better biological activity and lower toxicity to living cells than chemically synthesized AgNPs. Regarding these attributes, the synthesized nanoparticles hold considerable promise for diverse applications within the biomedical sector. Our findings indicate that silver nanoparticles have substantial potential for use in numerous industries, particularly in the food and pharmaceutical sectors. Further studies are needed to elucidate the precise molecular pathways involved in this protective response and to confirm the superiority of biologically synthesized AgNPs over their chemically synthesized counterparts.

Contributions of the authors

Conceived and designed the experiments: Deniz A. Ç.; Performed the experiments: All authors; Writing-review & editing: All authors; Supervision & funding acquisition: Deniz A. Ç.

Acknowledgment

This study was supported by the Erzincan Binali Yıldırım University Research Foundation. The authors thank Erzincan Binali Yıldırım University for financial support for the project [Project Number = FYL-2019-613].

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] P. K. Jain, X. Huang, I. H. El-Sayed, and M. A. El-Sayed, "Noble metals on the nanoscale: optical and photothermal properties and some applications in imaging, sensing, biology, and medicine," *Acc Chem Res.*, vol. 41, pp. 1578-1586, May 2008.

[2] A. Mbonyiryivuze, I. Omollo, B. D. Ngom, B. Mwakikunga, S. M. Dhlamini, E. Park, and M. Maaza "Natural dye sensitizer for Grätzel cells: Sepia melanin," *Mater Chem Phys.*, vol. 3, pp. 1-6, June 2015.

[3] T. Faunce, and A. Watal, "Nanosilver and global public health: international regulatory issues," *Nanomed.*, vol. 5, pp. 617-632, June 2010.

[4] N. Roy, A. Gaur, A. Jain, S. Bhattacharya, and V. Rani, "Green synthesis of silver nanoparticles: an approach to overcome toxicity," *Environ Toxicol Pharmacol.*, vol. 36, pp. 807-812, November 2013.

[5] M. P. Patil, and G. D. Kim, "Eco-friendly approach for nanoparticles synthesis and mechanism behind antibacterial activity of silver and anticancer activity of gold nanoparticles," *Appl Microbiol Biotechnol.*, vol. 101, pp. 79-92, December 2017.

[6] A. R. Shahverdi, S. Minaeian, H. R. Shahverdi, H. Jamalifar, and A. A. Nohi, "Rapid synthesis of silver nanoparticles using culture supernatants of Enterobacteria: a novel biological approach," *Process Biochem.*, vol. 42, pp. 919-923, May 2007.

[7] O. V. Kharissova, H. R. Dias, B. I. Kharisov, B. O. Pérez, and V. M. J. Pérez, "The greener synthesis of nanoparticles," *Trends Biotechnol.*, vol. 31, pp. 240-248, April 2013.

[8] G. Benelli, and C. M. Lukehart, "Applications of green-synthesized nanoparticles in pharmacology, parasitology and entomology," *J Clust Sci.*, vol. 28, pp. 1-2, January 2017.

[9] X. Hu, S. Cook, P. Wang, and H. Hwang, "In vitro evaluation of cytotoxicity of engineered metal oxide nanoparticles," *Sci Total Environ.*, vol. 407, pp. 3070-3072, April 2009.

[10] G. M. Rubin, E. B. Lewis, "A brief history of *Drosophila's* contributions to genome research," *Science.*, vol. 287, pp. 2216-2218, March 2000.

[11] B. H. Jennings, "*Drosophila*- a versatile model in biology and medicine," *Mater Today.*, vol. 14, pp. 190-195, May 2011.

[12] U. Graf, F. E. Würgler, A. J. Katz, H. Frei, H. Juon, C. B. Hall, and P. G. Kale, "Somatic mutation and recombination test in *Drosophila melanogaster*," *Environ Mutagen.*, vol. 6, pp. 153-188, 1984.

[13] D. L. Lindsley, and G. G. Zimm, "The Genome of *Drosophila melanogaster*," *Academic Press*, San Diego, CA, 1992.

[14] P. H. Davis, R.R. Miller and K. Tan. 1965-1985. Flora of Turkey and the Aegean Islands. vol. 1-9, Edinburgh University Press, Edinburgh.

[15] P. H. Davis, R.R. Miller and K. Tan. 1988. Flora of Turkey and the Aegean Islands. vol. 10 (Supplement I), Edinburgh University Press, Edinburgh.

[16] A. Güner, N. Özhatay, T. Ekim and K.H.C. Başer. 2000. Flora of Turkey and The East Aegean Islands. Vol. 11 (Supplement II), Edinburgh University Press, Edinburgh.

[17] O. Erel, "A new automated colorimetric method for measuring total oxidant status," *Clin Biochem.*, vol. 38, pp. 1103-1111. December 2005.

[18] G. A. Sega, "A Review of the genetic effects of ethyl methanesulfonate," *Mutat Res.*, vol. 134, pp. 113-142, September–November 1984.

[19] R. Socha, and F. Marec, "Genotoxicity of the anti-juvenile hormone agent precocene II as revealed by the *Drosophila* wing spot test," *Mutagenesis*, vol. 4, pp. 216-220, May 1989.

[20] H. Frei, and F. E. Würgler, "Statistical methods to decide whether mutagenicity test data from *Drosophila* assays indicate a positive, negative, or inconclusive result," *Mutat Res-Envir Muta.*, vol. 203, pp. 297-308, August 1988.

[21] N. A. Begum, S. Mondal, S. Basu, R. A. Laskar, and D. Mandal, "Biogenic synthesis of Au and Ag nanoparticles using aqueous solutions of Black Tea leaf extracts," *Colloids Surf B Biointerfaces.*, vol. 71, pp. 113-118, June 2009.

[22] A. García-Quintero, and M. Palencia, "A critical analysis of environmental sustainability metrics applied to green synthesis of nanomaterials and the assessment of environmental risks associated with the nanotechnology," *Sci Total Environ.*, vol. 793, pp. 148524, November 2021.

[23] J. Jeevanandam, S. F. Kiew, S. Boakye-Ansah, S. Y. Lau, A. Barhoum, M. K. Danquah, and J. Rodrigues, "Green approaches for the synthesis of metal and metal oxide nanoparticles using microbial and plant extracts," *Nanoscale.*, vol. 14, pp. 2534-2571. 2022.

[24] A. E. Fadiji, P. E. Mortimer, J. Xu, E. E. Ebenso, and O. O. Babalola, "Biosynthesis of nanoparticles using endophytes: a novel approach for enhancing plant growth and sustainable agriculture," *Sustainability.*, vol. 14, pp.10839, August 2022.

[25] M. A. Huq, M. Ashrafudoulla, M. M. Rahman, S. R. Balusamy, and S. Akter, "Green synthesis and potential antibacterial applications of bioactive silver nanoparticles: A review," *Polymers.*, vol. 14, pp. 742, February 2022.

[26] M. Sahin, and I. H. Gubbuk, "Green synthesis of palladium nanoparticles and investigation of their catalytic activity for methylene blue, methyl orange, and rhodamine B degradation by sodium borohydride," *Reac Kinet Mech Cat.*, vol.135, pp. 999-1010, February 2022.

[27] A. M. Awwad, and N. M. Salem, "Green synthesis of silver nanoparticles by Mulberry Leaves Extract," *J Nanosci Nanotechnol.*, vol. 2, pp. 125-128, 2012.

[28] H. Acay, and M. F. Baran, "Investigating antimicrobial activity of silver nanoparticles produced through green synthesis using leaf extract of Common grape (*Vitis vinifera*), "*Appl Ecol Environ Res.*, vol. 17, pp. 4539-4546, February 2019.

[29] G. Gnanajobitha, K. Paulkumar, M. Vanaja, S. Rajeshkumar, C. Malarkodi, G. Annadurai, and C. Kannan, "Fruit-mediated synthesis of silver nanoparticles using *Vitis vinifera* and evaluation of their antimicrobial efficacy," *J Nanostruct Chem.*, vol. 3, pp. 1-6, August 2013.

[30] A. K. M. Asaduzzaman, B. S. Chun, and S. R. Kabir, "*Vitis vinifera* assisted silver nanoparticles with antibacterial and antiproliferative activity against *Ehrlich ascites* carcinoma cells," *J Nanopart.*, vol. 2016, pp. 6898926, August 2016.

[31] X. Guo, Y. Li, J. Yan, T. Ingle, M. Jones, N. Mei, M. Boudreau, C. Cunningham, M. Abbas, A. Paredes, T. Zhou, M. Moore, P. Howard, and T. Chen. "Size- and coating-dependent cytotoxicity and genotoxicity of silver nanoparticles evaluated using in vitro standard assays. " *Nanotoxicology*, vol. 10, pp.1373 - 1384. August 2016.

[32] T. Kumkoon, M. Srisaisap, and P. Boonserm, (2023). "Biosynthesized silver nanoparticles using *Morus alba* (white mulberry) leaf extract as potential antibacterial and anticancer agents. " *Molecules, vol.* 28(3), pp. 1213. January 2023.

[33] C. Ong, L. Y. L. Yung, Y. Cai, B. H. Bay, and G. H. Baeg, "*Drosophila melanogaster* as a model organism to study nanotoxicity," *Nanotoxicology.*, vol. 9, pp. 396-403, July 2015.

[34] E. Demir, G. Vales, B. Kaya, A. Creus, and R. Marcos, "Genotoxic analysis of silver nanoparticles in *Drosophila*," *Nanotoxicology.*, vol. 5, pp. 417-424, November 2010.

[35] D. J. Gorth, D. M. Rand, and T. J. Webster, "Silver nanoparticle toxicity in *Drosophila*: size does matter," *Int J Nanomedicine.*, vol. 6, 343-350, February 2011.

[36] A. Ávalos, A. I. Haza, E. Drosopoulou, P. Mavragani-Tsipidou, and P. Morales, "In vivo genotoxicity assessment of silver nanoparticles of different sizes by the Somatic Mutation and Recombination Test (SMART) on *Drosophila*," *Food Chem Toxicol.*, vol. 85, pp.114-119. November 2015.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1293-1303 YEAR: 2024
DOI: 10.17798/bitlisfen.1556171

Assessing the Effectiveness of Machine Learning Techniques for Silver Price Prediction: A Comparative Study*

Erhan ERGİN[1**], Binali Selman EREN[2]

[1]*Bitlis Eren University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Bitlis, Türkiye.*
[2]*Bitlis Eren University, Vocational School of Tatvan, Department of Accounting and Tax, Bitlis, Türkiye.*
(ORCID: 0000-0001-6281-3654) (ORCID: 0000-0001-5136-6406)

Keywords: Silver Price, Financial Prediction, Machine Learning, Prediction Algorithms.

Abstract
Silver is considered an important asset in terms of economic indicators and a valuable investment asset in terms of the markets. Therefore, determining silver prices is critically important for both national economies and investors. However, the non-stationary and non-linear nature of silver prices makes predicting price movements challenging. The methods used for predicting silver prices must be suitable for capturing these volatile and complex behavioral characteristics. The silver market can be influenced by other commodities and investment assets. Factors affecting silver prices, such as gold prices, Brent crude oil prices, the US Dollar index, the VIX index, and the S&P 500 index, can play a significant role. In this context, these variables have been used as inputs for predicting silver prices in the study. Three different models have been developed to predict the prices one, two, and three days ahead. These models have been predicted using four different machine learning methods: linear regression, support vector regression (SMOReg), k-nearest neighbors (k-NN), and random forest (RF). The results show that the random forest and k-NN methods exhibit the highest performance. The random forest achieves the highest accuracy in the first two models, while k-NN excels in the third model. Linear regression and SMOReg methods are less successful compared to the others. Consequently, it can be concluded that random forest and k-NN methods can be preferred for long-term predictions, and that these results may provide valuable insights, especially for investors and decision-makers.

1. Introduction

For many years, individuals have utilized precious metals as a means of saving and investing, thereby making them a crucial component of the global financial system. The financialization of commodities markets permitted investors to diversify their portfolios and resulted in a considerable increase in investments associated with precious metals [1]. In both developed and developing economies, financial products employ precious metals as indispensable assets for industrial processes. Consequently, commodities play an important role in economic growth and development. Precious metals serve as a reliable source of stability during periods of political and economic uncertainty. Gold, silver, platinum, and palladium, in particular, are essential inputs in industrial production. This has led to a heightened interest among investors in these metals within the context of financial markets.

* This paper is an expanded and revised version of the abstract presented as an oral presentation at the 2nd International Rahva Conference on Technical and Social Research, held on December 3-4, 2022.

** Corresponding author: eergin@beu.edu.tr
Received: 25.09.2024, Accepted: 17.11.2024

Silver is a precious metal of considerable importance to both the global financial system and the industrial production system. In addition to its status as an investment vehicle and a store of value, silver is employed as a precious metal in a number of other contexts, including industrial production, art, medicine, chemistry, and photography. In addition to its role as a store of value against economic uncertainty, inflation, and exchange rate fluctuations, silver provides portfolio diversification for investors and facilitates trading transactions due to its high liquidity. These features have made silver a valuable financial instrument for both individual and institutional investors [2].

Throughout history, the price of silver has been influenced by a variety of economic and social factors. Historically, medieval Europe relied heavily on silver, a precious metal, as an essential economic resource. The Industrial Revolution of the nineteenth century and the "Silver Break" of 1873 had a significant impact on the price of silver. In the 20th century, global political and economic developments rendered silver a valuable asset. During the "Silver Crisis" of 1980, extensive silver purchases led to regulations that artificially inflated prices, ultimately causing a rapid decline. This crisis revealed the effects of speculative bubbles and market manipulations, leading to tighter financial regulations. Subsequently, silver prices increased significantly due to factors such as economic uncertainties, inflation concerns, rising investment demand, increased industrial usage, and the impact of the 2008 global economic crisis. In recent years, global financial crises and the COVID-19 pandemic have impacted the value and price of silver. Figure 1 also shows the historical fluctuations in silver prices.

Figure 1. Silver prices (1970-2024)

Source: tradingeconomics.com

Because of these characteristics, market participants closely monitor silver prices. Predicting silver prices is crucial for trading decisions, portfolio management, industrial production planning, economic indicators, and the overall health of financial markets. The purpose of this study is to predict the price of silver using a variety of variables, including the US Dollar Index, VIX Index, S&P 500 Index, Brent Crude Oil, and gold prices. The central research question is to identify the most effective machine learning method for predicting silver prices with the minimum prediction error. Previous studies in this field have employed a range of methods, including linear regression, support vector regression (SMOReg), random forest, and K-Nearest Neighbors (k-NN), for financial prediction. A review of the literature reveals a limited number of studies on the use of machine learning methods for predicting silver prices. It is expected that the techniques and variables utilized in this study will enhance and broaden the existing body of research. The structure of the subsequent sections is as follows: The second section provides a summary of relevant literature; the third section details the dataset and method explanations; the fourth section presents the experimental results; and the final section discusses the conclusions and implications.

2. Literature Review

In recent years, machine learning techniques have been extensively utilized in the field of financial prediction. The literature in this field is growing rapidly, with new studies contributing to a continually expanding body of knowledge. A significant number of studies have employed machine learning techniques to predict the prices of precious metals. However, it has been observed that the number of studies that predict silver prices with machine learning methods is limited. This section presents an overview of studies that predict silver prices and prices of other precious metals using machine learning methods, with a particular focus on studies conducted in recent years.

Çelik and Başarır (2017) predict the prices of precious metals such as gold, silver, platinum, and palladium using artificial neural networks (ANN). They use five performance metrics to assess the accuracy of the predictions. The study analyzes five years of various financial data. The results show that the error rates in market predictions are at acceptable levels [3].

Goel et al. (2022) use machine learning methods to predict gold and silver prices in the Indian market. They conducted the analysis using CNN and

CNN-RNN hybrid models, utilizing data from January 2021 to August 2022. The results show that the RNN model is only successful in gold price prediction, while other models generally provide acceptable accuracy [4].

Üntez and İpek (2022) predict silver prices using LSTM and ARIMA deep learning methods. The results of the study reveal that the ARIMA method shows more realistic predicts [5].

Öndin and Küçükdeniz (2023) predict silver prices using Google Trends data and the Latent Dirichlet Allocation (LDA) method. They predict silver prices using random forests, gaussian processes, support vector machines, regression trees, and artificial neural networks. The regression tree method yields the most successful predict results [6].

Alparslan and Uçar (2023) use machine learning methods to examine gold and silver commodity price predictions for the COVID-19 pandemic period. The performance of the models is evaluated using MAE, MAPE, and RMSE metrics. The findings show that the Long Short-Term Memory (LSTM) model provides more accurate predictions in the pre-COVID-19 period, while Support Vector Regression (SVR) provides the best predict results for gold and LSTM for silver in the COVID-19 period [7].

Wang et al. (2023) develop an innovative model to predict silver prices. This model includes conventional neural networks (CNNs), the self-attention mechanism (SA), and the new gated unit (NGU). The results show that the CNN-SA-NGU model outperforms other models with a MAE of 87.90, the explained variance score (EVS) of 0.97, the r-squared (R^2) of 0.97, and a training time of 332.78 seconds [8].

Gono et al. (2023) use the XGBoost machine learning method to predict silver prices. They develop two main models to predict silver price fluctuations. In the first model, it predicts that prices will fall on the first two days, rise on the third day, fall again on the fourth day, and rise on the fifth and sixth days. The first model performs best with a MAPE value of 5.98% and a RMSE value of 1.6998. In the second model, it predicts that prices will fall on the first three days, then rise until the sixth day. The model's RMSE value is 1.6967, and its MAPE value is 6.06%. The proposed models perform best when compared to other ensemble models such as CatBoost and random forest [9].

Gür (2024) evaluates the performance of a hybrid model consisting of CNN, LSTM, GRU, and a combination of these models for silver price prediction. The models are trained with historical silver price data, and their prediction accuracies are compared. The findings show that the CNN-LSTM-GRU hybrid model produces more successful predictions [10].

Jin and Xu (2024) use 13 years of data to predict silver prices. In the study, models were developed using Gaussian process regression and Bayesian optimization techniques, and relatively high-level predictions were determined. For the silver price, RRSE was calculated as 0.2257%, RMSE as 0.0515, MAE as 0.0389, and correlation coefficient as 99.967% [11].

3. Material and Method

Machine learning is a subfield of artificial intelligence that enables computers to learn and act in a manner analogous to humans, with the assistance of algorithms and data. One of the fundamental tenets of machine learning is to achieve enhanced performance by minimizing human involvement [12]. In this context, machine learning empowers computers to improve their performance and make informed judgments through data acquisition. The main goal is to create models using data to generate forecasts and predictions. Machine learning algorithms construct a model by studying data and generating predictions based on it. The problem's specific attributes and data collection features inform algorithm selection and use. The field of machine learning has experienced significant advancements over time, particularly in the area of data interaction and modeling methods. From its earliest stages to its present state, machine learning has undergone a notable evolution, progressing from its basic forms to the more sophisticated deep learning techniques we observe today. These advancements have enabled the efficient application of machine learning across various fields. Machine learning provides an extensive array of applications in the financial industry, significantly contributing to the resolution of diverse financial issues. Machine learning methods are also extensively applied in the financial field. The analytical and predictive functionalities it offers in domains including credit scoring, algorithmic trading, portfolio management, and market predicting facilitate more effective and efficient financial decision-making [13].

Many algorithms have been developed for machine learning. Some of these algorithms are particularly suitable for financial predicting and have high performance, as reported in the literature. This study employed linear regression, support vector regression (SMOReg), random forest, and K-Nearest Neighbor (k-NN) methods commonly utilized in the literature to predict silver prices [14].

Linear regression is a statistical method used to establish a relationship between a dependent variable and one or more independent variables. The purpose of linear regression is to find the best-fit line representing the linear relationship between the dependent variable and the independent variables [15].

In the simple linear regression algorithm, a single attribute is used to predict the response that emerges from this attribute. It is assumed that these two variables are linearly related. Therefore, a function is tried to be found that predicts the response value (y) as accurately as possible as a function of the attribute or independent variable (x) [16].

$$h(x_i) = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (1)$$

where,
$h(x_i)$ represents the predicted response value for the i th observation.
β_0 is regression coefficient (intercept/constant)
β_1 is regression coefficient (slope)
ε_i, represents the error term.

Support Vector Machines (SVM) are algorithms that can be used to solve both classification and regression problems [17]. SVM was proposed for solving classification problems in the study conducted by Cortes and Vapnik [18]. Smola and Scholkopf [19] and Shevade et al. [20] developed the Support Vector Regression method to address regression problems involving SMOs [21]. Regression analysis is a model that deals with the relationship between variables in the prediction problem. SMOreg addresses the chaotic and inefficient issues found in standard SMO and utilizes optimization guidelines for binary problems. By employing two threshold parameters, SMOreg operates more efficiently than the original SMO. The SMOreg algorithm further enhances the convergence value observed in SMO, thereby demonstrating improved efficiency and performance [22]. SMOreg uses structural risk minimization constraints as a model, which makes it a reliable way to predict regression and deal with small-sample data that is not linear. Consequently, SMOreg is capable of accurately predicting time series data [23].

The k-Nearest Neighbors (k-NN) algorithm is a method used for classification problems. When k-NN classifies a new data point, it does so based on the labels of the k nearest neighbors in the existing dataset [24]. Additionally, regression problems commonly use it due to its low computational cost and ease of interpretation. In regression tasks, k-NN determines the predicted value of a new data point by calculating the average of the target values of the k nearest neighbors. This process involves calculating the distances between data points, selecting the k nearest neighbors, and averaging the target values of these neighbors.

Random Forest (RF) is a machine learning technique that utilizes multiple decision trees and is employed in both regression and classification problems [25]. The Random Forest algorithm is a bagging-based machine learning method [26]. This algorithm constructs multiple decision trees to improve accuracy and stability by combining them. In Random Forest, each tree's growth increases the degree of randomness in the model. Specifically, rather than searching for the best feature during node splits, the best feature is selected from a randomly chosen subset of features. This approach enhances model diversity, resulting in improved overall performance [27].

3.1. Evaluation Metrics

Correlation Coefficient: The correlation coefficient is a statistical measure that quantifies the direction and strength of the relationship between a dependent variable and one or more independent variables. The correlation coefficient can assume a value between -1 and 1. A correlation coefficient of -1 indicates an inverse relationship between the two variables; a correlation coefficient of 0 indicates the absence of a relationship between the two variables; and a correlation coefficient of 1 indicates a perfect relationship between the two variables [28].

Mean Absolute Error (MAE): Mean absolute error attempts to find out how far the predicted values are from the actual values. The formula for Mean Absolute Error (MAE) is:

$$MAE = \frac{1}{n}\sum_i^n |P_i - A_i| \quad (2)$$

where:
n is the number of observations.
P_i represents the actual values.
A_i represents the predicted values.
$|P_i - A_i|$ is the absolute error for each observation.
In essence, MAE represents the average of the absolute differences between the actual and predicted values, thereby providing a measure of the degree to which the model's predictions align with the actual data.

Root Mean Square Error (RMSE): The root mean square error (RMSE) is a statistical measure that quantifies the discrepancy between the predicted and actual values. It provides a means of assessing the magnitude of the error in estimating the distance

between the two sets of values. The RMSE is calculated by taking the standard deviation of the prediction errors, thereby providing a measure of the spread of the prediction errors.
The formula for Root Mean Square Error (RMSE) is:

$$RMSE = \sqrt{\frac{1}{n}\sum_i^n (P_i - A_i)^2} \quad (3)$$

where:
n is the number of observations.
P_i represents the actual values.
A_i represents the predicted values.
$(P_i - A_i)^2$ is the squared error for each observation.
The RMSE is useful for assessing the accuracy of a model; it provides a measure of how much error is present in the predictions and tends to give more weight to larger errors due to the squaring step.

Relative Absolute Error (RAE): The relative absolute error is calculated by taking the sum of the difference between the predicted and actual values and dividing it by the sum of the difference between the actual value and the mean of the actual values.
The formula for Relative Absolute Error is:

$$RAE = \frac{1}{n}\sum_i^n \frac{\sum_i^n (P_i - A_i)}{\sum_i^n (P_i - A_m)} \quad (4)$$

where:
n is the number of observations.
P_i is the actual value for observation i.
A_i is the predicted value for observation i.
A_m is total of actual values.
RAE is a valuable tool for assessing the efficacy of a model by providing a benchmark against which its performance can be evaluated. It is particularly useful for comparing the added value of more complex models relative to a straightforward reference model.

Root Relative Squared Error (RRSE): The Root Relative Squared Error (RRSE) formula is utilized to assess the efficacy of a predictive model, particularly in regression tasks. It entails a comparison between the error associated with a model's predictions and the error associated with a baseline model that predicts the mean of the actual values.
The formula for RRSE is:

$$\text{RRSE} = \sqrt{\frac{1}{n}\sum_i^n \frac{\sum_i^n (P_{ij} - A_i)^2}{\sum_i^n (P_i - A_m)^2}} \quad (5)$$

where:
n is the number of observations.
P_{ij} is the predicted value of dataset j for observation i.
A_i is the predicted value for observation i.
A_m is total of actual values.

The RRSE is a useful tool for evaluating the relative performance of a regression model in comparison to a basic model that solely predicts the mean of the actual values. By normalizing the squared error with respect to the error of the mean prediction, the RRSE provides a measure of improvement or deterioration in predictive performance.

3.2. Data Set

The data set of the study consists of daily price data for silver, US dollar index, VIX index, S&P 500 index, Brent oil and gold between 02/01/2008-28/06/2024. The variables included in the study's daily closing prices include 4090 days of data. The dataset of the study is obtained from Yahoo Finance [29]. The open-source software WEKA, utilized in the field of data mining, has been selected for the analysis of the working data. The variables that are thought to affect silver prices and the studies in the literature on these variables are shown in Table-1.

Table 1. Variables

Variables	Abbreviation	Unit	Variable Type	Frequency	References
Silver	XAG	Ons/USD	Output	Daily	
USD Dollar Index	DXY	Index	Input	Daily	[30]
Volatility Index	VIX	Index	Input	Daily	[31];[32]; [33];[34];
S&P 500 Index	SPX	Index	Input	Daily	[30]
Brent Oil	XBR	Barrel/USD	Input	Daily	[35];[36]; [37];[38]
Gold	XAU	Ons/USD	Input	Daily	[39];[40];[41]; [42];43

4. Results and Discussion

In this section of the study, the historical prices of silver during the research period are presented. This section also includes descriptive statistics for the variables, a correlation matrix, and the study's analysis results.

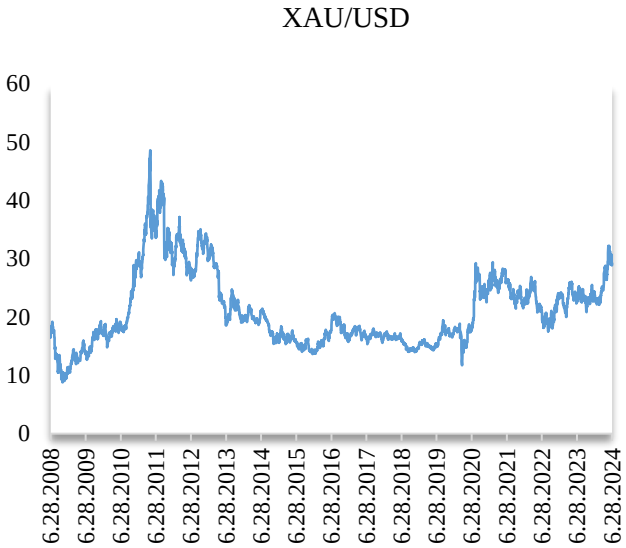


Figure 2. Silver prices in the research period
Source: created by the authors

Figure 2 illustrates that after the 2008 global economic crisis, silver prices entered a distinct upward trend. During this period, the price of silver reached approximately 50 USD per ounce, marking a historical peak. In the following years, particularly until 2020, silver prices exhibited relatively low volatility. However, in 2020, with the increased perception of global risk caused by the COVID-19 pandemic, there was a significant surge in silver demand, leading to a rise in silver prices. In 2021, prices remained elevated but showed some fluctuations. In 2022 and 2023, price variability continued due to economic recovery, monetary policies implemented by central banks, and other economic factors. In the first six months of 2024, silver prices fluctuated due to global economic developments, ranging from between 22 and 32 USD.

Table 2. Descriptive statistics

	DXY	VIX	SPX	XBR	XAU	XAG
Min.	71.33	9.14	676.53	19.33	713.30	8.79
Max.	114.11	82.69	5487.03	146.08	2433.90	48.58
Mean	90.30	19.99	2450.50	78.63	1439.60	20.86
Std.	9.84	9.12	1193.38	24.97	340.24	6.27
Obs.	4090	4090	4090	4090	4090	4090

Table 2 presents the descriptive statistics of the variables. According to the descriptive statistics table, the highest volatility is observed in the VIX, with a standard deviation (9.12) significantly higher than its mean (19.99). The SPX shows significant fluctuation, with a wide range and a high standard deviation (1193.38).

Correlation analysis is effectively used in machine learning-based modeling studies to identify input variables with high correlation in order to eliminate the problem of multicollinearity. A correlation coefficient greater than 0.8 indicates a significant relationship between the variables [44]. In this study, correlation analysis was conducted to examine whether there is a multicollinearity relationship among the variables. The analysis concluded that there are no barriers to including all selected input variables in the model. The results of the correlation analysis among the input variables are presented in Table 3. According to the findings, there is a strong positive relationship (0.7794) between DXY and SPX, indicating that the U.S. Dollar Index is in alignment with the S&P 500 Index. A moderately positive relationship (0.4443) is observed between DXY and XAU.

Table 3. Correlation table for the variables

	DXY	VIX	SPX	XBR	XAU	XAG
DXY	1					1
VIX	-0.1352	1				-0.1352
SPX	0.7794	-0.2140	1			0.7794
XBR	-0.5106	-0.1390	-0.1687	1		-0.5106
XAU	0.4443	-0.1383	0.7477	0.1537	1	0.4443

The strong positive correlation (0.7477) between SPX and XAU suggests that these two variables generally move together. On the other hand, there is a negative relationship (-0.5106) between DXY and XBR, indicating that the U.S. Dollar Index moves in the opposite direction of Brent crude oil prices.

Table 4. Analysis results

Linear Regression			
Error Terms	t+1	t+2	t+3
Correlation Coefficient	0.9210	0.9183	0.9167
MAE	1.7980	1.8168	1.8312
RMSE	2.5324	2.5734	2.5975
RAE%	34.6658	35.0281	35.3059
RRSE%	38.9844	39.6149	39.9858
SMOReg			
Error Terms	t+1	t+2	t+3
Correlation Coefficient	0.9181	0.9151	0.9136
MAE	1.7497	1.7671	1.7793
RMSE	2.6594	2.702	2.7307
RAE%	33.7345	34.0712	34.3061
RRSE%	40.9398	41.5948	42.0372
k-NN			
Error Terms	t+1	t+2	t+3
Correlation Coefficient	0.9954	0.9937	**0.9926**
MAE	0.4192	0.4696	**0.4713**
RMSE	0.6320	0.7316	**0.7878**
RAE%	8.0816	9.0535	**9.0873**
RRSE%	9.7298	11.2619	**12.1281**
Random Forest			
Error Terms	t+1	t+2	t+3
Correlation Coefficient	**0.9960**	**0.9941**	0.9937
MAE	**0.3969**	**0.4458**	0.4745
RMSE	**0.5877**	**0.7144**	0.7342
RAE%	**7.6518**	**8.5953**	9.1479
RRSE%	**9.0468**	**10.9971**	11.3019

In the current study, three distinct models have been developed to predict silver prices. As presented in Table 4, the first model is denoted as t+1, the second as t+2, and the third as t+3. The primary objective in constructing these models is to predict silver prices for one, two, and three days into the future. Each model uses the input variable values from day t, while the output variable-silver price-is represented by its value on days t+1, t+2, and t+3 for the first, second, and third models, respectively. The price predictions from these models are evaluated using four different machine learning algorithms. The use of multiple machine learning methods is intended to ensure the comparability and reliability of the results, facilitating the identification of the model that incorporates the most accurate method. The dataset is split into 80% training data and 20% test data. Predictions are made exclusively with the test data, and the models' performance is assessed based on these results. Model comparison is conducted through the error terms derived from the test data, method producing the lowest error being deemed the most successful.

Table 4 presents the prediction performance of four different machine learning methods individually:

Linear Regression

Correlation Coefficient: The correlation coefficient decreases slightly over time (0.9210→0.9167), indicating a gradual decline in the model's predictive accuracy as time progresses.
MAE and RMSE: Both MAE (1.7980→1.8312) and RMSE (2.5324→2.5975) show a slight increase over time. As errors increase, the model's predictions become less precise.
RAE and RRSE: Percentage errors also increase over time (RAE%34.66→%35.30, RRSE%38.98→%39.98), suggesting a decline in model performance over time.

SMOReg

Correlation Coefficient: Similar to linear regression, the correlation coefficient decreases over time (0.9181→0.9136).
MAE and RMSE: Over time, there have been slight increases in MAE (1.7497 → 1.7793) and RMSE (2.6594→2.7307). Although errors remain relatively low, they increase over time.
RAE and RRSE: Percentage errors increase over time (RAE %33.73→%34.31, RRSE %40.94→%42.04), indicating a decline in performance.

k-NN (k-Nearest Neighbors)

Correlation Coefficient: The correlation coefficient remains very high compared to other models (0.9954→0.9926). Although there is a very slight decrease over time, the model's accuracy remains quite stable.
MAE and RMSE: MAE (0.4192→0.4713) and RMSE (0.6320→0.7878) show a slight increase over time, but the errors are very low compared to other models.
RAE and RRSE: Percentage errors (RAE %8.08→%9.09, RRSE %9.73→%12.13) also increase over time but remain much lower than those of other models.

Random Forest

Correlation Coefficient: is very high (0.9960→0.9937), with a slight decrease over time, but the accuracy remains strong.
MAE and RMSE: MAE (0.3969→0.4745) and RMSE (0.5877→0.7342) increase over time, however these values remain relatively low compared to other models.
RAE and RRSE: Percentage errors (RAE % 7.65→% 9.15, RRSE % 9.05→% 11.30) are low, with a moderate increase.

k-NN and Random Forest exhibit the lowest error rates and highest correlation coefficients for t+1, t+2, and t+3 predictions. Although both models experience a minimal decrease in performance over time, they remain superior. In contrast, linear regression and SMOReg models show deteriorating performance and increasing errors over time, particularly in terms of MAE and RMSE. As time progresses (towards t+3), k-NN and Random Forest models provide better predictions, whereas linear regression and SMOReg models perform weaker.

5. Conclusion and Suggestions

In this study, silver prices are predicted using machine learning methods. In this context, four different machine learning algorithms commonly used in the literature are utilized. These algorithms are linear regression, support vector regression (SMOReg), k-nearest neighbors (k-NN), and random forest. Through these four algorithms, the predictive capabilities and performance of the models over time are demonstrated. This allows for a comparative assessment of the methods' performance in predicting silver prices. In general, Random Forest and k-NN (k-Nearest Neighbors) algorithms, respectively, perform better compared to other methods. These models exhibit higher correlation coefficients and significantly lower error rates than the other models. Random Forest produces the most accurate predictions in the first two models, while k-NN shows the highest prediction accuracy in the third model. Over time, the performance decline in these models remains minimal, and their predictive accuracy remains generally high. The high correlation coefficients and low error rates indicate that these two algorithms provide long-term reliability and accuracy in predicting silver prices. Based on these findings, it is recommended that k-NN and Random Forest algorithms be preferred for long-term predictions, especially in financial analysis, risk management, and investment strategies.

However, the linear regression and SMOReg models have shown a decline in performance over time. There has been a noticeable increase in correlation coefficients, MAE (mean absolute error), and RMSE (root mean square error) values. These increases indicate that these models' prediction accuracy decreases over time, resulting in a reduction in their overall performance. This suggests that the results produced by the linear regression and SMOReg algorithms may be less effective for long-term predicts. Therefore, it may be more appropriate to use these methods for shorter-term predictions or in periods of low volatility. Nevertheless, considering that SMOReg delivers acceptable results in certain cases in the short term, its performance could be improved through further model development and optimization.

In future studies, incorporating additional variables that influence silver prices into the model has the potential to enhance prediction accuracy. Specifically, integrating external factors such as geopolitical events, central banks' monetary policies, and macroeconomic indicators into the model could enrich the algorithms with more comprehensive datasets, leading to more consistent results. In addition, analyzing the interplay between these external factors and silver prices can provide deeper insights into market behavior and trends. Furthermore, exploring more advanced modeling techniques, such as deep learning, may yield superior outcomes compared to existing algorithms, thereby contributing to a more accurate predicting of financial markets.

Finally, we recommend updating the modeled data more frequently and periodically testing the accuracy of the results to improve the algorithms' applicability in real market conditions. This approach would enable the continuous optimization of the models, allowing them to better adapt to changing market dynamics. We anticipate that with more advanced parameter optimization techniques and

larger datasets, successful algorithms like k-NN and Random Forest could achieve even higher performance. Moreover, the creation of models with the capacity to adapt to real-time market fluctuations may serve to enhance their validity and effectiveness across a range of economic conditions. This would enable investors and analysts to make more informed decisions.

Contributions of the authors

Erhan Ergin: Data Collection, Literature Review, Data Analysis, Model Development, Editing.

Binali Selman Eren: Conceptualization, Model Development, Manuscript Writing, Interpretation of Results.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] L. A. Smales and B. M. Lucey, "The influence of investor sentiment on the monetary policy announcement liquidity response in precious metal markets," *Journal of International Financial Markets, Institutions and Money*, vol. 60, pp. 19-38, 2019.

[2] A. Charles, O. Darné, and J. H. Kim, "Will precious metals shine? A market efficiency perspective," *International Review of Financial Analysis*, vol. 41, pp. 284-291, 2015.

[3] U. Çelik and C. Başarır, "The prediction of precious metal prices via artificial neural network by using RapidMiner," *Alphanumeric Journal*, vol. 5, no. 1, pp. 45-54, 2017.

[4] S. Goel, M. Saxena, P. K. Sarangi, and L. Rani, "Gold and silver price prediction using hybrid machine learning models," in *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 2022, pp. 390-395.

[5] A. Üntez and M. İpek, "Developing financial forecast modeling with deep learning on silver/ons parity," Journal *of Advanced Research in Natural and Applied Sciences*, vol. 8, no. 1, pp. 35-44, 2022.

[6] S. Öndin and T. Küçükdeniz, "Latent Dirichlet allocation method-based nowcasting approach for prediction of silver price," Accounting, vol. 9, no. 3, pp. 131-1, 2023.

[7] S. Alparslan and T. Uçar, "Comparison of commodity prices by using machine learning models in the COVID-19 era, " *Turkish Journal of Engineering*, vol. 7, no. 4, pp. 358-368, 2023.

[8] H. Wang, B. Dai, X. Li, N. Yu, and J. Wang, "A novel hybrid model of CNN-SA-NGU for silver closing price prediction," *Processes*, vol. 11, no. 3, p. 862, 2023.

[9] D. N. Gono, H. Napitupulu, and Firdaniza, "Silver price forecasting using extreme gradient boosting (XGBoost) method," *Mathematics*, vol. 11, no. 18, p. 3813, 2023.

[10] Y. E. Gür, "Comparative analysis of deep learning models for silver price prediction: CNN, LSTM, GRU, and hybrid approach," *Akdeniz İİBF Dergisi*, vol. 24, no. 1, pp. 1-13, 2024.

[11] B. Jin and X. Xu, "Gaussian process regression-based silver price forecasts," *Journal of Uncertain Systems*, vol. 18, no. 1, 2024. [Online]. Available: https://doi.org/10.1142/S1752890924500132.

[12] M. M. Taye, "Understanding of machine learning with deep learning: architectures, workflow, applications and future directions," *Computers*, vol. 12, no. 5, p. 91, 2023.

[13] R. Gupta, A. Sharma, and T. Alam, "Building Predictive Models with Machine Learning," in *Data Analytics and Machine Learning: Navigating the Big Data Landscape*, Singapore: Springer Nature Singapore, 2024, pp. 39-59.

[14] E. Koçoğlu and F. Ersöz, "Data mining application with machine learning algorithms to manage interest rate risk," *Business & Management Studies: An International Journal*, vol. 10, no. 4, pp. 1545-1564, 2022.

[15] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*. Hoboken, NJ, USA: John Wiley & Sons, 2021.

[16] S. Savaş and S. Buyrukoğlu, Eds., *Teori ve Uygulamada Makine Öğrenmesi*, 1st ed. İstanbul, Türkiye: Beta Basım Yayım Dağıtım A.Ş., 2022.

[17] J. Cervantesa, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges, and trends," *Neurocomputing*, vol. 189, pp. 189-215, 2020.

[18] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.

[19] A. J. Smola and B. Scholkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, pp. 199–222, 2004.

[20] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to the SMO algorithm for SVM regression," *IEEE Transactions on Neural Networks*, vol. 11, no. 5, pp. 1188-1193, 2000.

[21] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, and M. Data, "Practical machine learning tools and techniques," in *Data Mining*, vol. 2, no. 4, pp. 403-413, 2005. Amsterdam: Elsevier.

[22] C. Li and L. Jiang, "Using locally weighted learning to improve SMOreg for regression," in *Pacific Rim International Conference on Artificial Intelligence*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 375-384.

[23] L. Wang, L. Tan, C. Yu, and Z. Wu, "Study and application of non-linear time series prediction in ground source heat pump system," in *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, 2012, pp. 3522-3525.

[24] P. Cunningham and S. Delany, "k-Nearest neighbour classifiers - A tutorial," *ACM Computing Surveys*, vol. 54, no. 6, pp. 1-25, 2021.

[25] Y. C. Liang, Y. Maimury, A. H. L. Chen, and J. R. C. Juarez, "Machine learning-based prediction of air quality," *Applied Sciences*, vol. 10, no. 24, p. 9151, 2020.

[26] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[27] A. Ibrahim, R. Kashef, and L. Corrigan, "Predicting market movement direction for Bitcoin: A comparison of time series modeling methods," *Computers & Electrical Engineering*. 2021. [Online]. Available: https://doi.org/10.1016/j.compeleceng.2020.106905.

[28] B. Marapelli, "Software development effort duration and cost estimation using linear regression and k-nearest neighbors machine learning algorithms," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 2, pp. 2278-3075, 2019.

[29] finance.yahoo.com (Accessed Date: 1 July 2024)

[30] D. G. Baur and L. A. Smales, "Hedging geopolitical risk with precious metals," *Journal of Banking & Finance*, vol. 117, p. 105823, 2020.

[31] A. A. Salisu, A. E. Ogbonna, and A. Adewuyi, "Google trends and the predictability of precious metals," *Resources Policy*, vol. 65, p. 101542, 2020.

[32] T. L. D. Huynh, "The effect of uncertainty on the precious metals market: New insights from transfer entropy and neural network VAR," *Resources Policy*, vol. 66, p. 101623, 2020.

[33] M. B. Tuncel, Y. Alptürk, T. Yılmaz, and İ. Bekci, "Korku endeksi (VIX) ile kıymetli madenler arasındaki ilişki üzerine ekonometrik bir çalışma," *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, vol. 35, no. 3, pp. 1069-1083, 2021.

[34] M. Dumlupınar and T. Kocabıyık, "Yatırım aracı olarak kıymetli metaller: Kıymetli metallerin fiyatını etkileyen unsurlar ve kıymetli metallerde nedensellik ilişkisi," *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, vol. 28, no. 2, pp. 241-266, 2023.

[35] M. Bildirici and C. Türkmen, "The chaotic relationship between oil return, gold, silver, and copper returns in Turkey: Non-linear ARDL and augmented non-linear Granger causality," *Procedia-Social and Behavioral Sciences*, vol. 210, pp. 397-407, 2015.

[36] I. Arif, L. Khan, and K. M. Iraqi, "Relationship between oil price and white precious metals return: A new evidence from quantile-on-quantile regression," *Pakistan Journal of Commerce and Social Sciences*, vol. 13, no. 2, pp. 515-528, 2019.

[37] M. Kamışlı, S. Kamışlı, and F. Temizel, "Emtia fiyatları birbirlerini etkiler mi? Asimetrik frekans nedensellik analizi," *Uluslararası Yönetim İktisat ve İşletme Dergisi*, vol. 13, no. 13, pp. 1079-1093, 2017.

[38] A. A. Salisu, A. E. Ogbonna, and A. Adewuyi, "Google trends and the predictability of precious metals," *Resources Policy*, vol. 65, p. 101542, 2020.

[39] B. Elmas and M. Polat, "Gümüş fiyatları ve Dow Jones endeksi'nin altın fiyatlarına etkisi üzerine eşbütünleşme ve nedensellik analizi," *Bingöl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, vol. 3, no. 6, pp. 33-48, 2013.

[40] A. F. Erol and S. Aytekin, "Altın fiyatlarını etkileyen faktörlerin nedensellik analizi ile incelenmesi," in *International Congress of Management, Economy and Policy (ICOMEP'20)*, 2020.

[41] B. R. Mishra, A. K. Pradhan, A. K. Tiwari, and M. Shahbaz, "The dynamic causality between gold and silver prices in India: Evidence using time-varying and non-linear approaches," *Resources Policy*, vol. 62, pp. 66-76, 2019.

[42] A. Açacak, E. Gülsar, and E. Meriç, "Kıymetli madenlerin birbirleriyle ilişkisi: Asimetrik nedensellik," *Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, vol. 21, no. 1, pp. 28-37, 2020.

[43] M. Aslan and L. Sizer, "Dow Jones endeksi, gümüş fiyatları ve altın fiyatları arasındaki ilişkinin ekonometrik açıdan incelenmesi," *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, vol. 23, no. 4, pp. 1035-1048, 2023.

[44] M. Uluskan and H. D. Şenli, "YSA sınıflandırma modellerinde korelasyon-hipotez testi tabanlı filtreleme yoluyla girdi seçimi," *Nicel Bilimler Dergisi*, vol. 6, no. 1, pp, 68-102, 2024.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1304-1313 YEAR: 2024
DOI: 10.17798/bitlisfen.1558069

SimCLR-based Self-Supervised Learning Approach for Limited Brain MRI and Unlabeled Images

Kazım FIRILDAK[1*], Gaffari ÇELİK[2], Muhammed Fatih TALU[3]
[1]*Firat University, Department of Computer Technology, Elazığ, Türkiye*
[2]*Ağrı İbrahim Çeçen University, Department of Computer Technology, Ağrı, Türkiye*
[3]*İnönü University, Department of Computer Science, Malatya, Türkiye*
(ORCID: 0000-0002-1958-3627) (ORCID: 0000-0001-5658-9529) (ORCID: 0000-0003-1166-8404)

Keywords: Unlabeled images, Brain tumor, Self-Supervised learning, SimCLR, Barlow twins, NnCLR

Abstract
In this study, a SimCLR-based model is proposed for the classification of unlabeled brain tumor images in medical imaging using a self-supervised learning (SSL) technique. Additionally, the performances of different SSL techniques (Barlow Twins, NnCLR, and SimCLR) are analyzed to evaluate the performance of the proposed model. Three different datasets, consisting of pituitary, meningioma, and glioma brain tumors as well as non-tumor images, were used as the dataset. Out of a total of 7,671 images, 6,128 were used as unlabeled data, and the model was trained with both labeled and unlabeled data. The proposed model achieved high performance with unlabeled data, reducing the need for manual labeling. As a result, the model demonstrated superior performance compared to other models, with high performance values such as 99.35% c_acc and 96.31% p_acc.

1. Introduction

Deep learning has become a widely used method in image classification [1],[2],[3], segmentation [4],[5],[6], audio processing [7],[8],[9], object detection [10],[11], and natural language processing (NLP) [12],[13] tasks. Significant improvements have been achieved, especially in challenging areas such as disease classification and organ segmentation in medical image analysis. These successes are often based on large amounts of manually labeled training data. While labels can be easily obtained through crowd sourcing in natural images, this method is limited in medical imaging due to the requirement of expert knowledge. This situation indicates that access to unlabeled medical images is often easier than access to labeled images [14].

Supervised learning approaches have reached their limits due to the challenges and costs associated with the manual annotation of labeled data. Additionally, image recognition systems learn image representations using large amounts of images along with semantic annotations for these images. These annotations can be provided in different formats such as class labels, hashtags, or bounding boxes. However, predefined semantic annotations often fail to encompass the long tail of diverse and rare visual concepts. This limitation hinders progress in the performance of image recognition systems [15],[16],[17].

While traditional methods rely on annotations in large datasets, self-supervised methods enable more efficient use of deep learning by providing the ability to learn without the need for these annotations. These methods further advance developments in the field of deep learning by facilitating feature learning from unlabeled data [15],[18].

Self-supervised learning (SSL) is defined as a learning paradigm that sits between unsupervised learning and supervised learning. Compared to supervised learning, it aims to learn without the requirement for labeled data. Instead, it allows the model to generate labels for itself by leveraging the natural structure and relationships within the data. In this way, the model extracts “supervisory signals” from the data itself. In supervised learning approaches, large amounts of labeled data are generally required for the model to learn. However, this labeling process can be time-consuming and costly. SSL aims to learn by utilizing the intrinsic

*Corresponding author: kfirildak@firat.edu.tr
Received: 29.09.2024, Accepted: 28.10.2024

structures within the data, allowing for efficient model training on large amounts of unlabeled data [14],[19].

SSL can be considered a subfield of unsupervised learning since it does not involve manual labeling. However, when viewed within a narrower framework, unsupervised learning generally aims to uncover specific data patterns such as clustering, community discovery, or anomaly detection, whereas SSL targets extracting information from data based on the fundamental principles of supervised learning [14]. Figure 1 clearly illustrates the differences between these two approaches [20].

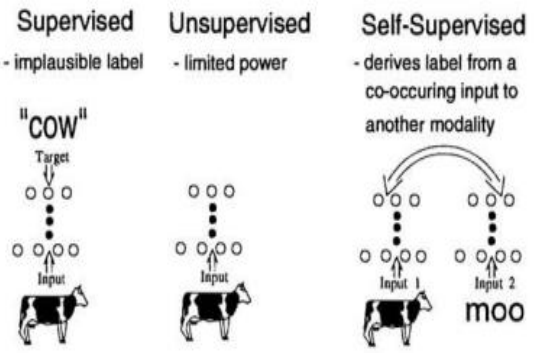


Figure 1. An Example of Supervised, Unsupervised, and Self-Supervised Learning.

Our main contributions in this study are given as follows.

- In the study, high success was achieved in the classification of unlabeled brain tumor images using the SimCLR-based SSL model. This approach effectively used unlabeled data and reduced the need for manual labeling.
- The proposed model achieved better results compared to other SSL techniques with high performance values such as 99.35% c_acc and 96.31% p_acc. This shows that the model is effective in medical image analysis.
- By using 6128 of the total 7671 images as unlabeled, a significant saving was achieved in the data labeling
- process and the need for manual labeling was minimized.
- By comparing different SSL techniques such as Barlow Twins, NnCLR and SimCLR, it was shown that the SimCLR-based model was superior. These comparisons are an important contribution in terms of evaluating the effectiveness of SSL techniques in medical image classification.
- The study showed that the SSL approach has a wide application potential, especially in areas such as medical image analysis where limited labeled data is available. This method can provide more efficient and faster medical image classification by reducing the need for expert knowledge.

2. Material and Method

2.1. Material

Although labels can be easily obtained through crowd sourcing from datasets, this method is limited in medical imaging because it requires expert knowledge. Therefore, this study was conducted to classify unlabeled data using the SSL technique with brain tumor images from medical images.

The brain is an organ that controls vital systems and tissues in the human body, such as hormonal regulation, memory, and reasoning. Brain tumors occur due to the abnormal proliferation of brain cells. Tumors, which are categorized as benign or malignant, increase intracranial pressure, complicating the control of vital systems and potentially leading to the patient's death in advanced stages. As with all types of cancer, early diagnosis and treatment increase survival rates for brain tumors. Changes in the brain can be detected using medical imaging techniques such as computed tomography and magnetic resonance imaging (MRI). The brain MRI images used in this study consist of four classes. These classes are specified as "glioma, meningioma, pituitary, and no_tumor."

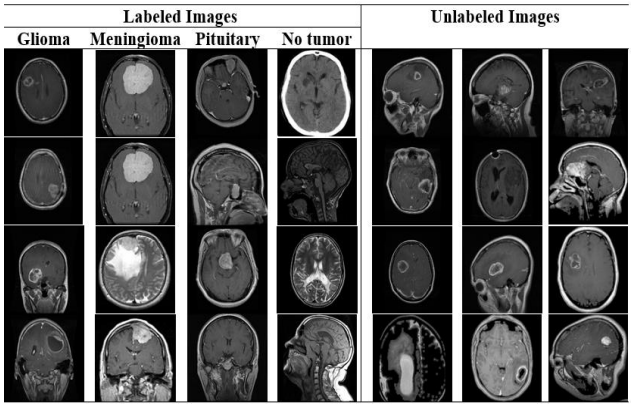


Figure 2. Some Labeled and Unlabeled Images Containing Brain Tumor Images.

In this study, three different datasets containing brain MRI images were used. The first dataset includes a total of 1311 MRI images, consisting of 300 Pituitary images, 306 Meningioma images, 300 Glioma images, and 405 No Tumor

images [21]. The second dataset contains a total of 3264 MRI images, comprising 901 Pituitary images, 937 Meningioma images, 926 Glioma images, and 500 No Tumor images [22]. The third dataset used contains a total of 3096 MRI images, including 844 Pituitary images, 913 Meningioma images, 901 Glioma images, and 438 No Tumor images [23]. To test the developed networks, the brain MRI datasets from three different sources were mixed and split into 1235 training data, 6128 unlabeled data, and 308 test data. Sample brain MRI images are shown in Figure 2.

2.2. Method

In this study, a SimCLR-based model is proposed for training on limited datasets with a small number of images. The performances of SimCLR [24], Barlow Twins [25] and NnCLR [26] techniques were also analyzed by examining these methods.

SimCLR, Barlow Twins, and NnCLR methods use twin networks. These networks aim to minimize the network error for representations of data belonging to the same class. Figure 3 (a) provides an example of a twin network architecture. The twin networks share the same network weights and attempt to learn to produce a minimum distance value for different samples belonging to the same class. This situation is the opposite for samples belonging to different classes. Figure 3 (b) illustrates this situation in detail for sample data from the MNIST dataset. The network error is calculated using methods like contrastive loss, and the network is trained.

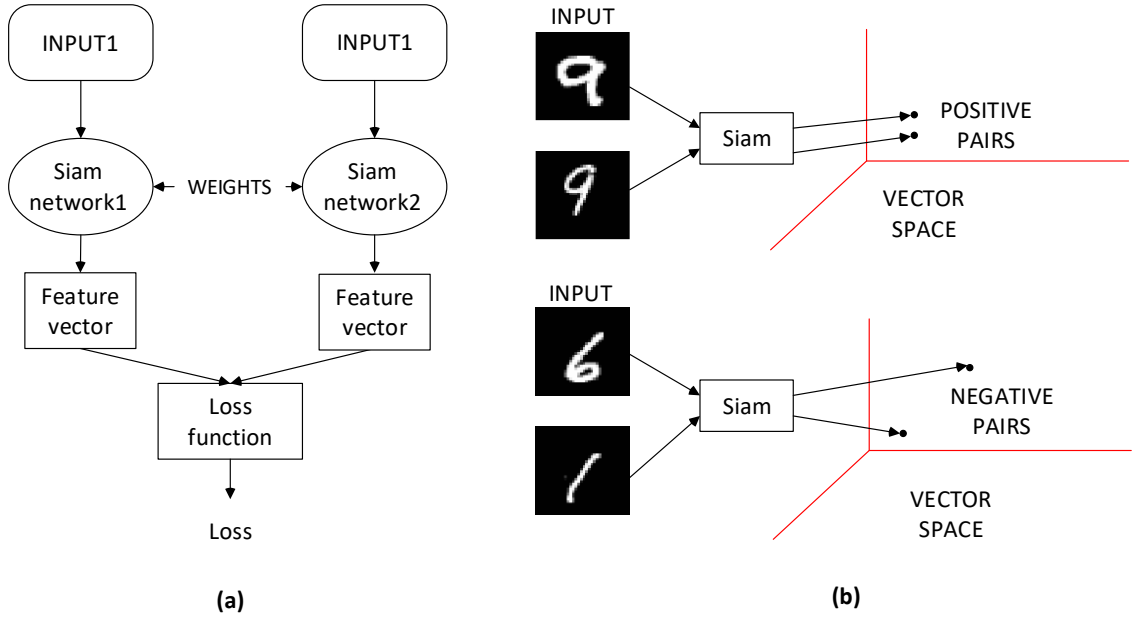


Figure 3. (a) Twin Network Architecture, (b) Positive and Negative Examples of Twin Networks in Vector Space

2.2.1. SimCLR

SimCLR fundamentally utilizes twin networks. These networks learn unlabeled data from datasets lacking sufficient amounts of data through classic data augmentation in four stages [24]. Figure 4 presents a schematic representation of the SimCLR method.

In the first stage of the SimCLR method, unlabeled data is augmented to create an increased data pool for each sample using classic data augmentation techniques (random resizing, random cropping, and adding Gaussian blur noise). In the second stage, the encoder network $f(.)$ obtains feature vectors for these augmented images x_i and x_j. In the third stage, a projection operation $g(.)$ is applied. At the end of this process, the contrastive loss value for the vectors z_i and z_j is calculated. The contrastive loss is given by the following equation.

$$l_{i,j} = -log \frac{exp\left(sim\left(z_i, z_j\right)/\tau\right)}{\sum_{k-1,indicator[k\neq i]}^{2N} exp\left(sim\left(z_i, z_j\right)/\tau\right)} \quad (1)$$

Here, N is the number of samples, and τ is the temperature parameter. The indicator function is defined as $\{Eğer\ [k \neq i]\ ise\ 1\ \}$. When calculating the loss, the contrastive loss is computed based on the cosine similarity between the selected number of positive data samples. The proposed model is trained

according to the backpropagation algorithm by taking the partial derivatives of the contrastive loss method.

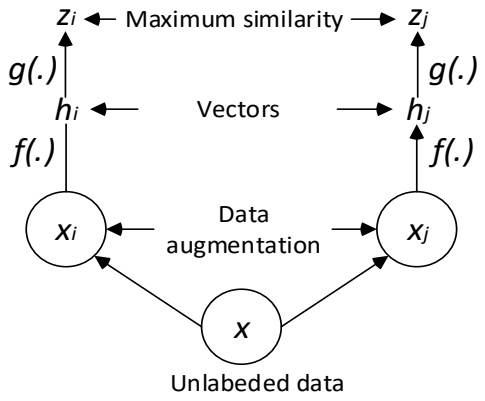


Figure 4. Schematic Representation of the SimCLR Method

2.2.2. Barlow twins

Barlow twins [25] is an innovative Contrastive Learning (CL) model that aims to perform classification on datasets with limited data using a self-supervised learning strategy with twin networks. This method generates augmented images T for all images in the sampled batch X. These images are provided as input in pairs y^a and y^b and y^a and y^b to the deep learning model f with parameters θ. The output of this model consists of feature vectors z^a and z^b. These vectors are assumed to be mean-centered over the batch size, resulting in an average output of 0 for each unit across the batch. The loss function of the Barlow Twins method is shown below.

$$L_{BT} \triangleq \sum_{\hat{1}} (1 - C_{ii}) + \lambda + \sum_{i} \sum_{j \neq i} C_{ij}^2 \qquad (2)$$

Here, λ is a positive constant that adjusts the importance of the first and second terms of the loss. C is the cross-correlation matrix calculated between the outputs of the twin networks over the batch size. This matrix is constructed according to the following equation.

$$C_{ij} \triangleq \frac{\sum_b z_{b,i}^A z_{b,j}^B}{\sqrt{\sum_b \left(z_{b,i}^A\right)^2} \sqrt{\sum_b \left(z_{b,j}^B\right)^2}} \qquad (3)$$

Here, b represents the batch samples. i, j are the vector dimensions of the network outputs. C has the output dimension of the network and takes values ranging from -1 (perfect anti-correlation) to 1 (perfect correlation) [25].

The objective function of the Barlow Twins method resembles the INFONCE method [27]. However, it stands out because it does not require a large number of negative samples and works effectively with high-dimensional vectors.
After calculating the Barlow Twins loss, the method trains the twin network according to the backpropagation algorithm.

2.2.3. NnCLR

Developed based on the SimCLR method, this approach is a modern self-supervised learning architecture that uses twin networks. The NnCLR method [26] employs the nearest neighbor clustering technique to increase the number of latent representations and obtain more diverse positive pairs rather than single positive examples [26]. This necessitates maintaining a vector support set that represents the entire data distribution. Figure 5 presents a schematic representation of the NnCLR algorithm.

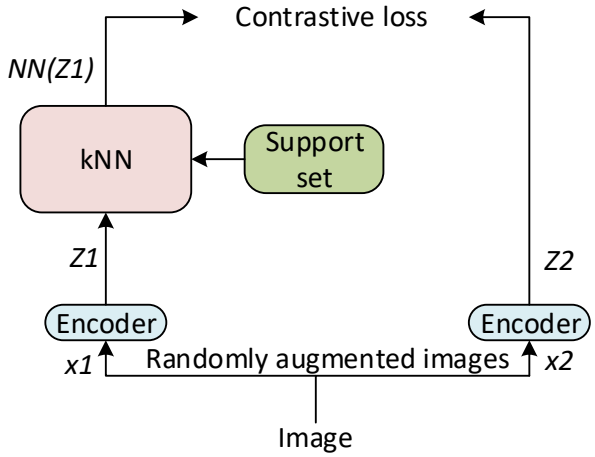


Figure 5. Schematic Representation of the NnCLR Method

SimCLR uses two augmented images z_i and $z_,^+$to create a positive pair. In contrast, NnCLR proposes using the nearest neighbors of z_i from the Q support set to form positive pairs. Similar to SimCLR, negative pairs are applied to the loss function in mini-batches. The loss function of the NnCLR algorithm is provided below.

$$L_i^{NnCLR} = -log \frac{exp(NN(z_i, Q) \cdot z_i^+ / \tau)}{\sum_k^n exp\left(NN(z_i, Q) \cdot z_i^+ / \tau\right)} \qquad (4)$$

2.2.4. Proposed Model

The overall structure of the proposed model is presented in Figure 6. The proposed model consists of the Encoder, Projection Head, and Classification stages. The model takes labeled and unlabeled images as input, and the Encoder module extracts feature maps. The Projection Head transforms the features extracted in the Encoder stage into a different space, creating similarity matrices between unlabeled and labeled data. In the Classification stage, the similarity matrices obtained in the Projection Head are used to predict which class the data may belong to.

In the Encoder stage, the input images undergo two convolution (Conv2D) operations followed by MaxPooling2D operations sequentially, repeated three times. The obtained features are then flattened into a vector using the Flatten operation and processed through three Dense layers. In the Conv2D operations, the number of filters (f) is set to 16, 32, and 64, with a filter size (kernel size) of 3x3. The activation function used after the Conv2D operations is ReLU. The Conv2D operation extracts feature maps from the images, while the MaxPooling2D operation reduces dimensionality by eliminating unnecessary features.

The Projection Head stage consists of three consecutive Dense layers. In these layers, the number of neurons is set to 256, 128, and 64, respectively. Additionally, the activation function used in each Dense layer is ReLU. The SimCLR loss function is employed for training the proposed Encoder and Projection architecture.

The Classification stage also consists of three consecutive Dense layers, with the number of neurons set to 256, 128, and 4, respectively. The activation function used in the first two Dense layers is ReLU. Categorical CrossEntropy is utilized for training the Classification module. The Categorical CrossEntropy (CE) cost function is as follows [28].

$$L_{CE} = -\sum_{i=1}^{N} y_i \,.\, log\, \hat{y}_i \tag{5}$$

Here, N represents the number of classes. y_i denotes the true classes, while $\hat{y}_i$ represents the predicted class.

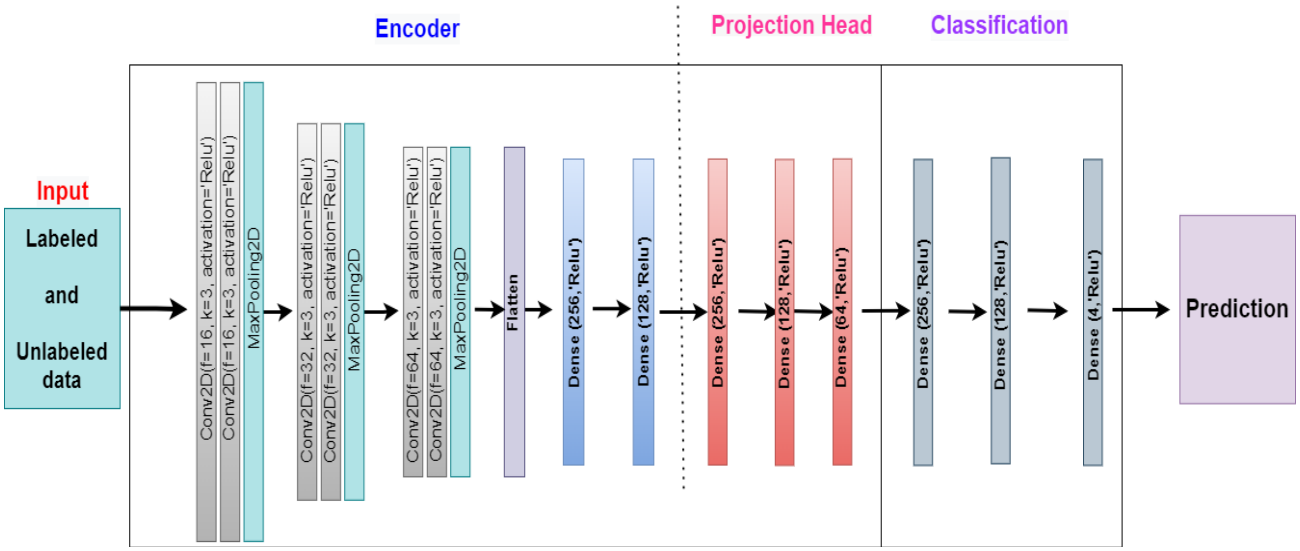


Figure 6. Overall Structure of the Proposed Model.

3. Results and Discussion

In this study, a SimCLR-based model was proposed to enhance the classification performance of unlabeled images. Three datasets containing brain tumor images were used, specifically comprising "pituitary," "meningioma," and "glioma" types, along with non-tumor images. A total of 7671 images were included in the datasets, out of which 6128 were designated as unlabeled data, while 1235 images were allocated for training and 308 for testing. The proposed model was trained using both labeled and unlabeled data.

For the training of the proposed model, the Adam optimization algorithm was employed, with SimCLR Loss and Categorical Crossentropy as the cost functions. The images were trained at a resolution of 96x96 over 100 epochs, with a temperature parameter τ set to 0.1. Data augmentation has been applied using the RandomFlip ("horizontal") method. This technique increases data diversity, allowing the model to generalize better. The

performance of the models is based on the Accuracy metric. This metric is given below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Where, TP denotes the true-positive value, TN stands for the true-negative value, FP refers to the false-positive value, and FN indicates the false-negative value.

In the initial experimental study, the performance of the proposed model was compared to that of the SimCLR model, and the results are presented in Table 1. Additionally, the accuracy and loss graphs of the models are displayed in Figure 7. p_acc represents the accuracy of the classifier output, c_acc denotes the accuracy of comparative learning, r_acc indicates correlation accuracy, and val_p_acc reflects the test accuracy.

Table 1. Test Results of the Proposed Model vs. SimCLR.

Model	c_acc (%)	p_acc (%)	r_acc (%)	val_p_acc (%)
SimCLR	99.28	58.86	00.78	54.87
Proposed Model	**99.35**	**96.31**	**57.76**	**72.40**

The findings indicate that the proposed model achieved a comparative accuracy (c_acc) of 99.35%, a classification accuracy (p_acc) of 96.31%, a correlation accuracy (r_acc) of 57.76%, and a validation accuracy (val_p_acc) of 72.40%. These results demonstrate a significant improvement over the SimCLR model, underscoring the effectiveness of the proposed self-supervised learning approach in classifying unlabeled images. The substantial gains in accuracy metrics suggest that the model can efficiently leverage unlabeled data, enhancing performance in medical image classification tasks.

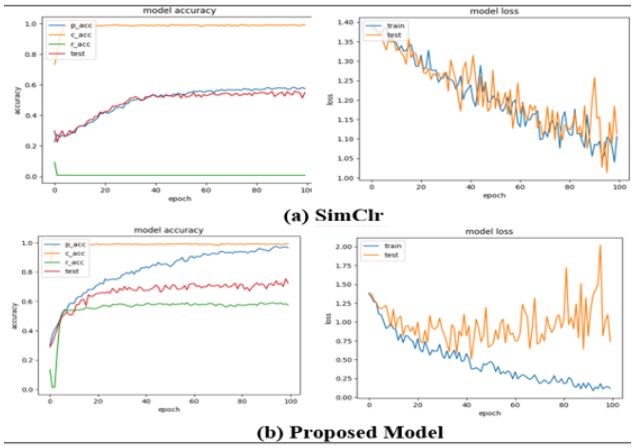


Figure 7. Accuracy and Loss Graphs of SimCLR and Proposed Model. (a) SimCLR Accuracy and Loss Graphs, (b) Proposed Model Accuracy and Loss Graphs.

This figure (Figure 7) illustrates the performance comparison between the SimCLR model and the proposed model. The accuracy and loss metrics for both models are displayed, providing insight into their training dynamics and overall effectiveness in classifying the datasets. The proposed model exhibits superior performance, as indicated by higher accuracy and lower loss across the training epochs.

The findings presented in this table summarize the performance metrics of the models employed in this study. Upon reviewing the results, it is evident that the proposed model predicts with significantly higher accuracy compared to the other models. This suggests that the incorporation of the self-supervised learning technique, particularly through the SimCLR framework, has contributed to enhanced classification performance in the context of brain tumor imaging.

Table 2. Test results of the models

Model	c_acc (%)	p_acc (%)	r_acc (%)	val_p_acc (%)
Barlow twins	98.65	56.63	35.94	53.25
NnCLR	98.36	62.36	36.26	58.12
SimCLR	99.28	58.86	00.78	54.87
Proposed Model	**99.35**	**96.31**	**57.76**	**72.40**

Accuracy and loss graphs of the models during training are presented in Figure 8. Additionally, the accuracy and loss graphs of the Barlow Twins and NnCLR models are shown in Figure 9. When examining the accuracy graph (a) in Figure 8, we can say that the proposed model demonstrates more stable and higher accuracy throughout the training.

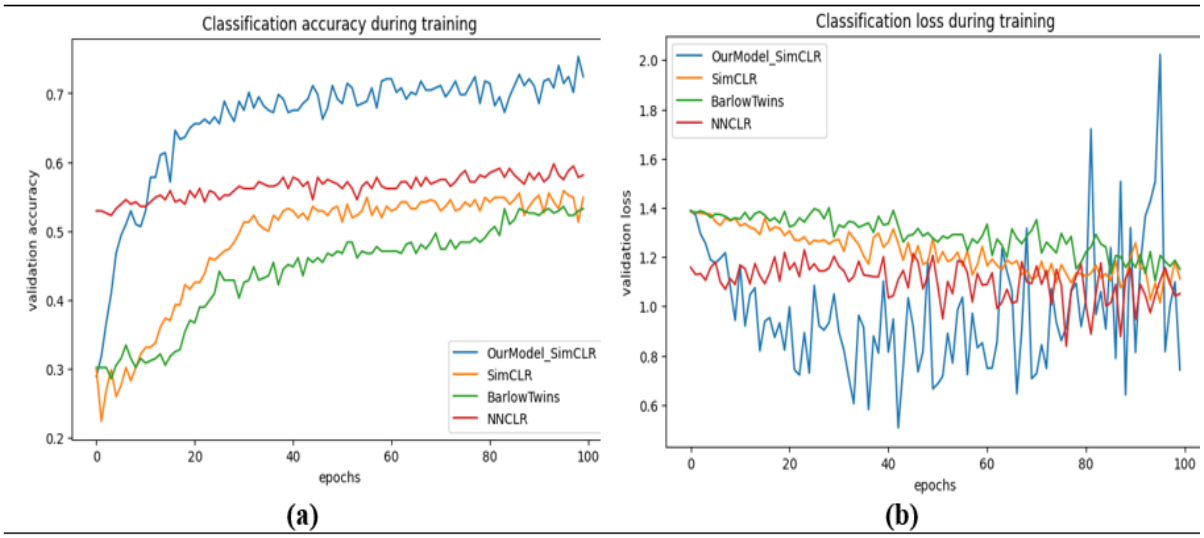


Figure 8. Accuracy (a) and loss (b) graphs of the models.

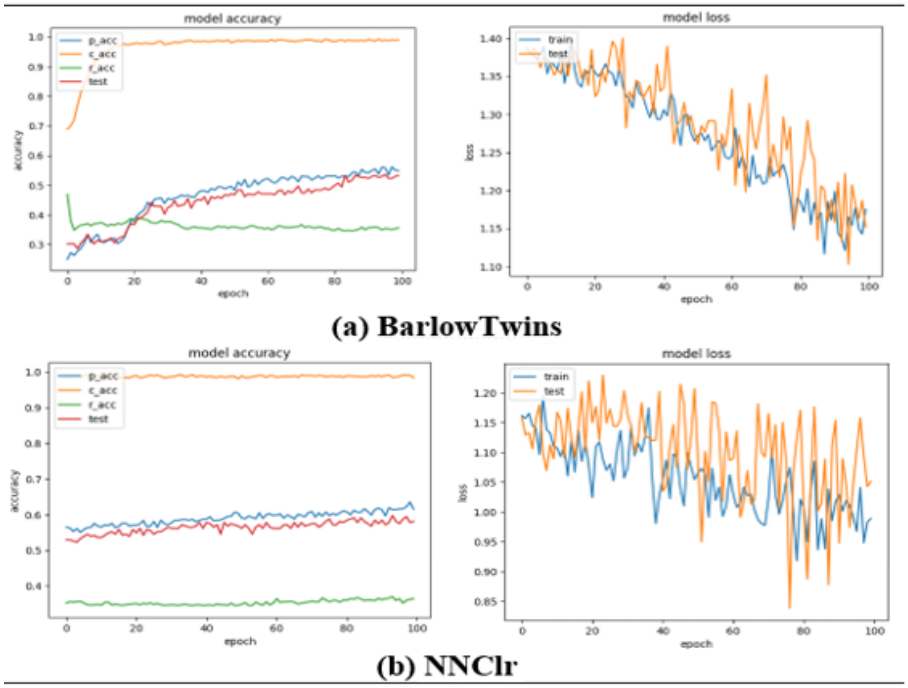


Figure 9. Accuracy and loss graphs of the BarlowTwins and NnCLR models. (a) BarlowTwins, (b) NnCLR model graphs

4. Conclusion and Suggestions

In this study, a SimCLR-based model was proposed using self-supervised learning (SSL) techniques to enhance the classification performance of unlabeled brain tumor images. The research was conducted on three types of brain tumors: pituitary, meningioma, and glioma, as well as non-tumor images. A dataset consisting of a total of 7,671 images was used, of which 6,128 were designated as unlabeled data, while the remaining images were utilized for training and testing. The proposed model was trained with both unlabeled and labeled data, effectively reducing the need for expert knowledge in medical imaging and minimizing the requirement for manual labeling.

As a result of performance analysis, it was observed that the proposed model achieved significant improvements compared to the SimCLR model, with values of 99.35% c_acc, 96.31% p_acc, 57.76% r_acc, and 72.40% val_p_acc. These findings demonstrate the effective utilization of unlabeled data through the SSL technique. It was concluded that this approach holds great potential for broader applications, especially in fields such as medical image analysis, where the labeled dataset is limited.

Contributions of the authors

Fırıldak: conceptualization, data collection, formal analysis, visualization, software development, interpretation of results
Çelik: methodology, literature review, interpretation of results, writing-first draft, revision
Talu: project administration, conceptualization, methodology, formal analysis, supervision, revision

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] M. Toğaçar, N. Muzoğlu, B. Ergen, B. S. B. Yarman, and A. M. Halefoğlu, "Detection of COVID-19 findings by the local interpretable model-agnostic explanations method of types-based activations extracted from CNNs," *Biomed. Signal Process. Control*, vol. 71, p. 103128, Jan. 2022, doi: 10.1016/j.bspc.2021.103128.

[2] G. Celik, "Detection of Covid-19 and other pneumonia cases from CT and X-ray chest images using deep learning based on feature reuse residual block and depthwise dilated convolutions neural network," *Appl. Soft Comput.*, vol. 133, p. 109906, Jan. 2023, doi: 10.1016/j.asoc.2022.109906.

[3] E. Başaran, "A new brain tumor diagnostic model: Selection of textural feature extraction algorithms and convolution neural network features with optimization algorithms," *Comput. Biol. Med.*, vol. 148, p. 105857, Sep. 2022, doi: 10.1016/j.compbiomed.2022.105857.

[4] G. Çelik and M. F. Talu, "A new 3D MRI segmentation method based on Generative Adversarial Network and Atrous Convolution," *Biomed. Signal Process. Control*, vol. 71, p. 103155, Jan. 2022, doi: 10.1016/j.bspc.2021.103155.

[5] S. Altun Güven and M. F. Talu, "Brain MRI high resolution image creation and segmentation with the new GAN method," *Biomed. Signal Process. Control*, vol. 80, p. 104246, Feb. 2023, doi: 10.1016/j.bspc.2022.104246.

[6] Z. Bozdag and M. F. Talu, "Pyramidal position attention model for histopathological image segmentation," *Biomed. Signal Process. Control*, vol. 80, p. 104374, Feb. 2023, doi: 10.1016/j.bspc.2022.104374.

[7] G. Celik and E. Başaran, "Proposing a new approach based on convolutional neural networks and random forest for the diagnosis of Parkinson's disease from speech signals," *Appl. Acoust.*, vol. 211, p. 109476, Aug. 2023, doi: 10.1016/j.apacoust.2023.109476.

[8] S. Mavaddati, "Voice-based age, gender, and language recognition based on ResNet deep model and transfer learning in spectro-temporal domain," *Neurocomputing*, vol. 580, p. 127429, May 2024, doi: 10.1016/j.neucom.2024.127429.

[9] M. A. Islam, M. Z. H. Majumder, M. A. Hussein, K. M. Hossain, and M. S. Miah, "A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets," *Heliyon*, vol. 10, no. 3, p. e25469, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25469.

[10] R. B. Rahman, S. A. Tanim, N. Alfaz, T. E. Shrestha, M. S. U. Miah, and M. F. Mridha, "A comprehensive dental dataset of six classes for deep learning based object detection study," *Data Br.*, vol. 57, p. 110970, Dec. 2024, doi: 10.1016/j.dib.2024.110970.

[11] B. Ganga, L. B. T., and V. K. R., "Object detection and crowd analysis using deep learning techniques: Comprehensive review and future directions," *Neurocomputing*, vol. 597, p. 127932, Sep. 2024, doi: 10.1016/j.neucom.2024.127932.

[12] K. Kantor and M. Morzy, "Machine learning and natural language processing in clinical trial eligibility criteria parsing: a scoping review," *Drug Discov. Today*, vol. 29, no. 10, p. 104139, Oct. 2024, doi: 10.1016/j.drudis.2024.104139.

[13] A. Montejo-Ráez, M. D. Molina-González, S. M. Jiménez-Zafra, M. Á. García-Cumbreras, and L. J. García-López, "A survey on detecting mental disorders with natural language processing: Literature review, trends and challenges," *Comput. Sci. Rev.*, vol. 53, p. 100654, Aug. 2024, doi: 10.1016/j.cosrev.2024.100654.

[14] L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, and D. Rueckert, "Self-supervised learning for medical image analysis using image context restoration," *Med. Image Anal.*, vol. 58, p. 101539, Dec. 2019, doi: 10.1016/j.media.2019.101539.

[15] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, "A Survey on Contrastive Self-Supervised Learning," *Technologies*, vol. 9, no. 1, p. 2, Dec. 2020, doi: 10.3390/technologies9010002.

[16] X. Liu *et al.*, "Self-supervised Learning: Generative or Contrastive," *IEEE Trans. Knowl. Data Eng.*, pp. 1–1, 2021, doi: 10.1109/TKDE.2021.3090866.

[17] I. Misra and L. van der Maaten, "Self-Supervised Learning of Pretext-Invariant Representations," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 6706–6716, doi: 10.1109/CVPR42600.2020.00674.

[18] Y. Wang, C. M. Albrecht, N. A. A. Braham, L. Mou, and X. X. Zhu, "Self-Supervised Learning in Remote Sensing: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 4, pp. 213–247, Dec. 2022, doi: 10.1109/MGRS.2022.3198244.

[19] S. Shurrab and R. Duwairi, "Self-supervised learning methods and applications in medical imaging analysis: a survey," *PeerJ Comput. Sci.*, vol. 8, p. e1045, Jul. 2022, doi: 10.7717/peerj-cs.1045.

[20] V. R. de Sa, "Learning Classification with Unlabeled Data," in *Adv. Neural Inf. Process. Syst.*, pp. 112–119, 1994. [Online]. Available: https://dl.acm.org/doi/10.5555/2987189.2987204

[21] S. Gupta, "Brain MRI Scans for brain tumor classification," Accessed: Jan. 25, 2024. [Online]. Available: https://www.kaggle.com/datasets/shreyag1103/brain-mri-scans-for-brain-tumor-classification

[22] S. Bhuvaji, A. Kadam, P. Bhumkar, S. Dedge, and S. Kanchan, "Brain Tumor Classification (MRI)," Accessed: Jan. 12, 2024. [Online]. Available: https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri/data

[23] Thomas, "Brain tumors," Accessed: Jan. 18, 2024. [Online]. Available: https://www.kaggle.com/datasets/thomasdubail/brain-tumors-256x256?select=Data

[24] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," Feb. 2020. [Online]. Available: http://arxiv.org/abs/2002.05709

[25] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny, "Barlow Twins: Self-Supervised Learning via Redundancy Reduction," Mar. 2021. [Online]. Available: http://arxiv.org/abs/2103.03230

[26] D. Dwibedi, Y. Aytar, J. Tompson, P. Sermanet, and A. Zisserman, "With a Little Help from My Friends: Nearest-Neighbor Contrastive Learning of Visual Representations," Apr. 2021. [Online]. Available: http://arxiv.org/abs/2104.14548

[27] A. van den Oord, Y. Li, and O. Vinyals, "Representation Learning with Contrastive Predictive Coding," Jul. 2018.

[28] J. C. Triana-Martinez, J. Gil-González, J. A. Fernandez-Gallego, A. M. Álvarez-Meza, and C. G. Castellanos-Dominguez, "Chained Deep Learning Using Generalized Cross-Entropy for Multiple Annotators Classification," *Sensors*, vol. 23, no. 7, p. 3518, Mar. 2023, doi: 10.3390/s23073518.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1314-1324 YEAR: 2024
DOI: 10.17798/bitlisfen.1558208

Exploring Radiation Shielding Properties of Lanthanide Elements

Nuray YAVUZKANAT[1*]

[1]*Bitlis Eren University, Faculty of Art and Sciences, Physics Department, Bitlis, Türkiye (ORCID: 0000-0001-5055-9185)*

Keywords: Lanthanides, LAC, HVL, EABF, EBF, GATE.

Abstract

In this work, the radiation shielding properties of Lanthanide elements were studied using the EpiXS program and GATE simulation, which agreed well with each other, based on some key parameters such as MAC, LAC, HVL, MFP, EABF, and EBF. It was observed that at lower energies of gamma-rays, the values of MAC and LAC are maximum, which decrease with the increase in energy due to reduced photoelectric interactions. Photoelectric absorption edges couple with peaks in attenuation values; peaks for elements of the lower atomic number, La, Ce, Pr, and Nd, appear as two while the peaks for elements of higher atomic number are three due to the additional absorptions by L-shell sub- levels or M-shell. These peaks take place when the energy of photons meets the energy level of electron binding. While Lutetium has the highest and Europium has the lowest LAC values, Lutetium also has the lowest HVL and MFP values; thus, it has the best radiation shielding properties. The EABF and EBF reach their maximum in the medium energy range and then decrease. Lutetium has the lowest photon buildup, and Lanthanum has the highest EABF and EBF values for all the studied elements at all penetration depths.

1. Introduction

The Rare earth elements (REEs) consist of a total of 17 elements, which include scandium, yttrium, and the lanthanides, which are crucial for digital and low-emission technologies [1]. Although they are called rare earth elements (REEs), they are actually quite abundant in the crust of the Earth, with light REEs being as abundant as copper [1]. Most REE mines are found in carbonatite-related deposits, with China dominating their production and controlling much of their extraction, which is why they are considered critical raw materials [1], [2]. REEs are utilized in advanced consumer electronics, eco-friendly products, industrial and medical equipment, as well as in the defense systems [2]. Their high reactivity posed difficulties in refining them into pure metals, and effective separation techniques were not developed until the 20th century because of their chemical similarities [3].

REEs are extensively utilized in radiation protection because they can be easily molded and enhance the physical and chemical properties of materials [4], [5]. With the growing emphasis on nuclear energy over traditional fossil fuels, the role of shielding materials has become more prominent. This is especially true in industries such as non-destructive testing and radiotherapy, where effective shielding in high-energy environments is crucial [6], [7]. In their study, Jing and colleagues investigated the shielding properties of REEs composite materials. The results showed that REEs effectively absorb thermal neutrons and gamma rays, while also improving the mechanical properties of the protective composites [11].

Lanthanides, occupying atomic numbers 57 to 71 on the periodic table, are categorized into light (57-64) and heavy (65-71) REEs based on their electron configurations [8]. This group of elements, including the lanthanides from Lanthanum (La) to Lutetium (Lu), plays a crucial role in digital and low-carbon technologies because of their distinctive magnetic and luminescent characteristics, making them essential in industries such as magnets,

[1]Corresponding author: nyavuzkanat@beu.edu.tr

Received: 30.09.2024, Accepted: 02.12.2024

hydrogen alloys, catalysts, and electronics, with growing demand driven by advanced technologies [1], [8].

An understanding of the different phenomena arising due to interaction of radiation with matter provides insight into the manner by which the radiation is transmitted or penetrated through a medium. This knowledge aids in selecting appropriate shielding materials based on the type of radiation [9,10]. Radiation shielding materials traditionally include lead, iron, and tungsten. While these materials can provide as an effective shield against gamma rays, they are less successful in shielding against neutrons. Besides, they have a number of limitations, such as opacity to visible light, and there are problems related to lead's toxicity, weight, and lack of flexibility [11]-[14]. In their work, Jing et al. studied the shielding features of REEs composite materials. The results indicated that REEs were effective in absorbing thermal neutrons and gamma rays; they also contribute to the mechanical properties of the protective composites [11]. In another study, Cui et al. evaluated the shielding properties of some rare earth/polypropylene materials at a tube voltage of 120 kV. The study found that increasing the rare earth filler content was more effective in enhancing the composite's shielding performance than merely increasing its thickness [5].

Incorporating lanthanides, such as Eu_2O_3, increases the photon interaction and linear attenuation coefficient (LAC), enhancing radiation shielding capacity [24]. The lead borate glasses with high concentrations of Pb are excellent glass materials that provide good radiation shielding characteristics; however, they present some problems in relation to transparency because of certain radiation-induced changes in glass [25]. Incorporating rare earth elements, such as cerium and dysprosium oxides, can improve both the transparency and radiation shielding effectiveness of these glasses [25].

In this investigation, the characteristics associated with radiation shielding of lanthanides, were investigated using the EpiXS program [26] and Monte Carlo simulations performed with the GATE framework. The radiation shielding capability of the material depends on a parameter known as the mass attenuation coefficient (MAC) (μ/ρ). Further, the shielding properties such as linear attenuation coefficient (LAC), half-value layer (HVL), mean free path (MFP), energy absorption buildup factor (EABF), and exposure buildup factor (EBF) have been studied using this parameter. The results obtained from both EpiXS and GATE simulation offer an in-depth insight into the shielding capabilities of the lanthanide elements.

2. Material and Method

Table 1 presents the key properties of the studied Lanthanide elements, ranging from Lanthanum to Lutetium, including density (g/cm^3) [27], effective atomic number, and electron density. These parameters are crucial for evaluating the shielding effectiveness of each element, as they directly influence photon interaction and attenuation capabilities.

Table 1. Key properties of Lanthanide elements relevant to radiation shielding effectiveness.

Elements	Density* (g/cm^3)	Effective Atomic Number	Electron Density
Lanthanum (La)	6.15	57	2.4712
Cerium (Ce)	6.77	58	2.4928
Praseodymium (Pr)	6.77	59	2.5216
Neodymium (Nd)	7.01	60	2.5050
Promethium (Pm)	7.26	61	2.5335
Samarium (Sm)	7.52	62	2.4831
Europium (Eu)	5.24	63	2.4966
Gadolinium (Gd)	7.90	64	2.4509
Terbium (Tb)	8.23	65	2.4630
Dysprosium (Dy)	8.55	66	2.4459
Holmium (Ho)	8.80	67	2.4464
Erbium (Er)	9.07	68	2.4483
Thulium (Tm)	9.32	69	2.4597
Ytterbium (Yb)	6.90	70	2.4359
Lutetium (Lu)	9.84	71	2.4437

*Density values are sourced from PubChem's Periodic Table database [27].

The intensity I of a photon beam passing through a material with a thickness of *x* (in cm) is determined using the Beer-Lambert law, which is expressed by the equation:

$$I = I_0\, e^{-\mu x} \tag{1}$$

where μ (cm^{-1}) represents the LAC, encompassing all interactions such as coherent scattering, Compton scattering, and the photoelectric effect. The MAC, denoted μ_m (cm^2/g), quantifies the amount of radiation absorbed by a material and is given by:

$$\mu_m = \frac{\mu}{\rho} = \sum w_i \left(\frac{\mu_i}{\rho_i}\right) \tag{2}$$

where ρ (g/cm^3) is the material's density, and w_i indicates the weight fraction of the i^{th} element. The MFP is defined as the average distance a photon travels before undergoing an interaction and is related to the HVL, which is the thickness of material required to reduce the radiation intensity by half. The HVL and MFP can be calculated as follows:

$$HVL = \frac{ln\,2}{\mu}\ ,\ MFP = \frac{1}{\mu} \tag{3}$$

In this study, MFP refers to the mean free path, representing the average distance a photon travels in a material before interaction. On the other hand, mfp denotes a multiple of the mean free path, used to express relative penetration depths, such as '40 mfp' indicating 40 times the mean free path. The buildup factor quantifies the ratio of the total photon flux at a given point to the photon flux that reaches that point without interacting. This factor is divided into two categories: the EABF and the EBF. EABF quantifies the energy absorbed by the material during radiation interaction, whereas EBF refers to the energy absorbed or retained in the air [28], [29]. The buildup factor is incorporated into the photon intensity equation:

$$I\,(x) = B\, I_o\, e^{-\mu\, x} \tag{4}$$

where B represents the buildup parameter. This parameter can be equal to or greater than 1. In cases where the interacting material has a low thickness and the photon is mono-energetic, the B value is 1. In other cases, the buildup value becomes greater than 1 [30], [31].

The Energy Absorption Buildup Factor (EABF) and Exposure Buildup Factor (EBF) are calculated using Harima's geometric-progression (G-P) fitting formula, which is widely applied to various materials [32]. This formula uses specific G-P fitting coefficients, denoted as a, b, c, d, and X_k, which are determined for each material and photon energy level according to the following expression.

$$P = \frac{P_1(log\ Z_2 - log\ Z_{eq}) + P_2(log\ Z_{eq} - log\ Z_1)}{log\ Z_2 - log\ Z_1} \tag{5}$$

The P values represent the G-P fitting coefficients that correspond to the atomic numbers Z at the specified energy level. Z_{eq} is the effective atomic number, which is used to characterize the material's overall response to photon interactions based on its atomic composition at that energy. The buildup factor is defined in terms of the mfp using the G-P fitting parameters, as expressed in the following equations:

$$B(E,x) = 1 + (b-1)X \begin{cases} \frac{K^x - 1}{K-1} & K \neq 1 \\ x & K = 1 \end{cases} \tag{6}$$

$$K(E,x) = c\,x^a + d\,\frac{tanh\left(\frac{x}{X_k} - 2 - -tanh(-2)\right)}{1 - tanh(-2)} \tag{7}$$

$$for\ x \leq 40 mfp$$

E represents the photon energy and *x* is the mfp value. The coefficient b is associated with the buildup factor at 1 mfp. The parameter K(E,x) indicates the photon dose multiplier and accounts for the alteration in the spectrum's shape at 1 mfp [32].

In this study, a comparison was made between the values obtained using two different methods: the EpiXS software program and the Monte Carlo simulations performed with the GATE framework. The EpiXS software program is an open-access tool designed for users in the field. It incorporates EPICS2017, which is the photon and electron library derived from ENDF/B-VIII. In contrast, EPDL97 is the photon library based on ENDF/B-VI.8. EPICS2017 features updated binding energies and cross sections and has been linearized to facilitate Lin-Lin interpolation, enhancing the data resolution within the photon library. This linearization has significantly increased the data density in the photon library, whereas the data density in the electron library remained unchanged [26].

The Monte Carlo simulation method is used across a wide range of fields, from medical

physics to particle physics. Among the most well-known Monte Carlo codes is Geant4, along with other specialized codes such as PENELOPE, FLUKA, MCNP, and GATE, which are tailored to specific application areas [33]. Geant4 is a software written in C++ that simulates interactions of particles with matter. GATE, on the other hand, is an interface program that runs Geant4 simulations and is continuously updated through international collaborations, with its development being closely tied to updates in Geant4 itself [33].

The GATE modeling used in this study is based on the detailed descriptions provided in the works of T. Şahmaran and Yavuzkanat, as shown in Figure 1 [28]. The NaI(Tl) scintillation detector (2-inch diameter x 2-inch height), modeled as a cylindrical structure with a thin MgO reflector to enhance light collection, was equipped with an aluminum shield for radiation protection. To evaluate the gamma shielding effectiveness, the investigated shielding material was placed between the detector and the lead block. Measurements labeled as I_0 represent open-field conditions, taken without shielding, while the values recorded as I were obtained after radiation passed through the shielding material. The exponential relationship between I and I_0 allows for the calculation of the attenuation coefficient of the material, as given by Equation 1.

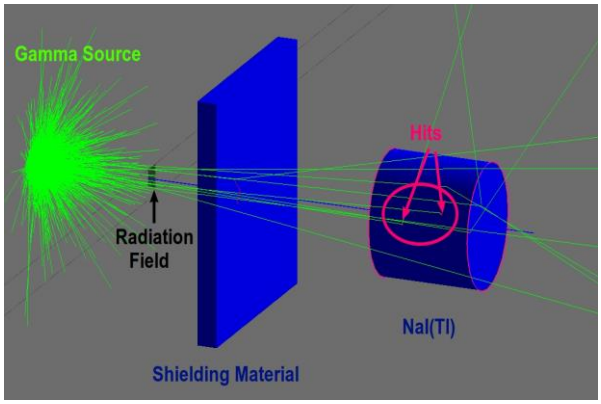


Figure 1. Configuration of the simulation setup, where gamma radiation is emitted isotropically from behind the blocks. The radiation field was arranged to be 1x1 cm² [28].

The percentage difference (PD) between the values was calculated using the formula:

$$PD = \left| \frac{LAC_{EpiXS} - LAC_{GATE}}{LAC_{EpiXS}} \right| x\ 100 \quad (8)$$

3. Results and Discussion

In Figures 2, as photon energy increases, the MAC and LAC for Lanthanide elements decrease. This trend indicates a decline in the likelihood of photoelectric interactions, as the energy becomes too high for photons to efficiently eject inner-shell electrons. Instead, Compton scattering and pair production processes become more significant, which are less energy-dependent. At very high photon energies, the MAC and LAC values stabilize, showing minimal variation. This stabilization occurs because the dominant attenuation mechanisms, such as Compton scattering and pair production, lead to a more constant attenuation behavior regardless of further increases in photon energy.

The MAC, LAC, HVL, and MFP for gamma energies ranging from 0.001 to 1000 MeV were determined using the EpiXS program for the Lanthanide group of elements. The results showed that these values were highest at low gamma ray energies and gradually decreased as the energy increased.

At lower photon energies (typically below 0.1 MeV or 100 keV), the sharp peaks are most likely due to photoelectric absorption edges, where the photon energy matches electrons' binding energy in the atom's inner shells. In Lanthanide group elements, the number of observed peaks (2 or 3) is related to their atomic structure. Elements with lower atomic numbers (La, Ce, Pr, and Nd) typically exhibit 2 peaks, corresponding to photon absorption in the K and L shells as observed in Figure 1. In contrast, elements with higher atomic numbers show 3 peaks due to additional absorption from the L-shell sub-levels (L1, L2, L3) or M-shell. The peaks occur when photon energy matches the binding energy of these electron shells, and as atomic number increases, more complex interactions lead to additional absorption edges and peaks.

As photon energy increases past these absorption edges, the effect of photoelectric absorption diminishes, leading to a gradual decline in both the MAC and LAC. This decrease is expected because the photoelectric absorption cross-section is inversely related to photon energy. At higher photon energies, beyond approximately 1 MeV, attenuation levels off as Compton scattering becomes the dominant interaction mechanism, with pair production also contributing at even higher energies. It was determined in Figures 2 (b) that Lutetium and Europium have the highest and lowest LAC values, respectively. In

Figure 2 (a), Lutetium and Lanthanum were found to have the highest and lowest MAC values, respectively.

According to Nagaraj et al. materials like Lanthanum polymer, Cerium polymer, Praseodymium polymer, Gadolinium texaphyrin polymer, Terbium co-ordination polymer and Erbium phosphate hydrate polymer were used to investigate X-ray/gamma, neutron, and EMI (electromagnetic interference) shielding properties [34]. Among them, Erbium phosphate hydrate showed the highest values in key shielding parameters such as mass attenuation coefficient (MAC) and effective atomic number (Z_{eff}). These findings suggest that Erbium phosphate hydrate polymer may be an ideal material for shielding against X-ray/gamma, neutron, and EMI radiation [34]. In our study, similar to this work, the MAC values are ranked from largest to smallest as follows: La < Ce < Pr < Gd < Tb < Er. Additionally, in our study examining all lanthanide group elements, the best results were observed in the order of Lutetium, Ytterbium, Thulium, and Erbium.

Figures 3 illustrate the variation in HVL and MFP values across the energy range of 0.001 to 1000 MeV. According to these figures, Lutetium exhibits the lowest HVL and MFP values as photon energy increases, indicating that it provides the most effective shielding capabilities among the Lanthanide elements. In contrast, Europium has the highest HVL and MFP values, reflecting its relatively lower shielding effectiveness. Lower HVL and MFP values generally signify better shielding capability, as these materials are more efficient at attenuating photon penetration. The observed differences in HVL and MFP values between Lutetium and Lanthanum can be attributed to their atomic numbers. Lutetium, with a higher atomic number (Z=71), has a denser electron cloud, which increases the likelihood of photon interactions, such as photoelectric absorption and Compton scattering. Consequently, this results in lower HVL and MFP values, reflecting a higher capacity for photon attenuation compared to Lanthanum (Z=57).

The differences between the Energy Absorption Buildup Factor (EABF) and the Exposure Buildup Factor (EBF) are primarily due to how each factor accounts for photon interactions in the material. EABF values are typically higher than EBF values because they consider not only the primary photon interactions but also cumulative effects from secondary photons produced through scattering processes. This cumulative interaction is especially noticeable in elements with higher atomic numbers (Z), where the increased electron density leads to enhanced photoelectric absorption and Compton scattering.

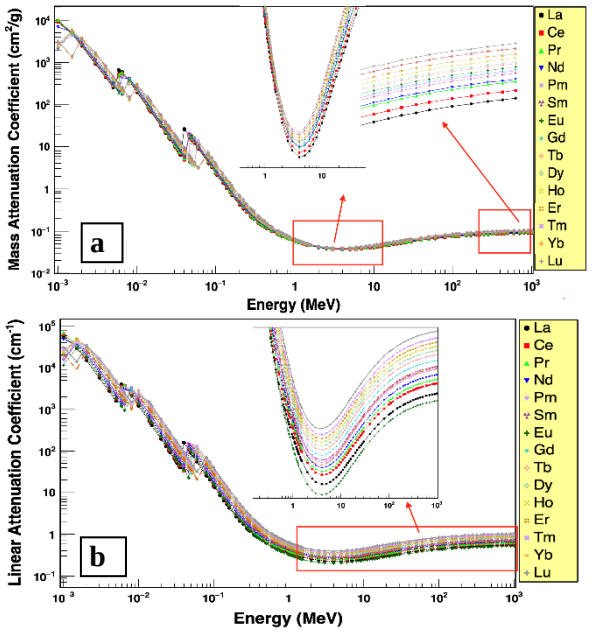


Figure 2. (a) MAC and **(b)** LAC values for Lanthanide group elements in the energy range of 0.001 to 1000 MeV.

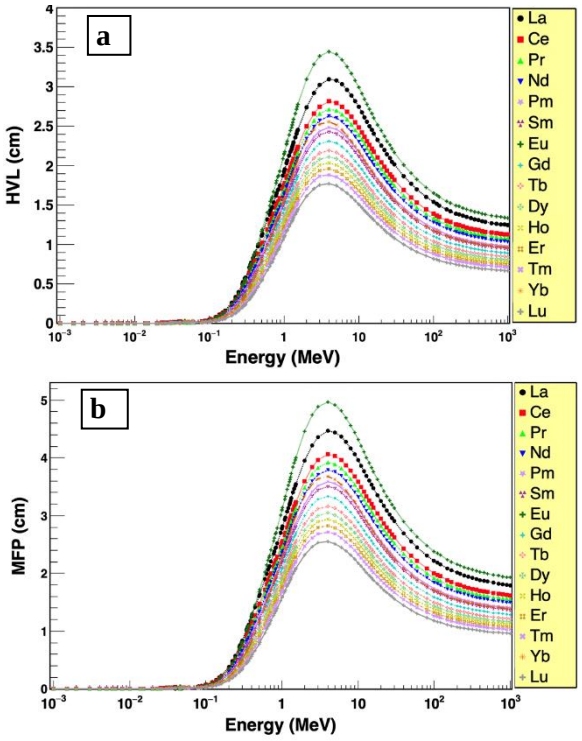


Figure 3. (a) HVL and **(b)** MFP values for Lanthanide group elements across the energy range of 0.001 to 1000 MeV.

Consequently, materials with higher Z values tend to exhibit higher EABF values, as their atomic structure promotes more significant photon attenuation and secondary photon production. In contrast, EBF values do not fully account for these secondary interactions, resulting in lower buildup values in comparison. This distinction highlights the role of atomic structure in influencing buildup factors, particularly as photon energy increases.

In Figure 4 and 5, the variation of the EABF and EBF for Lanthanide elements as a function of photon energy at various penetration depths (1, 10, 20, and 40 mfp). The graphs reveal that both EABF and EBF values initially increase, reaching a peak within the intermediate energy range, and then decrease as photon energy continues to rise. In the regions of low and high energy, the interactions responsible for the complete absorption of photons are photoelectric effects and pair production. As a result, photons have a shorter lifetime in the material. In contrast, Compton scattering predominates in the medium energy range, causing photons to stay in the material longer and resulting in higher EABF and EBF values. As shown in Figure 4 and 5, Lutetium exhibits the lowest photon buildup across all penetration depths, while Lanthanum shows the highest EABF and EBF values. This is because photons experience more scattering in Lanthanum, leading to greater photon buildup. As the photon path length through the material increases, or as the material thickness grows, scattering also increases, further raising the EABF and EBF values.

During the pair production process, the cross-section is proportional to Z^2, causing most photons to disappear, as in the photoelectric effect. Therefore, the EABF and EBF values are smaller in these energy regions. Photon energies above 1.02 MeV (the threshold for pair production) allow the pair production process to occur, with its interaction cross-section increasing in proportion to the square of the atomic number (Z^2). However, Compton scattering remains a significant competing process in this energy range, affecting photon interactions alongside pair production. Although pair production contributes to photon attenuation, its influence on the energy absorption buildup factor (EABF) and exposure buildup factor (EBF) is limited due to the concurrent impact of Compton scattering.

As a result, there is a less pronounced reduction in these factors, even in the energy range where pair production becomes more prominent. This is particularly evident in the graph (Figure 3), where no clear decrease in EABF values is observed above 1.2 MeV, despite pair production starting to take effect. This can be attributed to the ongoing significant contribution of Compton scattering at these higher energies, which offsets the potential reduction in EABF that might be expected solely from pair production.

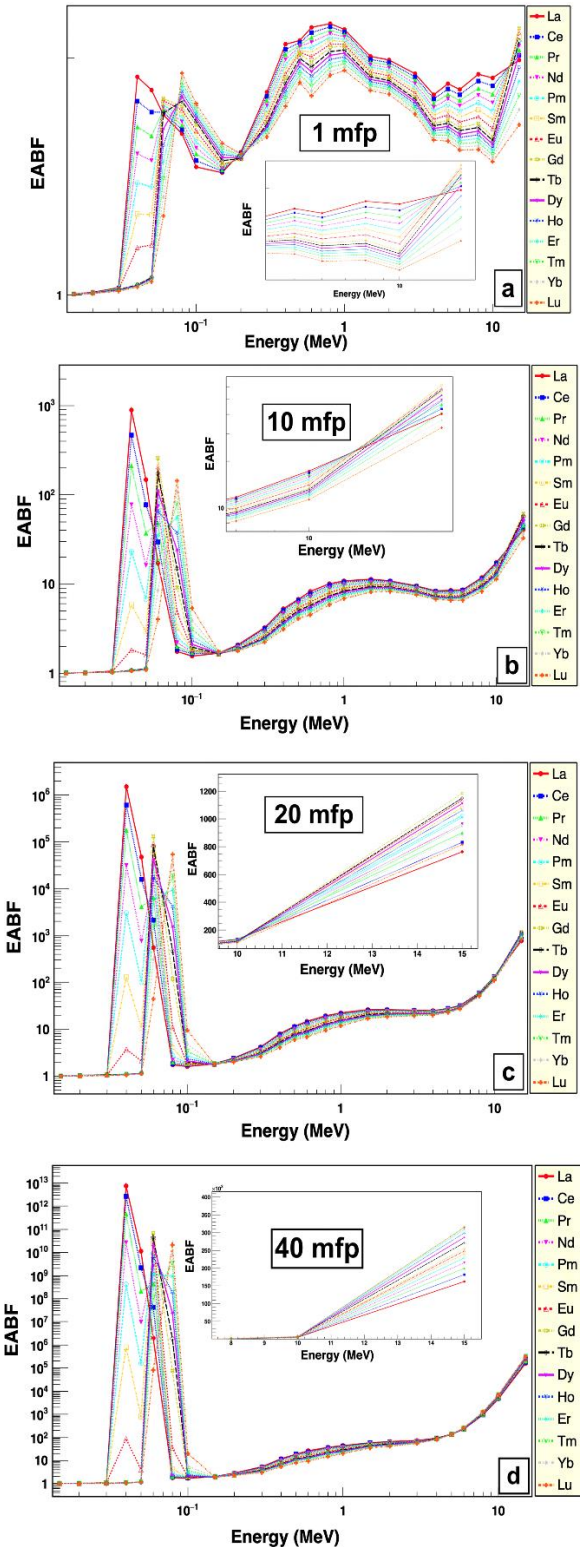


Figure 4. Variation of EABF with photon energy in the range of 0.015 to 15 MeV for Lanthanide group elements at various penetration depths for **(a)**1 mfp, **(b)** 10 mfp, **(c)** 20 mfp, **(d)** 40 mfp.

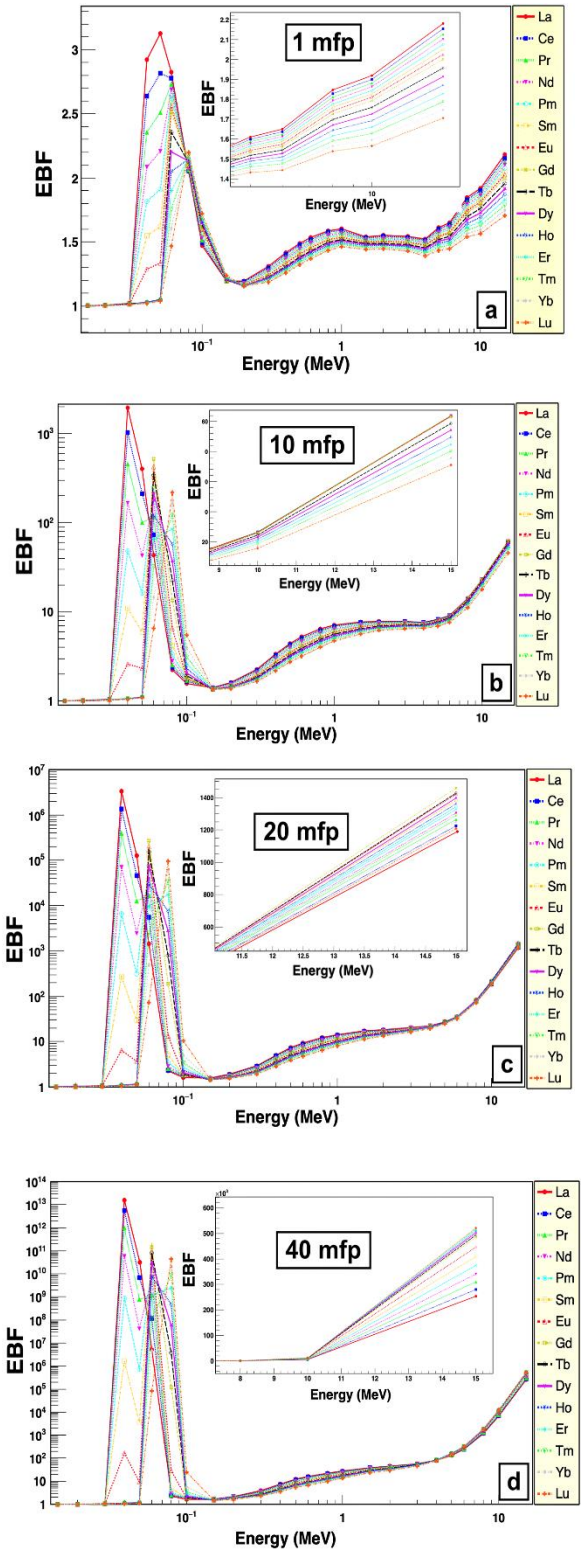


Figure 5. Variation of EBF with photon energy in the range of 0.015 to 15 MeV for Lanthanide group elements at various penetration depths for **(a)**1 mfp, **(b)** 10 mfp, **(c)** 20 mfp, **(d)** 40 mfp.

The Linear Attenuation Coefficient (LAC) values for each lanthanide element within the energy range of 200–2000 keV were compared with the results obtained from EpiXS and GATE simulations. These are presented in Table 2. The percentage difference was calculated using Eq. (8) and was found to be less than 5%.

According to Pyngrope's work, Erbium has the highest mass attenuation coefficient (0.03907 cm/g), while Lanthanum has the lowest (0.03646 cm/g), both measured with 4 MeV gamma radiation [35]. Erbium also has the lowest HVL (1.956 cm) and MFP (2.823 cm), making it the most effective lanthanide for radiation shielding at this energy level [35]. In our study, the results are consistent with the literature; based on the GATE simulation, the mass attenuation coefficient (MAC) of Erbium was found to be 0.039107 cm/g, while Lanthanum's MAC value was 0.036405 cm/g. Furthermore, according to both the GATE simulation and EpiX calculation, which show good agreement, the lowest HVL and MFP values are observed for Lutetium, with an HVL of 1.7707 cm and an MFP of 2.5545 cm. The highest values are found for Europium, with an HVL of 3.3924 cm and an MFP of 4.8943 cm.

4. Conclusion and Suggestions

The findings of this study underscore the varying radiation shielding capabilities of Lanthanide elements, with key parameters such as the MAC, LAC, HVL, and MFP. The observed decrease in MAC and LAC values with increasing photon energy underscores the transition from photoelectric interactions at lower energies to the predominance of Compton scattering and pair production at higher energies. The distinct peaks in attenuation values, correlated with the atomic structure of the elements, illustrate how atomic number influences shielding effectiveness, with Lutetium demonstrating superior performance due to its lowest HVL and MFP values. Conversely, Europium exhibited less effective shielding capabilities. Furthermore, the analysis of the EABF and EBF reveals critical insights into photon interactions, particularly the impact of energy range on these factors. Overall, this research provides valuable data that can inform the selection and optimization of Lanthanide materials for radiation shielding applications, paving the way for enhanced safety and efficiency in the nuclear and medical fields. Additionally, it establishes a foundation for investigating how the incorporation of rare earth elements into glass structures may alter their properties. Future studies should focus on exploring the practical applications of these findings in real-world settings, further advancing our understanding of radiation protection materials.

Table 2. Comparison of LAC Values for Lanthanide Elements: EpiXS vs. GATE Simulations.

Elements		Photon Energy (keV)				
		200	500	1000	1500	2000
		LAC (cm^{-1})				
Lanthanum (La)	**GATE**	2.5501	0.6311	0.3515	0.2864	0.2453
	EpiXS	2.6194	0.6238	0.3605	0.2848	0.2530
Cerium (Ce)	**GATE**	2.8613	0.6666	0.3928	0.3165	0.2743
	EpiXS	2.9784	0.6929	0.3963	0.3125	0.2776
Praseodymium (Pr)	**GATE**	3.0364	0.7100	0.4075	0.3353	0.2868
	EpiXS	3.1875	0.7250	0.4101	0.3227	0.2868
Neodymium (Nd)	**GATE**	3.2844	0.7574	0.4323	0.3378	0.3051
	EpiXS	3.4094	0.7586	0.4246	0.3334	0.2963
Promethium (Pm)	**GATE**	3.6717	0.7845	0.4314	0.3552	0.3150
	EpiXS	3.7147	0.8087	0.4480	0.3509	0.3119
Samarium (Sm)	**GATE**	3.8053	0.8212	0.4467	0.3465	0.3293
	EpiXS	3.9214	0.8361	0.4581	0.3580	0.3182
Europium (Eu)	**GATE**	2.7552	0.5754	0.3294	0.2404	0.2183
	EpiXS	2.8555	0.5959	0.3233	0.2521	0.2241
Gadolinium (Gd)	**GATE**	4.2475	0.8865	0.4586	0.3909	0.3291
	EpiXS	4.3897	0.9000	0.4821	0.3750	0.3333
Terbium (Tb)	**GATE**	4.6163	0.9108	0.4922	0.4021	0.3486
	EpiXS	4.7724	0.9587	0.5085	0.3946	0.3507
Dysprosium (Dy)	**GATE**	4.9806	0.9729	0.5124	0.4072	0.3493
	EpiXS	5.1120	1.0096	0.5288	0.4094	0.3638
Holmium (Ho)	**GATE**	5.2767	1.0191	0.5310	0.4063	0.3852
	EpiXS	5.4583	1.0599	0.5488	0.4238	0.3765
Erbium (Er)	**GATE**	5.7529	1.0885	0.5462	0.4268	0.3787
	EpiXS	5.8416	1.1146	0.5711	0.4397	0.3907
Thulium (Tm)	**GATE**	6.1296	1.1461	0.5856	0.4422	0.4031
	EpiXS	6.2485	1.1746	0.5951	0.4566	0.4056
Ytterbium (Yb)	**GATE**	4.6430	0.8474	0.4278	0.3301	0.2938
	EpiXS	4.7464	0.8774	0.4402	0.3369	0.2991
Lutetium (Lu)	**GATE**	6.9433	1.2390	0.6108	0.4613	0.4308
	EpiXS	7.0377	1.2836	0.6356	0.4850	0.4305

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] F. Wall, "Rare earth elements," in *Encyclopedia of Geology*, 2nd ed., D. Alderton and S. Elias, Eds., Elsevier, pp. 680–693, 2021. [Online]. https://doi.org/10.1016/B978-0-08-102908-4.00101-6.

[2] J. A. Giacalone, "The market for the 'not-so-rare' rare earth elements," *J. Int. Energy Policy*, vol. 1, no. 1, pp. 11–18, 2012. [Online]. Available: https://doi.org/10.19030/jiep.v1i1.7013.

[3] S. B. Castor and J. B. Hedrick, "Rare earth elements," in *Materials Science and Chemistry*, G. E. Totten, Ed., CRC Press, 2006. [Online]. Available: https://doi.org/10.1002/0470862106.id683.

[4] P. F. Lou, X. B. Teng, Q. X. Jia, Y. Q. Wang, and L. Q. Zhang, "Preparation and structure of rare earth/thermoplastic polyurethane fiber for X-ray shielding," *J. Appl. Polym. Sci.*, vol. 136, no. 17, p. 47435, 2018.

[5] T. Cui, S. Duan, R. Chen, R. Wang, and Q. Jia, "Monte Carlo simulation study of rare earth/polypropylene composite shielding 120 KV medical X-ray," *J. Phys.: Conf. Ser.*, vol. 2539, no. 1, p. 012070, Jul. 2023.

[6] Y. Liu, X. Li, Y. Yin, Z. Li, H. Yao, Z. Li, and H. Li, "Design and computational validation of γ-ray shielding effectiveness in heavy metal/rare earth oxide–natural rubber composites," *Polymers*, vol. 16, no. 15, p. 2130, 2024.

[7] F.-H. Mai, Q.-P. Zhang, R. Wang, L.-C. Meng, Y. Zhang, J.-L. Li, P.-Q. Liu, Y.-T. Li, and Y.-L. Zhou, "Polymer fibers highly filled with styrene maleic anhydride-modified PbWO4 for improved wear comfort of γ-ray-shielding articles," *ACS Appl. Polym. Mater.*, vol. 4, pp. 6394–6402, 2022.

[8] M. Şahiner, Y. Z. Akgök, M. Arslan, and M. H. Ergin, "Dünyada ve Türkiye'de nadir toprak elementleri (NTE)," *Maden Tetkik ve Arama Genel Müdürlüğü*, Fizibilite Etütleri Daire Başkanlığı, 2017.

[9] Y. S. Rammah, A. A. Ali, R. El-Mallawany, and F. I. El-Agawany, "Fabrication, physical, optical characteristics and gamma-ray competence of novel bismuth-borate glasses doped with Yb_2O_3 rare earth," *Physica B: Condens. Matter*, vol. 583, p. 412055, 2020.

[10] Y. S. Rammah, F. I. El-Agawany, K. A. Mahmoud, R. El-Mallawany, E. Ilik, and G. Kilic, "FTIR, UV–Vis–NIR spectroscopy, and gamma rays shielding competence of novel ZnO-doped vanadium borophosphate glasses," *J. Mater. Sci.: Mater. Electron.*, vol. 31, no. 12, pp. 9099–9113, 2020.

[11] H. Jing, L. Geng, S. Qiu, H. Zou, M. Liang, and D. Deng, "Research progress of rare earth composite shielding materials," *J. Rare Earths*, vol. 41, no. 1, pp. 32–41, 2023.

[12] P. Fernandez-Arias, D. Vergara, and J. A. Orosa, "A global review of PWR nuclear power plants," *Appl. Sci.*, vol. 10, no. 13, p. 4434, 2020.

[13] P. Wu, M. Liu, J. G. Deng, J. J. Fu, Z. H. Deng, and Y. Q. Song, "Research progress of neutron shielding material," *Nucl. Chem. Mater.*, vol. 45, no. 9, pp. 4–8, 2017.

[14] A. A. Suresh, P. Vinothkumar, M. Mohapatra, M. Dhavamurthy, and P. Murugasen, "The effect of rare earth on the radiation shielding properties of transparent lead-free alumino-borophosphate glass system," *Radiat. Phys. Chem.*, vol. 193, p. 109941, 2022.

[15] E. S. A. Waly, G. S. Al-Qous, and M. A. Bourham, "Shielding properties of glasses with different heavy elements additives for radiation shielding in the energy range 15–300 keV," *Radiat. Phys. Chem.*, vol. 150, pp. 120–124, 2018.

[16] G. Lakshminarayana et al., "Investigation of structural, thermal properties and shielding parameters for multicomponent borate glasses for gamma and neutron radiation shielding applications," *J. Non-Cryst. Solids*, vol. 471, pp. 222–237, May 2017.

[17] R. Florez, A. Loaiza, C. H. C. Giraldo, and H. A. Colorado, "Calcium silicate phosphate cement with samarium oxide additions for neutron shielding applications in the nuclear industry," *Prog. Nucl. Energy*, vol. 133, p. 103650, 2021.

[18] M. Ma, M. E. Thomas, P. McGuiggan, and J. B. Spicer, "Weak absorption and scattering losses from the visible to the near-infrared in single-crystal sapphire materials," *Opt. Eng.*, vol. 59, no. 8, p. 87101, 2020.

[19] K. M. Mahmoud and Y. S. Rammah, "Investigation of gamma-ray shielding capability of glasses doped with Y, Gd, Nd, Pr, and Dy rare earth using MCNP-5 code," *Physica B: Condens. Matter*, vol. 577, p. 411756, 2020.

[20] P. Kaur, D. Singh, and T. Singh, "Gamma rays shielding and sensing application of some rare earth doped lead-alumino-phosphate glasses," *Radiat. Phys. Chem.*, vol. 144, pp. 336–342, 2018.

[21] W. F. Yang, Y. Liu, L. Yang, D. Li, and J. Li, "Research progress in shielding materials for nuclear radiation," *Mater. Rev.*, vol. 21, no. 5, pp. 82–87, 2007.

[22] B. Gan, S. C. Liu, Z. He, F. C. Chen, H. X. Niu, J. C. Cheng et al., "Research progress of metal-based shielding materials for neutron and gamma rays," *Acta Metall. Sin.*, vol. 7, pp. 68–75, 2021.

[23] S. Zhao, Z. P. Huo, G. Q. Zhong, H. Zhang, and L. Q. Hu, "Research progress of neutron and gamma-ray composite shielding materials," *J. Funct. Mater.*, vol. 52, no. 3, pp. 3001–3010, 2021.

[24] G. Jagannath, A. G. Pramod, K. Keshavamurthy, B. N. Swetha, B. Bheemaiah, R. Eraiah, R. Rajaramakrishna, R. Ramesh, P. Prajwal, V. Hegde, S. C. Prashantha, and M. S. Alhuthali, "Nonlinear optical, optical limiting and radiation shielding features of Eu^{3+} activated borate glasses," *Optik*, vol. 232, p. 166563, 2021, doi: 10.1016/j.ijleo.2021.166563.

[25] O. I. Sallam, Y. S. Rammah, I. M. Nabil et al., "Enhanced optical and structural traits of irradiated lead borate glasses via Ce^{3+} and Dy^{3+} ions with studying radiation shielding performance," *Sci. Rep.*, vol. 14, p. 24478, 2024, doi: 10.1038/s41598-024-73892-w.

[26] F. C. Hila, A. Asuncion-Astronomo, C. A. M. Dingle, J. F. M. Jecong, A. M. V. Javier-Hila, M. B. Z. Gili, C. V. Balderas, G. E. P. Lopez, N. R. D. Guillermo, and A. V. Amorsolo Jr., "EpiXS: A Windows-based program for photon attenuation, dosimetry and shielding based on EPICS2017 (ENDF/B-VIII) and EPDL97 (ENDF/B-VI.8)," *Radiat. Phys. Chem.*, vol. 182, p. 109331, 2021.

[27] PubChem, "Periodic Table," *National Center for Biotechnology Information*, [Online]. Available: https://pubchem.ncbi.nlm.nih.gov/periodic-table/. [Accessed: Aug. 3, 2024].

[28] T. Şahmaran and N. Yavuzkanat, "Evaluation of gamma and neutron radiation shielding parameters of some glass materials by Monte Carlo and theoretical methods," *Radiat. Eff. Defects Solids*, vol. 179, no. 3–4, pp. 489–500, 2024.

[29] American National Standards Institute (ANSI), "Gamma-ray attenuation coefficient and buildup factors for engineering materials," *ANS/ANSI-6.4.3*, 1991.

[30] Y. Karabul, "Bazalt numunelerde EABF ve EBF parametrelerinin yeni bir metot ile tayini," M.S. thesis, İstanbul Univ., Inst. of Science, Istanbul, 2014, pp. 13–14.

[31] P. S. Singh, T. Singh, and P. Kaur, "Variation of energy absorption buildup factors with incident photon energy and penetration depth for some commonly used solvents," *Ann. Nucl. Energy*, vol. 35, pp. 1093–1097, 2008.

[32] S. Bilici, A. Bilici, and F. Külahcı, "Comparison photon exposure and energy absorption buildup factors of CR-39 and Trivex optical lenses," *Turk. J. Sci. Technol.*, vol. 17, no. 1, pp. 23–35, 2022.

[33] N. Yavuzkanat, "Fosfor-sandviç tipi dedektör sisteminin toplam gama verimi için modellenen Geant4 tabanlı GATE simülasyonu," *ALKÜ Fen Bilimleri Dergisi,* vol. 2, no. 3, pp. 150–162, 2020, doi: 10.46740/alku.804473.

[34] N. Nagaraj, H. C. Manjunatha, Y. S. Vidya, L. Seenappa, K. N. Sridhar, and P. S. Damodara Gupta, "Investigations on lanthanide polymers for radiation shielding purpose," *Radiat. Phys. Chem.*, vol. 199, p. 110310, 2022, doi: 10.1016/j.radphyschem.2022.110310.

[35] A. Pyngrope, Study of Radiation Shielding Properties of Lanthanides, *Solvent and Glass*, 2023, ch. 4, pp. 26–38. ISBN: 978-620-6-76775-6.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1325-1334 YEAR: 2024
DOI: 10.17798/bitlisfen.1561303

An Investigation of the Fresh and Hardened Properties of Nano Zinc Oxide Reinforced 3D Printed Geopolymer Mortars

Maksut SELOĞLU[1*]

[1] *Firat University, Elazığ Organize Sanayi Vocational School, Construction Department, Elazig, Türkiye*
(ORCID: 0000-0002-0200-8423)

Keywords: 3D-printed geopolimer mortar, Nano zinc oxide, Compressive strength, Flexural strength, Buildability.

Abstract
It is well known that even small amounts of nanomaterials can improve the mortar structure and enhance its fresh state and hardened properties. This paper investigates the fresh state and hardened properties of nano zinc oxide-reinforced 3D-printed geopolymer mortars. The mechanical properties of 7, 28, 90, and 180 days of 3D-printed geopolymer mortars cured at ambient temperature were investigated. For this purpose, 3D-printed geopolymer mortar samples containing 0%, 0.25%, 0.50%, and 0.75% nano zinc oxide were produced. Flow table and buildability tests were applied to these samples to determine the fresh state properties. Ultrasonic pulse velocity, flexural strength, and compressive strength tests were applied to the hardened 3D-printed geopolymer mortar samples. The best mechanical test results were obtained from 3D-printed geopolymer mortar samples containing 0.5% nano zinc oxide at the end of all curing times. In the ZN 50 series cured for 28 days, approximately 29% higher strength was obtained in FS and 66% higher in compressive strength compared to the ZN 0 series without nanomaterials. It has been noted that incorporating a tiny quantity of nano zinc oxide into 3D-printed geopolymer mortars improves their mechanical performance.

1. Introduction

Concrete is the most commonly used building material in the world, especially in the construction sector. One of the most commonly used materials in building structures is mortar. Mortar is obtained by mixing binders such as sand and cement with water. Mortar and concrete are the second most consumed materials in the world after water. One of the main industries that produce greenhouse gases is the cement industry. This cement production has a large carbon footprint and has grown to be a major source of pollution worldwide in recent decades [1].

Geopolymer (GP) is inorganic polymeric binder materials, first developed by Davidovits in the 1970s [2]. GP is an aluminum and silicate type binder material formed by activating natural and waste pozzolanic materials with various alkaline activators [3]. GPs are highly acknowledged in the concrete industry as superior cement alternatives, providing an environmentally benign and sustainable building solution. GP composites are environmentally friendly because they are sustainable, use waste materials, and do not contain Portland cement. GP mortar or concrete is known as environmentally friendly mortar or environmentally friendly concrete because it uses one or more components that are considered industrial waste materials with pozzolanic quality and limited landfill areas during production. GP mortars and concretes, can be produced by activating one or more of the following materials with alkali activators: industrial wastes such as fly ash (FA), rice husk ash, silica fume, ground granulated blast furnace slag, fired clays and shales, heat-treated materials such as metakaolin (MK), natural pozzolans such as volcanic ashes, trass, and diatomite soils, pumice, which is a type of volcanic glassy rock and also known as pumice stone, and ground perlite, which is a volcanic rock [4].

*Corresponding author: mseloglu@firat.edu.tr
Received: 04.10.2024, Accepted: 03.12.2024

Three-dimensional printing technology (3DPT), which has a very widespread use, has also found a place in the construction sector. The first inventor of 3DPT C. Hull, received his patent in 1986 and has shown considerable development and progress to this day [5]. With global warming and the deterioration of the ecological balance, which are among the biggest problems in the world, solutions have been sought in the construction sector for environmentally friendly production techniques, and innovative methods have been developed thanks to the opportunities provided by technological developments [6], [7]. One of these is the start of production with 3DPT. With 3D printing, a three-dimensional object modeled in a computer environment is divided into layers, and during the production phase, each layer is built on top of the previous layer and maintains its shape. Despite its high cost and complex structure, it has become widespread enough to be used in almost every field today compared to the period when it was first produced. So much so that even products that cannot be produced with classical manufacturing methods can be easily produced with 3D printers [8].

It is imperative to apply sophisticated techniques to improve the structural performance of GP mortars by adding nanomaterials. Nanomaterials have gained importance in the building sector as a way to enhance mortar performance. It has been found that the structural performance and lifespan of GP mortars can be greatly enhanced by the addition of tiny amounts of nanomaterials [1]. Many properties vary according to the different nanomaterials used. Nanomaterials are added to GP mortars to improve their mechanical properties. Theoretically, all materials in the aluminosilicate class can be activated by alkali activators and exhibit binding properties. It is known from the literature that nanomaterials added to very small amounts of alkali activators increase the mechanical performances of GP mortars [9]. The use of nanomaterials in three dimensional printed geopolimer (3DGP) mortar and concrete samples has increased recently [10]–[17]. Nano zinc oxide (n-ZN) is a promising nanomaterial because of its good semiconductor capabilities, and its ability to improve microstructure by acting as a filler and increasing mechanical strength [18]. It is also easy and economical to produce and safe for the environment [19]. However, the use of n-ZN in GP mortar samples is limited [18], [20]–[24]. Also, the use of n-ZN in 3DGP mortar samples is very limited [25]. More researchers have been examining the mechanical strength and durability properties of GP composites utilizing n-ZN in recent years [26], [27]. It is known from the literature that n-ZN has an improving effect on mechanical performance [9]. In their investigation, Raj et al. [28] found that n-ZN containing GP composites had a greater mechanical strength than GP composites. The fire resistance of plywood coated with a GP containing n-ZN was investigated by Wang et al [29]. They discovered that a flame is not necessary when utilizing huge volumes of n-ZN. Slag-based GPs were compared to n-ZN, nano-iron, and hybrid nanomaterials in terms of their physical characteristics, mechanical strength, and radiation shielding capabilities by Mohsen et al [30]. Comparing the hybrid nanomaterial series against the n-ZN series, they discovered that the hybrid nanomaterial series produced superior results in mechanical strength and radiation protection qualities. Seloglu et al. [9] compared the mechanical strength of different nanomaterials in MK-FA-based GP mortars. They found that the best results were obtained in the GP mortar series containing 0.5% n-ZN that were cured at ambient temperature for 7 and 28 days. Zidi et al.'s investigation [19] looked at how nano alumina, n-ZN, and nano silica affected the mechanical characteristics of GP composites. They stated that using n-ZN to get outcomes that were both cost-effective and highly strength. Tan et al.'s investigation [31] examined how n-ZN affected the slag-based GP's engineering characteristics. They stated that the mechanical qualities of the composite were reduced when the percentage exceeded 2%. Tanyildizi, et al. [25] investigated the effects of using n-ZN in 3DGP mortars on mechanical and durability effects. They found that the highest mechanical properties were obtained from samples containing %0.5 n-ZN.

In this experimental study, fresh and hardened properties of n-ZN reinforced FA-MK-based 3DGP mortar samples were investigated. Within the scope of the study, four different n-ZN reinforced 3DGP mortars were produced and added to the composite by volume at 0%, 0.25%, 0.50%, and 0.75% ratios. For the fresh state properties of 3DGP mortar samples, the mini-slump spread test and buildability test, which is an important test in 3D printing, were performed. To examine the hardened properties, Ultrasonic pulse velocity (UPV), flexural strength (FS), and compressive strength (CS) tests were performed by the relevant standards. The study aims to investigate the rheological and mechanical properties of 3DGP mortars produced without molds and without using cement as a binder by curing at ambient temperatures for 7, 28, 90, and 180 days. The result of this research will be helpful to structural designers who want to build sustainable 3DGP mortar.

2. Materials and Method

2.1. Materials

In the current study, FA and MK were employed as binders. Table 1 presents the properties and composition of binder materials.

To manufacture the geopolymer (GP) mortar, a binary activator consisting of Na_2SiO_3 (SM) and NaOH (SH) solution was used. The SH pellets were dissolved in water and added to the SM before the creation of the GP mortars. This process began 24 hours before the production of the GP mortar to ensure the development of a 12 M alkaline solution. Fine river Palu sand was added to this mix as an aggregate. Table 2 presents the properties of aggregate.

MK requires extra water in the GP mortar because of its clay nature. For this purpose and to ensure workability, additives were added to these GP mortars at a rate of 1% of the binder amount. n-ZN was added in increments of %0, %0.25, %0.50, and %0.75. These mixtures containing n-ZN are expressed as ZN 0, ZN 25, ZN 50, and ZN 75 in the study, respectively. Table 3 presents the properties of additive, Table 4 presents the properties of n-ZN, and Table 5 presents the GP mortar mix design in detail.

Table 1. Properties and composition of binder materials

Binder	LOI	MgO	Al_2O_3	Na_2O	K_2O	CaO	Fe_2O_3	SiO_2	Specific gravity
FA	0.91	2.42	22.95	0.92	0.99	2.58	7.25	58.25	2.31 (g/cm^3)
MK	1.11	0.16	40.25	0.24	0.55	0. 91	0.85	56.10	2.52 (g/cm^3)

Table 2. Properties of aggregate

	Grain diameter	Water absorption	Specific gravity
Fine sand	0. 6 mm	2.1	2.72 (g/cm^3)

Table 3. Properties of additive

	Type	Color	Specific gravity
Additive	Viscosity regulator	Green liquid	1.01 (g/cm^3)

Table 4. Properties of n-ZN

Nano-ZnO	
Chemical formula	ZnO
Surface area (m^2/g)	20-65
Color	White
Average particle diameter (nm)	18
Chemical structure	Crystal
Purity (%)	99.99
Shape	Near spherical

Table 5. GP mortar mix design (by mass ratio)

Mix	FA	MK	Nano-ZnO (%)	SM	SH	SS/SH	AAS/B	Aggregate	VRA (%)
ZN 0	0.50	0.50	0.00	0.55	0.25	2.20	0.80	2.57	0.01
ZN 25	0.50	0.50	0.25	0.55	0.25	2.20	0.80	2.57	0.01
ZN 50	0.50	0.50	0.50	0.55	0.25	2.20	0.80	2.57	0.01
ZN 75	0.50	0.50	0.75	0.55	0.25	2.20	0.80	2.57	0.01

***SM:** Na_2SiO_3**, SH:** NaOH**, AAS/B:** alkaline activator solution/binder**, VRA:** viscosity regulator additive

2.2 Mix Design and Method

The GP mortar mixing procedure used in this study follows the steps used in the author's previous study [25]. While printing, the printer speed was set to 80 cm/minute, and a 0.9 x 1 cm rectangular nozzle was employed. The samples were exposed to ambient conditions for 7, 28, 90, and 180 days. Five layers of GP mortar specimens were fabricated by 3D printer, with the layer height of each GP mortar sample designed as approximately 0.8 cm. Figure 1 shows the flow chart of the GP mortar sample produced with a 3D printer.

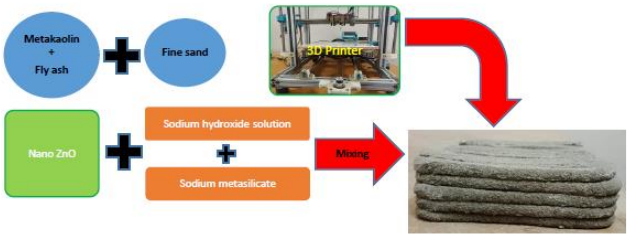


Figure 1. The flow chart of 3DGP mortar

For this study, forty-eight prismatic GP mortar mixture samples of four different types, each measuring 4 cm x 4 cm x 16 cm, were prepared. These samples were then kept under 23±2 °C ambient temperature curing conditions and kept until they reached the specified test ages (7, 28, 90, and 180 days). At the end of the curing periods, the GP mortar samples produced with the 3D printer were subjected to fresh state and hardened property tests.

2.3 Experimental Procedure

2.3.1 Fresh State Testing Procedure

The flowability properties of the produced GP mortars were determined by the flow table test. It was carried out according to ASTM C230 [32]. Fresh GP mortar mixtures were filled into the spreading table in two layers. Each layer was compacted 20 times with a tamper. Then, it was slowly removed and the flow table was lowered 25 times in 15 seconds to ensure spreading on the table. The diameter of the spread fresh GP mortar mass was measured in two perpendicular directions and the measured values were recorded in cm. Then, the flow table was determined by taking the arithmetic average of the values obtained in these two perpendicular directions for each fresh GP mortar sample. The test device used to determine the flowability properties and the GP mortar on which the flow table test was applied are given in Figure 2.

a) **b)**

Figure 2. a) The flow table test device [25], **b)** GP mortar sample

Buildability is the ability to 3D print by adding mortar layer by layer continuously to the required level without significant deformation or collapse of the freshly printed component [33]. The buildability tests, which are of great importance in terms of extrudability and printability properties of 3DGP mortar specimens produced with a 3D printer, were applied. It was carried out according to ASTM C1437 [4]. This test determines the shape retention of fresh GP mortar samples subjected to static load. Each fresh GP mortar mixture was placed in a mini-slump mold shortly after the first mixing. The mold was removed after a waiting period of 60 seconds. To ensure a smooth and equal distribution of the load, a glass plate was placed on the fresh GP mortar sample. Then, a 0.6 kg load was applied with the glass plate weight for 60 seconds. At the end of 60 seconds, the deformation of the fresh GP mortar samples was measured in two perpendicular directions, taking into account the final heights of the samples. The arithmetic averages of these two measured values were taken and the average height values were recorded. At the end of the test, it is accepted that the mixtures with higher final height values can be constructed better. The GP mortar on which the buildability test was applied is given in Figure 3.

Figure 3. The buildability test GP mortar sample.

2.3.2 Hardened State Testing Procedure

UPV and mechanical strength (FS and CS) tests were applied to hardened GP mortar samples produced with a 3D printer. Forty-eight prismatic samples, each

measuring 4 cm x 4 cm x 16 cm, for four different GP mortars made up the test for this study. Prismatic specimens measuring 4 x 4 x 16 cm were subjected to FS testing according to ASTM C348 [35] standards using a loading rate of 0.2 kN/s. Three samples were tested for each mixture. At the end of the curing periods, the FS of the GP mortar specimens produced with the 3D printer belonging to each series were obtained by taking the arithmetic average of three samples. Then, the CS of the 3DGP mortar samples were determined in an automatically controlled press according to the ASTM C349 [36] standard. The 3DGP mortar samples, which were cured at ambient temperature and subjected to CS tests at the end of the 7, 28, 90, and 180-day curing periods, were the samples subjected to FS tests and were divided into two for each sample after the test. The arithmetic mean of six specimens was used to determine the CS of the 3DGP mortar specimens that belonged to each series. Figure 4 shows 3DGP mortar samples subjected to FS and CS tests using a mechanical testing device.

a) **b)**

Figure 4. **a**) 3DGP mortar samples subjected to FS, **b)** 3DGP mortar samples subjected to CS

3. Results and Discussion

3.1 Fresh Properties Test Results

3.1.1 Flowability Test Results

The flowability test results of 3DGP mortar samples are given in Figure 5.

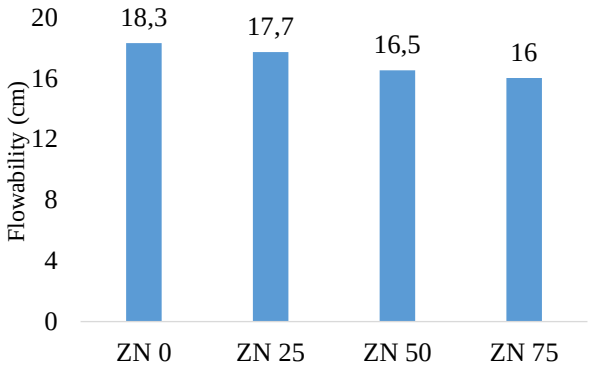


Figure 5. The flowability test results of 3DGP mortar samples

When Figure 5 is examined, the flow tables in the ZN 0, ZN 25, ZN 50, and ZN 75 series were found to be 18.3, 17.7, 16.5, and 16 cm, respectively. n-ZN reinforcement reduces the flowability properties and decreases the flow table values in GP mortars. Nanomaterials have reduced workability due to their large surface area. They also react rapidly with alkaline activators, providing the creation of a viscous, and extremely adhesive GP matrix [28].

3.1.2 Buildability Test Results

Figure 6 displays the outcomes of the 3DGP mortar samples' buildability test.

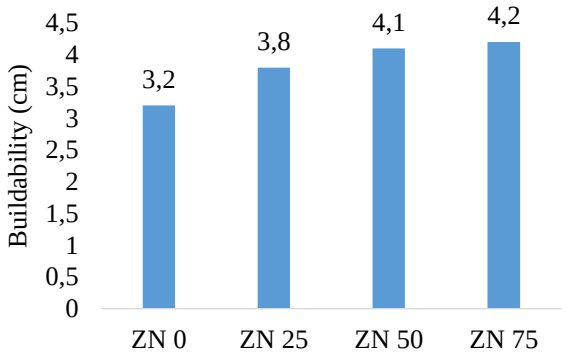


Figure 6. The buildability test results of 3DGP mortar samples

When Figure 6 is examined, the buildability test results of the ZN 0, ZN 25, ZN 50, and ZN 75 series were found to be 3.2, 3.8, 4.1, and 4.2 cm, respectively. Buildability benefits from the quick improvement in yield strength of GP composites that comes from the addition of nanoparticles [33]. In this study, n-ZN was utilized to improve buildability and strength. The literature has investigated the utilization of nanoparticles to boost the yield strength of 3DGP samples. To enhance the buildability of GP composites without considerably compromising their pumpability, it was acceptable to incorporate the nanoparticles during the mixing stage [37]. n-ZN reinforcement increases the buildability properties in 3DGP mortars [25]. This increase was found to be 18.75%, 28.13%, and 31.25% in the series containing n-ZN (ZN 25, ZN 50, and ZN 75) compared to the ZN 0 control series without n-ZN, respectively. As a result, it was discovered that n-ZN use in GP composites was crucial for buildability.

3.2 Hardened Properties Test Results

3.2.1 UPV Test Results

To learn about the strength characteristics of 3DGP mortar specimens, UPV tests were used. The UPV

test results of 3DGP mortar samples are given in Figure 7.

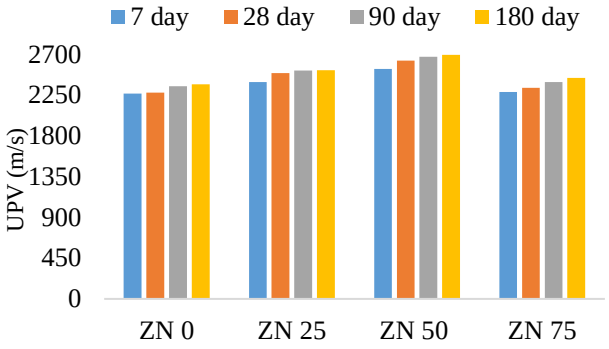


Figure 7. The UPV test results of 3DGP mortar samples

When Figure 7 is examined, the UPV test results of the ZN 0, ZN 25, ZN 50, and ZN 75 series increased in all series in parallel with the increase in cure times. This increase was found to be 10%, 15%, and 2%, respectively, in the series containing n-ZN (ZN 25, ZN 50, and ZN 75) cured at ambient temperature for 28 days, compared to the ZN 0 control series without n-ZN. The highest UPV test values were recorded in the ZN 50 series containing 0.5% n-ZN at the end of all cure times. These results are also consistent with the literature [20]. In the literature, the CS of the specimens improved by 0.5% with the addition of n-ZN, but the CS of the samples reduced with the addition of 0.7% [20]. UPV outcomes are often parallel to CS and are assumed to be related to CS [25]. The UPV test results of the ZN 50 series cured at ambient temperature for 7, 28, 90, and 180 days were found to be 2538 m/s, 2631 m/s, 2674 m/s, and 2695 m/s, respectively.

3.2.2 FS Test Results

The FS test results of 3DGP mortar samples are given in Figure 8.

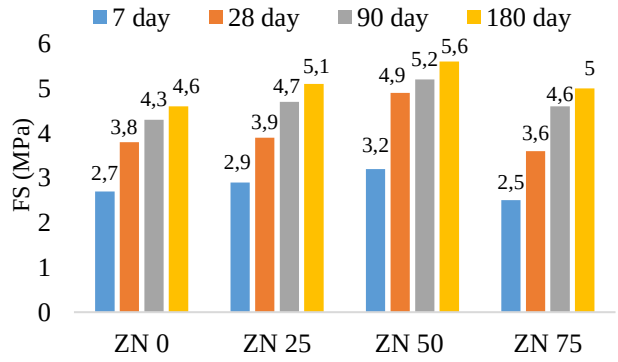


Figure 8. The FS test results of 3DGP mortar samples

When Figure 8 is examined, the FS test results of the ZN 0, ZN 25, ZN 50, and ZN 75 series increased in all series in parallel with the increase in all cure times. This increase was found to be 75%, 63%, and 53%, respectively, in the series containing n-ZN (ZN 25, ZN 50, and ZN 75) cured at ambient temperature for 28 days, compared to the ZN 0 control series without n-ZN. The highest FS test results were obtained in the ZN 50 series containing 0.5% n-ZN at the end of all cure times. Sobhy et al. [38] found that the inclusion of 0.5% n-ZN increased the FS at the end of 28-day curing times by approximately 25%. The worst effect of NZ is that it reduces the mechanical strength and hydration degree when used in composites at a rate higher than 0.5%; all of these are due to the reaction between ZnO and CH. Therefore; the reinforcement of NZ at a rate higher than 0.5% has a bad effect on the mechanical properties of the composite. The literature also has instances of this circumstance. [39, 40].

3.2.3 CS Test Results

The CS test results of 3DGP mortar samples are given in Figure 9.

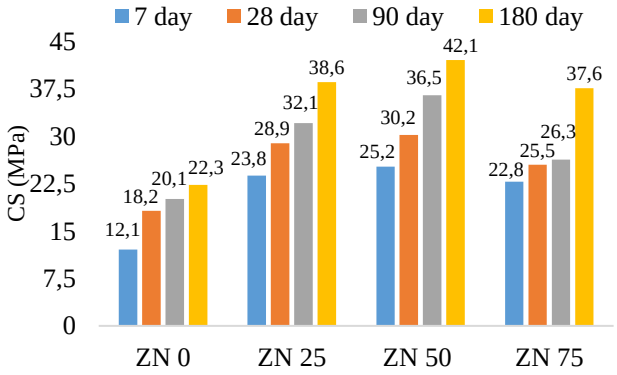


Figure 9. The CS test results of 3DGP mortar samples

When Figure 9 is examined, the CS test results of the ZN 0, ZN 25, ZN 50, and ZN 75 series increased in all series in parallel with the increase in all cure times. This increase was found to be 67.1%, 44.8%, and 19.8%, respectively, in the series containing n-ZN (ZN 25, ZN 50, and ZN 75) cured at ambient temperature for 28 days, compared to the ZN 0 control series without n-ZN. The highest CS test results were obtained in the ZN 50 series containing 0.5% n-ZN at the end of all cure times. n-ZN contributed to the increase in strength by filling the void structures. If homogeneous distribution is provided, no sedimentation and no agglomeration is observed, the type of nanomaterial increases the mechanical performance of the GP mortar depending

on the type and rate of addition. The results obtained from this study are supported by the literature [20], [25], [28], [40]. Therefore, the optimum amount of nano zinc oxide to be used in such composites is thought to be 0.5%.

Zidi et al.'s study [20], indicated that when the n-ZN was 0.5%, the CS test was at its highest. They claimed that the decrease in the interfacial transition zone caused by the n-ZN particles' and GPs matrix's interfacial adhesion is the cause of this. Additionally, they stated that at a dosage of 0.7% for n-ZN, the CS dropped as a result of improperly disseminated nanoparticle aggregation [20]. Other research, in addition to this one, has shown that the mechanical properties are decreased when n-ZN is used in amounts more than 0.5% [28], [41]. Zailan et al. [18] found the highest CS test results in 2.5% n-ZN reinforced GP composites cured at ambient temperature for 28 days. In this study, the highest CS was obtained in GP mortars containing 0.5% n-ZN. Raj et al. [28] found that the inclusion of 0.5% n-ZN increased the CS by approximately 26% [23]. Zidi et al. [19], in their paper comparing the cost analyses of different nanomaterials, found that GP samples containing 0.5% n-ZN increased the CS by 26.7%. As in previous research, the maximum CS in this sample came from 3DGP mortar containing 0.5% n-ZN. In this study, approximately 20% CS increase was recorded in 3DGP mortar samples. The CS of the ZN 50 series cured at ambient temperature for 28 days was found to be 30.2 MPa. n-ZN created a filler effect, filled the existing voids, and created a more impermeable structure, causing an increase in the strength of 3DGM samples. This increase is due to the high specific surface area and reactivity of nano-sized materials. As a result of pozzolanic reactions, the increase in strength became more pronounced over time. Therefore, as the curing time increased, the increase in strength continued at 7, 28, 90, and 180 days. In addition, since production is done without using molds, it is more economical and faster, which are other advantageous results in addition to its CS. This strength is sufficient and acceptable for many construction applications.

4. Conclusion

This study investigated the fresh state properties and hardened mechanical strength properties of GP mortars produced with n-ZN reinforcement and cured at ambient temperature conditions. ZN 0 series without nanomaterials were accepted as control samples and the experimental results obtained with n-ZN reinforced 3DGP mortar samples were compared. The obtained results are given below:

1- Buildability test results increased and GP mortar sample flowability values decreased when n-ZN was added to mixtures. The ZN 0 samples produced the greatest flowability and lowest buildability findings. The ZN 75 samples yielded the highest buildability and lowest flowability scores as well. The rheological characteristics of 3DGP mortar samples were found to be improved by the usage of n-ZN in this investigation.

2- The highest UPV results were obtained from the ZN 50 series, with approximately a 15% increase compared to the ZN 0 control samples cured for 28 days. It was determined that UPV results and CS results support each other as in traditional mortar. Therefore, it was concluded that the non-destructive test methods developed to determine the mechanical properties of traditional mortars are valid in 3DGP mortar samples.

3- The greatest FS results were obtained from the ZN 50 series, with approximately a 63% strength increase compared to the ZN 0 control samples cured for 28 days. This increase in FS of GP can be explained by the microstructure developed by the addition of zinc to the GP mortar and the formation of a denser GP mortar as a result of the polymerization reaction.

4- The peak CS of n-ZN reinforced 3DGP mortar samples is obtained using 0.5% nano zinc oxide. This can be attributed to the potential of nanomaterials to fill or minimize voids and the ability of nanozinc to increase the polymerization rate. 3DGP mortar samples containing 0.5% n-ZN gave the highest CS results at all cure times. Approximately 66% higher CS results were obtained from the ZN 50 series than the control samples cured for 28 days.

It has been observed that the fresh state and mechanical properties of 3DGP mortar samples exhibit good performance with a very small amount of n-ZN reinforcement. It is also thought that it would be useful to examine the durability properties of 3DGP mortar samples with n-ZN reinforcement.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] M. Usman, M. Arshad, and A. Raza, "Araştırma Makalesi Mechanical Performance of Microfiber-Reinforced Geopolymer Mortar with Nano-Titania," no. June, 2024.

[2] J. Davidovits and S. A. Cordi, "Synthesis of new high temperature geo-polymers for reinforced plastics/composites," *Spe Pactec*, vol. 79, pp. 151–154, 1979.

[3] H. Xu and J. S. J. Van Deventer, "The geopolymerisation of alumino-silicate minerals," *Int. J. Miner. Process.*, vol. 59, no. 3, pp. 247–266, 2000, doi: 10.1016/S0301-7516(99)00074-5.

[4] M. Seloğlu, "Investigation of the Mechanical and Durability Properties of Geopolymer Mortars Produced With a 3D Printer Based on Metakaolin and Fly Ash Containing Nanomaterial," Dicle University, 2024.

[5] B. Khoshnevis, "Automated construction by contour crafting - Related robotics and information technologies," *Autom. Constr.*, vol. 13, no. 1, pp. 5–19, 2004, doi: 10.1016/j.autcon.2003.08.012.

[6] D. Ahlers, "Development of a Software for the Design of Electronic Circuits in 3D-Printable Objects," no. December 2015, 2015, doi: 10.13140/RG.2.2.14755.60963.

[7] O. Diegel, A. Nordin, and D. Motte, *Additive Manufacturing Technologies*, 2019, doi: 10.1007/978-981-13-8281-9_2.

[8] S. Ö. Felek, "Mimari Yapılarda 3 Boyutlu Yazıcıların Kullanımı," *Int. J. 3D Print. Technol. Digit. Ind.*, vol. 3, no. 3, pp. 289–296, 2019. [Online]. Available: https://dergipark.org.tr/en/pub/ij3dptdi/issue/51591/630599

[9] M. Seloğlu, H. Tanyildizi, and M. E. Öncü, "An Investigation of the Strength Properties of Fly Ash and Metakaolin-Based Geopolymer Mortars Containing Multi-Wall Carbon Nanotube, Nano Silica, and Nano Zinc," *Bitlis Eren Üniversitesi Fen Bilim. Derg.*, vol. 12, no. 3, pp. 842–852.

[10] H. G. Şahin and A. Mardani-Aghabaglou, "Assessment of materials, design parameters and some properties of 3D printing concrete mixtures; a state-of-the-art review," *Constr. Build. Mater.*, vol. 316, no. November 2021, 2022, doi: 10.1016/j.conbuildmat.2021.125865.

[11] B. Panda, S. Ruan, C. Unluer, and M. J. Tan, "Investigation of the properties of alkali-activated slag mixes involving the use of nanoclay and nucleation seeds for 3D printing," *Compos. Part B Eng.*, vol. 186, no. November 2019, p. 107826, 2020, doi: 10.1016/j.compositesb.2020.107826.

[12] B. Panda, C. Unluer, and M. J. Tan, "Extrusion and rheology characterization of geopolymer nanocomposites used in 3D printing," *Compos. Part B Eng.*, vol. 176, no. July, p. 107290, 2019, doi: 10.1016/j.compositesb.2019.107290.

[13] G. X. Zhou *et al.*, "3D printing geopolymer nanocomposites structure: Graphene oxide size effects on a reactive matrix," *Carbon N. Y.*, vol. 164, pp. 215–223, 2020, doi: 10.1016/j.carbon.2020.02.021.

[14] H. Zhong and M. Zhang, "3D printing geopolymers: A review," *Cem. Concr. Compos.*, vol. 128, no. January, p. 104455, 2022, doi: 10.1016/j.cemconcomp.2022.104455.

[15] S. Qaidi, A. Yahia, B. A. Tayeh, H. Unis, R. Faraj, and A. Mohammed, "3D printed geopolymer composites: A review," *Mater. Today Sustain.*, vol. 20, p. 100240, 2022, doi: 10.1016/j.mtsust.2022.100240.

[16] M. H. Raza, R. Y. Zhong, and M. Khan, "Recent advances and productivity analysis of 3D printed geopolymers," *Addit. Manuf.*, vol. 52, no. January, p. 102685, 2022, doi: 10.1016/j.addma.2022.102685.

[17] S. H. Bong, M. Xia, B. Nematollahi, and C. Shi, "Ambient temperature cured 'just-add-water' geopolymer for 3D concrete printing applications," *Cem. Concr. Compos.*, vol. 121, no. April, p. 104060, 2021, doi: 10.1016/j.cemconcomp.2021.104060.

[18] S. N. Zailan, A. Bouaissi, N. Mahmed, and M. M. A. B. Abdullah, "Influence of ZnO nanoparticles on mechanical properties and photocatalytic activity of self-cleaning ZnO-based geopolymer paste," *J. Inorg. Organomet. Polym. Mater.*, vol. 30, pp. 2007–2016, 2020.

[19] Z. Zidi, M. Ltifi, and I. Zafar, "Comparative study: nanosilica, nanoalumina, and nanozinc oxide addition on the properties of localized geopolymer," *J. Aust. Ceram. Soc.*, vol. 57, no. 3, pp. 783–792, 2021.

[20] Z. Zidi, M. Ltifi, Z. Ben Ayadi, L. E. L. Mir, and X. R. Nóvoa, "Effect of nano-ZnO on mechanical and thermal properties of geopolymer," *J. Asian Ceram. Soc.*, vol. 8, no. 1, pp. 1–9, 2020.

[21] A. Nazari and S. Riahi, "The effects of ZnO2 nanoparticles on properties of concrete using ground granulated blast furnace slag as binder," *Mater. Res.*, vol. 14, pp. 299–306, 2011.

[22] C. B. Nayak, P. P. Taware, U. T. Jagadale, N. A. Jadhav, and S. G. Morkhade, "Effect of SiO2 and ZnO Nano-Composites on Mechanical and Chemical Properties of Modified Concrete," *Iran. J. Sci. Technol. Trans. Civ. Eng.*, vol. 46, no. 2, pp. 1237–1247, 2022, doi: 10.1007/s40996-021-00694-9.

[23] M. Rustan, Subaer, and Irhamsyah, "Studi Tentang Pengaruh Nanopartikel ZnO (Seng Oksida) Terhadap Kuat Tekan Geopolimer Berbahan Dasar Metakaolin," *J. Sains dan Pendidik. Fis.*, vol. 11, no. 3, pp. 286–291, 2015.

[24] M. Sarkar, M. Maiti, M. A. Malik, and S. Xu, "Development of anti-bio deteriorate sustainable geopolymer by SiO2 NPs decorated ZnO NRs," *Adv. Mater. Lett.*, vol. 10, no. 2, pp. 128–131, 2019.

[25] H. Tanyildizi, M. Seloglu, and A. Coskun, "The effect of nano zinc oxide on freeze-thaw resistance of 3D-printed geopolymer mortars," *J. Build. Eng.*, vol. 96, no. July, p. 110431, 2024, doi: 10.1016/j.jobe.2024.110431.

[26] M. F. Ali, M. T. Rashed, M. A. Bari, and K. M. Razi, "Effect of Zinc Oxide Nanoparticle on Properties of Concrete," *Int. Res. J. Eng. Technol.*, vol. 7, no. 2, pp. 1026–1029, 2020.

[27] R. S. Raj, G. P. Arulraj, N. Anand, B. Kanagaraj, E. Lubloy, and M. Z. Naser, "Nanomaterials in geopolymer composites: A review," *Dev. Built Environ.*, vol. 13, no. December 2022, p. 100114, 2023, doi: 10.1016/j.dibe.2022.100114.

[28] R. S. Raj, G. P. Arulraj, N. Anand, B. Kanagaraj, E. Lubloy, and M. Z. Naser, "Nanomaterials in geopolymer composites: A review," *Dev. Built Environ.*, vol. 13, p. 100114, 2023.

[29] L. Wang, W. Xiao, Q. Wang, H. Jiang, and G. Ma, "Freeze-thaw resistance of 3D-printed composites with desert sand," *Cem. Concr. Compos.*, vol. 133, p. 104693, 2022.

[30] A. Mohsen, H. A. Abdel-Gawwad, and M. Ramadan, "Performance, radiation shielding, and anti-fungal activity of alkali-activated slag individually modified with zinc oxide and zinc ferrite nano-particles," *Constr. Build. Mater.*, vol. 257, p. 119584, 2020, doi: 10.1016/j.conbuildmat.2020.119584.

[31] J. Tan, Z. Sierens, B. Vandevyvere, H. Dan, and J. Li, "Zinc oxide in alkali-activated slag (AAS): retardation mechanism, reaction kinetics and immobilization," *Constr. Build. Mater.*, vol. 371, p. 130739, 2023, doi: 10.1016/j.conbuildmat.2023.130739.

[32] C. Astm, "230, Standard specification for flow table for use in tests of hydraulic cement," *West Conshohocken, PA ASTM Int.*, 2008.

[33] S. Muthukrishnan, S. Ramakrishnan, and J. Sanjayan, "Technologies for improving buildability in 3D concrete printing," *Cem. Concr. Compos.*, vol. 122, p. 104144, 2021.

[34] C. ASTM, "Standard test method for flow of hydraulic cement mortar," *C1437*, 2007.

[35] A. ASTM, "C348-14 Standard Test Method for Flexural Strength of Hydraulic-Cement Mortars," *ASTM Int.*, West Conshohocken.

[36] A. ASTM, "C349-08: Standard test method for compressive strength of hydraulic-cement mortars (using portions of prisms broken in flexure)," *ASTM Int.*, West Conshohocken, PA, USA, 2008.

[37] B. Panda, S. Ruan, C. Unluer, and M. J. Tan, "Improving the 3D printability of high volume fly ash mixtures via the use of nano attapulgite clay," *Compos. Part B Eng.*, vol. 165, pp. 75–83, 2019.

[38] C. S. Sobhy *et al.*, "Insights on the influence of nano-Titanium dioxide and nano-Zinc oxide on mechanical properties and inhibiting of steel reinforcement," *Case Stud. Constr. Mater.*, vol. 16, p. e01017, 2022.

[39] M. Kumar, B. Manjeet, and G. Rishav, "An overview of beneficiary aspects of zinc oxide nanoparticles on performance of cement composites," *Mater. Today: Proc.*, vol. 43, pp. 892–898, 2021.

[40] M. R. Arefi, S. Rezaei-Zarchi, and S. Imani, "Synthesis of ZnO nanoparticles and their antibacterial effects," *African J. Biotechnol.*, vol. 11, no. 34, pp. 8520–8526, 2012.

Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1335-1354 YEAR: 2024
DOI: 10.17798/bitlisfen.1561354

Intrusion Detection and Performance Analysis Using Copula Functions

Mehmet BURUKANLI[1], Musa ÇIBUK[2 *]

[1] *Bitlis Eren University, Rectorate, Department of Common Courses, Bitlis, Türkiye*
[2] *Bitlis Eren University, Department of Computer Engineering, Bitlis, Türkiye*
(ORCID: 0000-0003-4459-0455) (ORCID: 0000-0001-9028-2221)

Keywords: Intrusion Detection, Copula Functions, KDD'99 Dataset, Naive Bayes Classifier, mRMR.

Abstract

Nowadays, interest in technology is growing as technology advances and makes our jobs easier. These rapid technological advancements bring with them a slew of unwanted negative attacks, such as cyber-attacks and unauthorized access. To prevent such negative attacks, intrusion detection systems are frequently used. In this research, we make some suggestions for novel and reliable classifiers for intrusion detection systems that are based on copulas. Using copula-based classifiers, we hope to detect intrusion in computer networks. Student's-t, Gumbel, Clayton, Gaussian, Independent and Frank classifiers, which are frequently used in the literature, have been preferred as copula-based classifiers. These classifiers were used to perform classification on the KDD'99 dataset. The 10-fold cross-validation method has been used in the classification phase. When the experimental results were examined, the proposed Gaussian copula-based classifier outperformed state-of-the-art basic methods on the KDD'99 dataset with a success rate of 99.41%. As a direct consequence of this, classifiers based on the copula have shown promising results in the field of intrusion detection. Classifiers that are based on the copula have been found to be a competitive alternative to the most recent and cutting-edge fundamental approaches.

1. Introduction

Today, the internet is an important communication tool that provides the flow of information between both personal and business relationships. This communication tool has also brought security risks. In particular, e-commerce applications made over the internet are exposed to serious attacks. These attacks cause significant damage to companies by causing loss of workforce, time and product in critical business applications [1]. Computer viruses and malware are examples of a few of these attacks. As a result of the attacks, information is lost and information that should be kept confidential may be disclosed. Security vulnerabilities on the Internet can cause great harm to web-based companies and public services. For this reason, companies and public service institutions increase their security measures day by day and have to make larger investments in order to take precautions against new threats [1]–[3]. Therefore, the tools that ensure the security of computer systems are gaining more and more importance, and especially the importance of Intrusion Detection Systems (IDS) is increasing. IDS helps to protect information systems against all kinds of attacks made over the network and is called any software or hardware components that have warning characteristics [2], [4]. By using IDS, attacks made over the network can be detected and prevented by activating the relevant mechanisms. There are many methods for performing IDSs. Some of these methods can be listed as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Tree (DT), and Ensemble Learning (EL). Apart from these methods, methods such as copula functions have also started to be applied [5]. In this study, attack detection has been performed by using copula functions, which is a new approach for the IDS domain. The classification algorithm inspired in this study has been first proposed by R. Salinas-Gutierrez et al. [6]. This classification algorithm, which is Gaussian copula-based, has been later generalized by M. Scavnicky [7] and a copula-based classification algorithm has been obtained. In this study, the use and performance of the

* Corresponding author: mcibuk@beu.edu.tr
Received: 04.10.2024, Accepted: 25.12.2024

algorithms developed on the basis of the copula-based classification approach, which has been generalized by M. Scavnicky [7], in the field of IDS have been examined. There have been many studies on intrusion detection in the literature. We can list some of these studies as follows. B. W. Masduki et al. [2], in their study, made attack detection using SVM. The success rate in R2L attacks was 96.08%. Ş. Sağıroğlu et al. [4] developed an ANN-based intelligent IDS in their study. The clever IDS they developed yielded very successful results. They tested the intelligent IDS using the KDD'99 dataset. They used 65536 samples from the KDD'99 dataset. The highest success rate they achieved was 97.92% and the lowest success rate was 81.93%. H. A. Sonawane et al. [8] proposed two methods in their study, namely Neural Networks (NN) and NN-based principal component analysis (PCA). The NN-based PCA method used a few features of the KDD'99 dataset, while the NN method used all the features of the KDD'99 dataset. By comparing these two methods, they observed that NN gave better results than PCA. The highest success rate of NN was 90.20%. M. Govindarajan et al. [9] performed attack detection using ensemble classifier (radial basis function (RBF)+SVM) in their study. The best success rate of RBF+DVM was 85.19%. A. Dastanpour et al. [10] performed attack detection using SVM, ANN and Genetic Algorithm (GA) algorithms in their study. When they applied the GA algorithm on SVM, they achieved 100% performance using 24 features of the KDD'99 dataset, while when they applied the GA algorithm on ANN, they achieved 100% performance on 18 features of the KDD'99 dataset. ANN+GA algorithm achieved better performance with fewer features. W. Wang et al. [11] performed attack detection using PCA in their study. The best success rate was 98.80%. S. Kumar et al. [12] used ANN to detect attacks in their study. The success rate on the KDD'99 dataset was 91.90%. They used 494021 samples for training and 311027 sample data for testing. J. Esmaily et al. [13] performed attack detection using DT and ANN in their study. In addition, DT and ANN algorithms were compared with each other. On the KDD'99 dataset; ANN gave better results with 99.71%. The success rate of the DT algorithm was 97.93%. Y. B. Bhavsar et al. [14] performed attack detection using DVM in their study. Normally, the success rate was 94.18%, while using 10-fold cross validation and RBF kernel, the success rate increased to 98.57%. G. Poojitha et al. [15] performed attack detection using ANN in their study. They used a total of 12723 samples from the KDD'99 dataset, 6363 samples for training and 6360 samples for testing. The success rate was 94.93%. B. Huyot et al. [5] performed online unsupervised attack detection on the Defense Advanced Research Projects Agency (DARPA) dataset by using copula theory or functions in their study. They achieved a success rate of 79% on the DARPA dataset. Although many [6], [7], [16]–[21] studies using copulas have been examined in the literature, no study on attack detection using copula functions has been found in the literature, except for this study. The contributions of this article have been given below.

- We propose novel and reliable classifiers (Student's-t, Gumbel, Clayton, Gaussian, Independent and Frank classifiers) for intrusion detection systems.
- We use the MrMR feature extraction technique for feature extraction.
- The proposed gaussian copula-based classifier outperformed state-of-the-art basic methods on the KDD'99 dataset with a success rate of 99.41%.
- We use the 10-fold cross-validation method to measure the performance of the proposed models.
- •We conclude that the copula-based classifier is a competitive alternative to several state-of-the-art methods.

2. Material and Method

2.1. Copula Functions

The term copula has been used in Latin to mean connection, relationship [7], [22]. The term copula was first proposed by Abe Sklar in 1959 [23]. Copulas have often been used to express (measure) the dependence between variables [5], [24]. The main purpose of copulas is to describe the interrelationship (dependency) of several random variables [25]. In addition, copulas are used to examine dependency structures among random variables and to obtain a multivariate distribution function [26]. Copula functions with their margins are used to construct the multivariate joint distribution function [6], [27]. In statistics science; copula are multivariate functions that provide the relationship between the common distribution functions of the random variable vector and the marginals of this distribution [22]. In other words; copulas are functions used to get a common distribution using marginal distributions [5], [22]. Copulas have been used in many application areas. At the beginning of these application areas are insurance [28], finance [29], statistics [30], economics [31], risk management [32] and security [3], [6], [33]. Copulas have an important place in applications because their elements in a class can be easily constructed, contain

large variables, and have good algebraic properties [34]. Copula functions have generally divided into two basic categories, elliptical and Archimedean copulas [35], [36].

$u = (u_1, u_2, u_3, \dots, u_n; \theta) \in [0,1]^n$ is the random variable vector and C is the joint distribution function. C has been shown in equation (1).

$$\mathrm{C}(\mathrm{u}_1, \mathrm{u}_2, \mathrm{u}_3, \dots, \mathrm{u}_\mathrm{n}; \theta) = \mathrm{P}(\mathrm{U}_1 \leq \mathrm{u}_1, \mathrm{U}_2 \leq \mathrm{u}_2, \mathrm{U}_3 \leq \mathrm{u}_3, \dots, \mathrm{U}_\mathrm{n} \leq \mathrm{u}_\mathrm{n}). \quad (1)$$

where θ is the copula parameter. Copulas can take a single parameter or more than one parameter according to their type. A two-dimensional (variable) copula function C, a continuous distribution function whose marginals are defined as $u = [0,1]$ and $C: [0,1]^2 \rightarrow [0,1]$ has the following properties.
$C(0,u) = C(u,0) = 0$ for $\forall_n \in [0,1]$
$C(1,u) = C(u,1) = u$ for $\forall_n \in [0,1]$
$C(v_1, v_2) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_2) \geq 0$ for each $(u_1, u_2), (v_1, v_2) \in [0,1] * [0,1]$ with $u_1 \leq v_1$ and $u_2 \leq v_2$ [5], [37][22] [3].

On the other hand, the Sklar theorem is one of the most important theorems of copulas [37]. The Sklar theorem mentions the function of copulas in linking multivariate joint distribution functions with its (univariate) marginal distribution functions [22], [38]. This theorem has made copulas more understandable and has easily expressed the relationship between copulas and joint distribution functions [39], [40]. The Sklar theorem has been given in equation (2).

$$C(u_1, u_2, \dots, u_d) = F\left({F_1}^{-1}(u_1), {F_2}^{-1}(u_2), \dots, {F_d}^{-1}(u_d)\right) \quad (2)$$

where the marginal distributions of random variables $x_1, x_2, \dots, x_d$ have expressed as $u_1 = F_1(x_1), u_2 = F_2(x_2), \dots, u_d = F_d(x_d)$, respectively [41], [42]. Although there are many types of copula families; Gumbel, Independent, Clayton, Gaussian, Student's-t, and Frank copula families, which have been widely used in the literature, have been used in this study [18], [22], [43]. The copula density function has often been used to estimate the parameter of a copula [16]. Let F be given as the joint distribution function. Let f be the density function of the joint distribution function F. Let C, be the copula function. $F_1, F_2, F_3, \dots, F_n$ are the marginal distributions of the F joint distribution function. The copula density function c has been alculated as in equation (3) [23], [44].

$$c(u_1, u_2, \dots, u_n) = \frac{\partial^n C(u_1, u_2, \dots, u_n)}{\partial u_1, \partial u_2, \dots, \partial u_n} = \frac{f({F_1}^{-1}(u_1), {F_2}^{-1}(u_2), \dots, {F_n}^{-1}(u_n))}{\prod_{i=0}^{n} f_i({F_i}^{-1}(u_i))} . \quad (3)$$

The relationship between the multivariate density function $f(u_1, u_2, \dots, u_n)$ and the copula density function has been shown in equation (4) [37], [45].

$$f(u_1, u_2, \dots, u_n) = c(F_1(u_1), F_2(u_2), \dots, F_n(u_n); \alpha) \prod_{i=1}^{n} f_i(u_i) . \quad (4)$$

where f_i is the univariate density function of the marginal distribution function F_i. c is the copula density function. α is the copula parameter [3]. The definitions of copulas families used in this study are given below. Gaussian (Normal) copula is obtained from multivariate Gaussian distribution [37], [46], [47] [3]. Gaussian copula is radially symmetric in its dependence structure [21], [44], [48]. Gaussian copula cannot model tail dependence because it has upper tail dependence ($\lambda_u = 0$) and lower tail dependence ($\lambda_l = 0$) [21]. The general representation of Gaussian copula is expressed as $C_\rho^{Gaussian}$. Student's-t copula is obtained from multivariate Student's distributions [46], [47]. The general representation of Student's-t copula is expressed as $C_{v,\rho}^{Student's-t}$. Student's-t copula is radially symmetric in its dependence structure [21], [44], [48]. Student'-t copula models both lower tail and upper tail dependency [21]. The Archimedean copula family is one of the most important copula families. Archimedean copulas are frequently preferred because they are easy to construct, can model various dependency properties of copula families, and are easy to use in applications [48], [50]. One of the advantages that makes Archimedean copulas useful is that they have a closed form [51], [52]. Archimedean copula families are generated using the formula in equation (5) [44], [53].

$$C_\theta(u_1, u_2, \dots, u_n) = \varphi^{-1}\left(\sum_{i=1}^{n} \varphi(u_i)\right) \quad (5)$$

where θ is the dependency parameter and φ is the generator function. In this study, Archimedean copulas (Clayton, Gumbel, Frank and Independent) were used for intrusion detection [3]. Naive Bayes theorem is a method for calculating (overcome) data uncertainty [54], [55]. Naive bayes is a classification technique frequently used in machine learning and data mining [19], [56]. Naive bayes classifier is based on bayes rule and probability theorem [57]. This

classifier is a supervised learning method used to predict the class from the features of the dataset [58], [59]. Copula-based classifiers often make use of the naive bayes theorem. The naive bayes theorem has been given in equation (6). The expression given in equation (6) has been used as a classification tool.

$$P(S_d|X) = \frac{P(S_d)P(X|S_d)}{P(X)} \quad (6)$$

Maximum A Posterior (MAP) is an estimation method used in naive bayes theorem [60], [61]. The MAP classifier can be designed by comparing the posterior probability $P(S_d|X)$. In other words, the MAP can be constructed by looking for the class that maximizes the posterior probability [62]. Also, MAP is used to select the best probability [63]. Let $X = (X_1, X_2, , X_3, \dots, X_K)$ features of an object belong to class D. When classifying this object, X features is assigned to the class with the posterior highest probability in class D [6]. For continuous features; A Gaussian copula function can be used to model the dependency structure in the probability function. In this case; probability can be calculated as in equation (7) by using the Gaussian copula density in the expression given in equation (6).

$$P(S_d|X) = \frac{\Phi(F_1(X_1),F_2(X_2),\dots,F_K(X_K)|S_d)P(S_d)\prod_{k=1}^{K} f_k(X_k|S_d)}{f(X)} \quad (7)$$

where F_i, is the marginal distribution functions and f_i is the marginal density of the features. Φ denotes Gaussian copula density [6], [7]. It can be generalized as in equation (8) by writing any (Student's-t, Clayton, Frank, Gumbel, Independent) copula density instead of Gaussian copula density in equation (7) [6], [7].

$$P(S_d|X) = \frac{\mathrm{c}(F_1(X_1),F_2(X_2),\dots,F_K(X_K)|S_d)P(S_d)\prod_{k=1}^{K} f_k(X_k|S_d)}{f(X)} \quad (8)$$

where c represents any copula density. The expression given in equation (8) is expressed as a naive bayesian classifier. The naive bayesian classifier can be obtained using the Independent copula. Therefore, the copula-based classifier is its generalized version. Based on the expression given in equation (8), a copula-based MAP classifier can be constructed as in equation (9) [6], [7].

$$\widehat{S_D} = \arg\max_{S_d \in Y} \mathrm{c}(F_1(X_1), F_2(X_2), \dots, F_K(X_K)|S_d) \prod_{k=1}^{K} f_k(x_k|S_d)P(S_d) \quad (9)$$

While constructing the MAP classifier, if the Inference Functions for Margins (IFM) method has been applied; any (Gaussian, Student's-t, Clayton, Frank, Gumbel, Independent, etc.) copula, the marginals estimated during the application of the IFM method, their densities and previous experimental probabilities have been used equation (9) to obtain a MAP classifier. In the case of applying the *Canonical Maximum Likelihood* (CML) method; any copula has been obtained by substituting the experimental cumulative distribution functions, experimental densities, and previous experimental probabilities in equation (9) [7] [3].

2.2. KDD'99 (KDD Cup 1999) Dataset

The first study sponsored by DARPA was carried out by the Massachusetts Institute of Technology (MIT) Lincoln laboratory in 1998 [64], [65]. DARPA is a data set used to perform both training/learning and testing [64]. KDD'99 dataset has been obtained by preprocessing the DARPA dataset (feature extraction etc.) [66], [67]. KDD'99 has been widely used in the research of IDSs in recent years [66], [68]. The purpose of the widespread use of the KDD'99 dataset is to facilitate training and testing for intrusion detection [4], [69]. A lot of preprocessing has been needed before the DARPA dataset can be used for IDSs. In this study; the KDD'99 dataset, which has obtained by preprocessing the DARPA dataset, has been used. By using the KDD'99 dataset, training and test results can be obtained faster [69]. There are 38 attack types in total, 24 attack types in the KDD'99 training dataset and 14 attack types in the test dataset [65], [69]. KDD'99 is a dataset consisting of 41 features, 9 basic and 32 derived [65]. The KDD'99 dataset is divided into four categories: basic features, content features, time-based traffic features, and host-based traffic features [65], [70]. In this study, two different KDD'99 datafiles, KDD100 (kddcup.data.gz) and KDD10 (kddcup.data_10_percent.gz) have been used [71]. KDD100 consists of 4898431 samples and KDD10 consists of 494021 samples [71], [72]. The quantities and categories of normal and attack types in the KDD10 and KDD100 datasets have been shown in Table 1.

Table 1. Quantities and categories of normal and attack types found in the KDD10 and KDD100 datasets [12], [3]

	KDD10		KDD100	
Attack Type	**Quantity**	**Quantity**	**Quantity**	**Category**
apache2	-	-	-	-
back	2203	DoS	2203	DoS
buffer_overflow	30	U2R	30	U2R
ftp_write	8	R2L	8	R2L
guess_passwd	53	R2L	53	R2L
httptunnel	-	-	-	-
imap	12	R2L	12	R2L
ipsweep	1247	probe	12481	probe
land	21	DoS	21	DoS
loadmodule	9	U2R	9	U2R
mailbomb	-	-	-	-
mscan	-	-	-	-
multihop	7	R2L	7	R2L
named	-	-	-	-
neptune	107201	DoS	1072017	DoS
nmap	231	probe	2316	probe
normal	97278	normal	972781	normal
perl	3	U2R	3	U2R
phf	4	R2L	4	R2L
pod	264	DoS	264	DoS
portsweep	1040	probe	10413	probe
processtable	-	-	-	-
ps	-	-	-	-
rootkit	10	U2R	10	U2R
saint	-	-	-	-
satan	1589	probe	15892	probe
sendmail	-	-	-	-
smurf	280790	DoS	2807886	DoS
snmpgetattack	-	-	-	-
snmpguess	-	-	-	-
spy	2	R2L	2	R2L
sqlattack	-	-	-	-
teardrop	979	DoS	979	DoS
udpstorm	-	-	-	-
warezclient	1020	R2L	1020	R2L
warezmaster	20	R2L	20	R2L
worm	-	-	-	-
xlock	-	-	-	-
xsnoop	-	-	-	-
xterm	-	-	-	-
Total	**494021**		**4898431**	

On the other hand, the data amounts of KDD10 and KDD100 data sets and the percentages of normal and attack (DoS, Probe, U2R, R2L) types have been shown in Table 2.

Table 2. The data amounts of the data types in the KDD'99 dataset and the percentage ratios of normal and attack (U2R, R2L, DoS, Probe) types [71], [3]

Dataset Name	**Data Amount**	**Attack Types**				**Normal**
		U2R	**R2L**	**DoS**	**Probe**	
KDD10	494021	%0.01	%0.22	%79.23	%0.83	%19.69
KDD100	4898431	%0.001	%0.02	%79.27	%0.83	%19.85

2.3. Preprocessing Stages of Datasets

In this study, two datasets, KDD10 and KDD100, which are under the KDD'99 dataset, have been used. These datasets have consisted of labeled data. Therefore, it is also clear to which attacks the data in the KDD10 and KDD100 datasets belong. The “protocol_type”, “service” and “flag” fields in these data sets have been in text (string) format, while the other fields have been in numerical format. In order to be able to operate on the data sets in our study, all the fields in the data sets must be in numerical format. Therefore, numerical values have been given to each of the “protocol_type”, “service” and “flag” fields in text format. In the "attack_type" field in the KDD10 and KDD100 data sets, a numerical value of 1 has been given for "normal traffic", while a numerical value of 2 has been given to other "all attack types". The purpose of doing this is to determine whether it is an attack rather than the type of attack. While converting the "protocol_type" names in the KDD10 and KDD100 datasets to digital format, icmp, tcp and udp protocols have been digitized 1, 2 and 3 respectively. The conversion of the "flag" names in the KDD10 and KDD100 datasets to numeric format has been shown in Table 3.

Table 3. Converting "flag" names in KDD10 and KDD100 datasets to numeric format [3]

Flag	Numerical Value
OTH	1
REJ	2
RSTO	3
RSTOS0	4
RSTR	5
S0	6
S1	7
S2	8
S3	9
SF	10
SH	11

The conversion of "service" names in the KDD10 and KDD 100 datasets to numeric format has been shown in Table 4.

Table 4. Converting "service" names in KDD10 and KDD100 datasets to numeric format [3]

KDD10 dataset		KDD100 dataset	
Service Name	**Numerical Value**	**Service Name**	**Numerical Value**
IRC	1	IRC	1
X11	2	X11	2
Z39_50	3	Z39_50	3
auth	4	aol	4
bgp	5	auth	5
courier	6	bgp	6
csnet_ns	7	courier	7
ctf	8	csnet_ns	8
daytime	9	ctf	9
discard	10	daytime	10
domain	11	discard	11
domain_u	12	domain	12
echo	13	domain_u	13
eco_i	14	echo	14
ecr_i	15	eco_i	15
efs	16	ecr_i	16
exec	17	efs	17
finger	18	exec	18
ftp	19	finger	19
ftp_data	20	ftp	20
gopher	21	ftp_data	21
hostnames	22	gopher	22
http	23	harvest	23
http_443	24	hostnames	24
imap4	25	http	25
iso_tsap	26	http_2784	26
klogin	27	http_443	27
kshell	28	http_8001	28
ldap	29	imap4	29
link	30	iso_tsap	30
login	31	klogin	31
mtp	32	kshell	32
name	33	ldap	33
netbios_dgm	34	link	34
netbios_ns	35	login	35
netbios_ssn	36	mtp	36
netstat	37	name	37
nnsp	38	netbios_dgm	38
nntp	39	netbios_ns	39
ntp_u	40	netbios_ssn	40
other	41	netstat	41
pm_dump	42	nnsp	42
pop_2	43	nntp	43
pop_3	44	ntp_u	44
printer	45	other	45
private	46	pm_dump	46
red_i	47	pop_2	47
remote_job	48	pop_3	48
rje	49	printer	49
shell	50	private	50
smtp	51	red_i	51
sql_net	52	remote_job	52

Table 4 (Continuous). Converting "service" names in KDD10 and KDD100 datasets to numeric format [3]

ssh	53	rje	53
sunrpc	54	shell	54
supdup	55	smtp	55
systat	56	sql_net	56
telnet	57	ssh	57
tftp_u	58	sunrpc	58
tim_i	59	supdup	59
time	60	systat	60
urh_i	61	telnet	61
urp_i	62	tftp_u	62
uucp	63	tim_i	63
uucp_path	64	time	64
vmnet	65	urh_i	65
whois	66	urp_i	66
		uucp	67
		uucp_path	68
		vmnet	69
		whois	70

2.4. Application of mRMR Method to KDD10 Data Set

mRMR is a method for selecting the most relevant features among the features in a dataset and reducing redundancy [73], [74]. mRMR is an entropy (disorder) based feature selection method. Entropy calculates the uncertainty in a random feature. Entropy produces values between 0-1. As seen in Table 5, the mRMR method has been applied on the KDD10 dataset, and the features in the dataset have been ranked according to their importance. According to these listed features, classification has been done using copula functions.

Table 5. Ranking of features according to importance when mRMR_miq criterion is applied to KDD10 dataset [3]

Raw Order		**After implemented mRMR**	
Feature Name	**Feature No**	**Feature Name**	**mRMR_miq**
duration	1	count	23
protocol_type	2	dst_bytes	6
service	3	duration	1
flag	4	dst_host_count	32
src_bytes	5	src_bytes	5
dst_bytes	6	srv_count	24
land	7	dst_host_srv_count	33
wrong_fragment	8	flag	4
urgent	9	service	3
hot	10	protocol_type	2
num_failed_logins	11	land	7
logged_in	12	wrong_fragment	8
num_compromised	13	urgent	9
root_shell	14	hot	10
su_attempted	15	num_failed_logins	11
num_root	16	logged_in	12
num_file_creations	17	num_compromised	13
num_shells	18	root_shell	14
num_access_files	19	su_attempted	15
num_outbound_cmds	20	num_root	16
is_host_login	21	num_file_creations	17
is_guest_login	22	num_shells	18
count	23	num_access_files	19
srv_count	24	num_outbound_cmds	20
serror_rate	25	is_host_login	21
srv_serror_rate	26	is_guest_login	22
rerror_rate	27	serror_rate	25
srv_rerror_rate	28	srv_serror_rate	26
same_srv_rate	29	rerror_rate	27
diff_srv_rate	30	srv_rerror_rate	28
srv_diff_host_rate	31	same_srv_rate	29
dst_host_count	32	diff_srv_rate	30
dst_host_srv_count	33	srv_diff_host_rate	31

Table 5 (Continuous). Ranking of features according to importance when mRMR_miq criterion is applied to KDD10 dataset [3]

dst_host_same_srv_rate	34	dst_host_same_srv_rate	34
dst_host_diff_srv_rate	35	dst_host_diff_srv_rate	35
dst_host_same_src_port_rate	36	dst_host_same_src_port_rate	36
dst_host_srv_diff_host_rate	37	dst_host_srv_diff_host_rate	37
dst_host_serror_rate	38	dst_host_serror_rate	38
dst_host_srv_serror_rate	39	dst_host_srv_serror_rate	39
dst_host_rerror_rate	40	dst_host_rerror_rate	40
dst_host_srv_rerror_rate	41	dst_host_srv_rerror_rate	41

2.5. Feature Selection

As can be seen in Table 5, after the features have been ranked in order of importance using the mRMR_miq feature selection criterion on the KDD10 dataset, feature selection has been started based on the first (23rd) feature. The first feature has been taken and the relationship status of the other features has been examined according to this feature. For example, the 1st feature (23rd) has been chosen at first. Then, the accuracy rate in the classification process has been i by taking the 1st feature and the 2nd feature (6th). Accuracy rates have been obtained using this method when each subsequent feature has been added. This situation has continued until the last feature (feature 41). The classification process has been completed by obtaining the three best performance rates and the features used for each dataset. In Figure 1 has been shown the relationship between the total elapsed time and the calculated percent while selecting the feature on the KDD10 dataset.

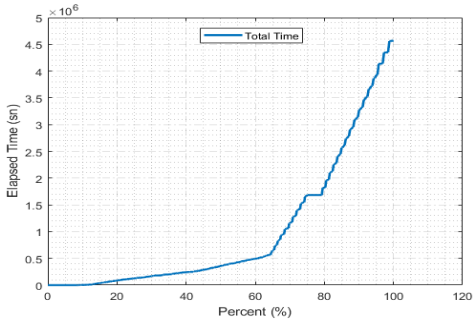


Figure 1. The relationship between the total calculated percentage depending on the total elapsed time while selecting the feature on the KDD10 dataset

As seen in Figure 1, the total time taken for all the features was while classifying on the KDD10 dataset has been calculated. During the calculation, as has stated above, a new feature has been added to the feature set each time, and the accuracy rates have been calculated. The amount of time required to calculate anything rises proportionally with the number of additional features provided. In particular, it has been observed that the calculated features spend much more time after the percentage rate is 80%. The calculation of the time elapsed between two features according to the use of IFM/CML methods with each copula family on the KDD10 dataset has been shown in Figure 2.

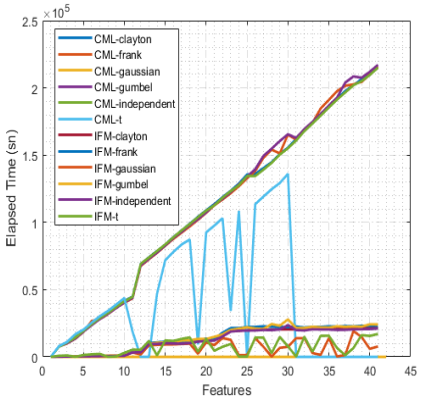


Figure 2. Calculation of the elapsed time between two features according to the use of IFM/CML methods with each copula family on the KDD10 dataset

As can be seen in Figure 2, the elapsed time between two features has been calculated at each step used while classifying the KDD10 dataset. While making the calculation, the accuracy rates have been calculated by adding the features separately as shown in Figure 2. The variation of features according to total elapsed time according to the use of IFM/CML methods with each copula family on the KDD10 data set has been shown in Figure 3.

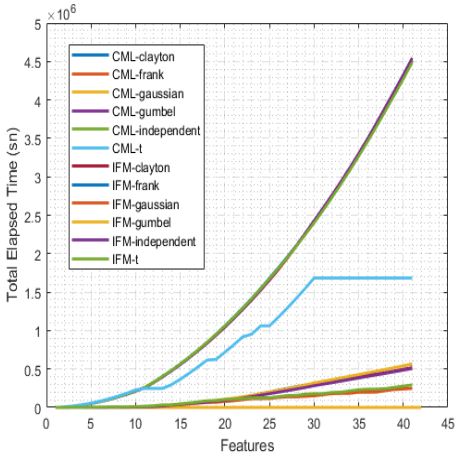


Figure 3. The variation of the characteristics according to the total elapsed time according to the use of IFM/CML methods with each copula family on the KDD10 dataset

As can be seen in Figure 3, while classifying on the KDD10 dataset, the total elapsed time for each copula family has been calculated by adding the time elapsed as a new feature has been added (to the previously calculated times). When all features have been used on the KDD10 dataset, the copula family that achieved the best performance has been the Gaussian copula using the IFM method. When all of the features were applied to the KDD10 dataset, the family of copulas known as the Gaussian copula utilizing the IFM approach produced the best results. The performance rates of the Gaussian copula family, which has shown the best performance for 41 features of the KDD10 dataset, according to the feature series ranked by mRMR have been shown in Figure 4.

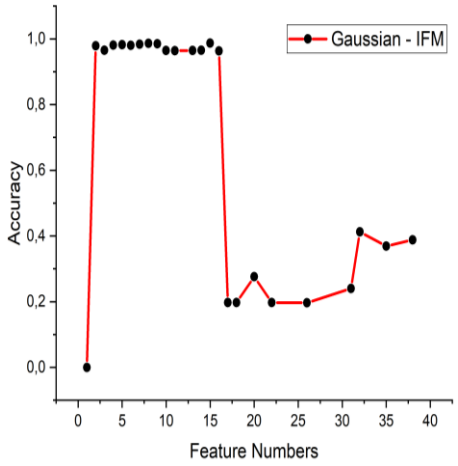


Figure 4. Performance ratios of the best performing Gaussian copula family for 41 features of the KDD10 dataset by feature series ranked by mRMR

As can be seen in Figure 4, the accuracy rates obtained depending on the number of Gaussian copula and IFM method features have been calculated on the KDD10 dataset. When taking into consideration the accuracy rates of the Gaussian copula, it has been shown that the best success rates have been reached between the second and the fifteenth features. This was discovered after taking into consideration the accuracy rates. This demonstrates to us that the first 15 features in the feature set, when applied with the IFM approach and the Gaussian copula, will produce relevant results when applied to the problem-solving process.

3. Results and Discussion

In this paper, intrusion detection has been made using copula-based classifiers and it has been investigated which copula has the best performance. The performance criterion used in this study has been given in equation (10) [3], [75].

$$Accuracy = \frac{True\ Positive\ (TP)+True\ Negative\ (TN)}{True\ Positive\ (TP)+False\ Negative\ (FN)+False\ Positive\ (FP)+True\ Negative\ (TN)} \quad (10)$$

In Figure 5 has been shown the application steps of copula-based classifiers to the KDD'99 dataset.

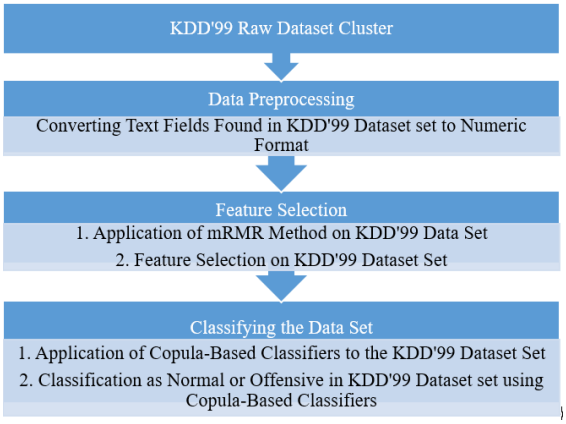


Figure 5. Application stages of copula-based classifiers to KDD'99 dataset

3.1. Application 1: KDD10

The KDD10 dataset was used in the first application because it contains less data but includes all samples. As a result, time was saved in feature selection. Gumbel, Independent, Clayton, Gaussian, Student's-t, and Frank copula families, as well as CML and IFM methods, were used to detect attacks in Application 1. 1%, 5%, 10%, and 50% of the KDD10 dataset were trained on a computer equipped with an Intel CPU (Xeon E5620), 16 GB RAM and Quadro K2000 GPU. Due to the large amount of data, 100% of the KDD10 dataset was trained on an HP-Z840 workstation with 2 10-cores Intel (Xeon E52687Wv3) processors, 64GB RAM and Quadro P5000 GPU. The trainings were held in the MATLAB environment. In the classification phase, the 10-fold cross-validation method was used. Using the confusion matrix found in Table 8 as the basis for the classifier evaluation metrics allowed for its acquisition. In the KDD10 dataset, certain percentages of each normal and attack type have been taken. 1% of the KDD10 dataset normally consists of 4940 data. But in this study, 4956 data have been used by taking 1% of each attack type. Purpose of this; it is to ensure that examples of all attack types are found while training the dataset. If the percentages of each attack type have been not taken at the same rate, low learning situations would occur in some attack types and excessive learning situations in others. This is true for 100%, 50%, 10% and 5% of the dataset. In Table 6, it has been demonstrated that the quantities of each attack type included in the KDD10 dataset correspond to the predetermined percentages (%) of the dataset.

Table 6. Quantities of each attack type found in the KDD10 dataset [3]

Attack Type	KDD10 (%100)	KDD10 (%50)	KDD10 (%10)	KDD10 (%5)	KDD10 (%1)
smurf	280790	140395	28079	14040	2808
neptune	107201	53601	10721	5361	1073
normal	97278	48639	9728	4864	973
back	2203	1102	221	111	23
satan	1589	795	159	80	16
ipsweep	1247	624	125	63	13
portsweep	1040	520	104	52	11
warezclient	1020	510	102	51	11
teardrop	979	490	98	49	10
pod	264	132	27	14	3
nmap	231	116	24	12	3
guess_passwd	53	27	6	3	1
buffer_overflow	30	15	3	2	1
land	21	11	3	2	1
warezmaster	20	10	2	1	1
imap	12	6	2	1	1
rootkit	10	5	1	1	1
loadmodule	9	5	1	1	1
ftp_write	8	4	1	1	1
multihop	7	4	1	1	1
phf	4	2	1	1	1
perl	3	2	1	1	1
spy	2	1	1	1	1
Total	**494021**	**247016**	**49411**	**24713**	**4956**

The amount of data has been calculated by taking the same percentages of each of the 23 attack types in the KDD10 dataset. For example; the data amount of the back attack type consists of 2203 samples in data set. Decimal numbers have been rounded. For example; While 1% of 2203, 22.03 samples should be taken from the back attack type, 23 samples have been taken due to rounding. The same is true for any other attack type. In Table 7, the success rates of the copula families with the best three performances have been shown by using 1% of KDD10 data set. Here, among the copulas that showed the same success, those who achieved this success at least have been considered more successful.

Table 7. Success rates of the best three performing copula families using the 1% rates of KDD10 dataset

Copula Family	Method	TP	TN	FP	FN	Accuracy (%)	Features Used
gumbel	**IFM**	**973**	**3920**	**63**	**0**	**98.73**	**“23 6 1 32 5 24 33 4 3 2”**
gumbel	IFM	973	3920	63	0	98.73	“23 6 1 32 5 24 33 4 3 2 7”
independent	IFM	973	3920	63	0	98.73	“23 6 1 32 5 24 33 4 3 2 7”
gaussian	IFM	973	3920	63	0	98.73	“23 6 1 32 5 24 33 4 3 2 10 12 22”
gaussian	IFM	973	3919	64	0	98.71	“23 6 1 32 5 24 33 4 3 2 10 12”
independent	IFM	973	3919	64	0	98.71	“23 6 1 32 5 24 33 4 3 2”
gaussian	IFM	973	3918	65	0	98.69	“23 6 1 32 5 24 33 4 3 2 10 12 22 29 31”
clayton	IFM	973	3918	65	0	98.69	“23 6 1 32 5 24 33 4 3 2 7”

As seen in Table 7, the best success rate has been obtained as 98.73% with Gumbel, Independent and Gaussian copulas using the IFM method. For the Gumbel copula family, this performance has been achieved using the “23 6 1 32 5 24 33 4 3 2” and “23 6 1 32 5 24 33 4 3 2 7” feature sets. Between these two feature sets, it should be preferred that shows the same performance with less features. While the Independent copula has achieved this success rate with the characteristics of “23 6 1 32 5 24 33 4 3 2 7”,

the Gaussian copula family has achieved this with the characteristics of "23 6 1 32 5 24 33 4 3 2 10 12 22". In this case, the Gumbel copula family has achieved the best performance by using fewer features than the Independent copula family and the Gaussian copula family. For 1% of the KDD10 dataset, Gumbel copula family, IFM method and "23 6 1 32 5 24 33 4 3 2" features should be preferred. In Table 8, the success rates of the copula families with the best three performances have been shown by using 5% of KDD10 dataset.

Table 8. The success rates of the best three performing copula families using 5% of KDD10 dataset

Copula Family	Method	TP	TN	FP	FN	Accuracy (%)	Features Used
gaussian	**IFM**	**4788**	**19707**	**142**	**76**	**99.12**	**"23 6"**
frank	**IFM**	**4788**	**19707**	**142**	**76**	**99.12**	**"23 6"**
independent	**IFM**	**4788**	**19707**	**142**	**76**	**99.12**	**"23 6"**
clayton	**IFM**	**4788**	**19707**	**142**	**76**	**99.12**	**"23 6"**
t	**IFM**	**4788**	**19707**	**142**	**76**	**99.12**	**"23 6"**
gumbel	**IFM**	**4788**	**19707**	**142**	**76**	**99.12**	**"23 6"**
independent	IFM	4853	19637	212	11	99.10	"23 6 1 32 5 24 33 4 3"
gumbel	IFM	4794	19692	157	70	99.08	"23 6 1"

As seen in Table 8, the best success rate with 99.12% has been obtained by using the IFM method with the characteristics of "23 6" for the Gaussian, Frank, Clayton, Gumbel, Independent and Student_t copula families. For 5% of the KDD10 data set, any of the Student_t, Gumbel, Clayton, Gaussian, Independent, and Frank copula families, IFM method and "23 6" features should be preferred. In Table 9, the success rates of the best three performing copula families have been shown by using 10% of KDD10 dataset.

Table 9. Success rates of the best three performing copula families using 10% of KDD10 dataset

Copula Family	Method	TP	TN	FP	FN	Accuracy (%)	Features Used
gaussian	**IFM**	**9585**	**39373**	**310**	**143**	**99.08**	**"23 6"**
frank	**IFM**	**9585**	**39373**	**310**	**143**	**99.08**	**"23 6"**
clayton	**IFM**	**9585**	**39373**	**310**	**143**	**99.08**	**"23 6"**
gumbel	**IFM**	**9585**	**39373**	**310**	**143**	**99.08**	**"23 6"**
independent	**IFM**	**9585**	**39373**	**310**	**143**	**99.08**	**"23 6"**
t	**IFM**	**9585**	**39373**	**310**	**143**	**99.08**	**"23 6"**
gaussian	IFM	9585	39373	310	143	99.08	"23 6 1"
frank	IFM	9585	39373	310	143	99.08	"23 6 1"
independent	IFM	9585	39373	310	143	99.08	"23 6 1"
clayton	IFM	9585	39373	310	143	99.08	"23 6 1"
gumbel	IFM	9585	39373	310	143	99.08	"23 6 1"
t	IFM	9584	39374	309	144	99.08	"23 6 1"
gumbel	IFM	9685	39260	423	43	99.06	"23 6 1 32 5 24 33 4 3"
independent	IFM	9617	39309	374	111	99.02	"23 6 1 32 5 24 33 4 3"

As seen in Table 9, the best success rate with 99.08% has been obtained by using the "23 6" and "23 6 1" feature sets for the Gumbel, Independent, Clayton, Gaussian, Student_t, and Frank copula families and the IFM method. For the Gaussian, Frank, Clayton, Gumbel, Independent and Student_t copula families, 23th and 6th features with a smaller number of features should be preferred. For 10% of the KDD10 dataset, any of the Gumbel, Independent, Clayton, Gaussian, Student_t, and Frank copula families, IFM method and "23 6" features should be preferred. In Table 10, the success rates of the best three performing copula families have been shown by using 50% of KDD10 dataset.

Table 10. Success rates of the best three performing copula families using 50% of KDD10 dataset

Copula Family	Method	TP	TN	FP	FN	Accuracy (%)	Features Used
gumbel	**CML**	**48174**	**196287**	**2090**	**465**	**98.97**	**"23 6 1 32 5 24 33 4 3 2 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 25 26 27 28 29 30"**
gumbel	CML	48250	196171	2206	389	98.95	"23 6 1 32 5 24 33 4 3 2 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 25 26 27 28"
gaussian	IFM	48128	196301	2076	511	98.95	"23 6 1 32 5 24 33 4 3 2 7 10 11 12"

As seen in Table 10, the Gumbel copula family has the best success rate with 98.97% by using "23 6 1 32 5 24 33 4 3 2 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 25 26 27 28 29 30" features and the CML method. The Gaussian copula family, on the other hand, has achieved a success rate of 98.95% using the IFM method with the characteristics of "23 6 1 32 5 24 33 4 3 2 7 10 11 12". For 50% of KDD10 dataset, Gumbel copula family, CML method and "23 6 1 32 5 24 33 4 3 2 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 25 26 27 28 29 30" features should be preferred. In Table 11, the success rates of the copula families with the best three performances have been shown by using the 100% rates of KDD10 dataset.

Table 11. Success rates of the best three performing copula families using 100% of KDD10 dataset

Copula Family	Method	TP	TN	FP	FN	Accuracy (%)	Features Used
gaussian	**IFM**	**96325**	**391193**	**5550**	**953**	**98.68**	**"23 6 1 32 5 24 33 4 3 2 7 9 10 11"**
gaussian	IFM	95710	391601	5142	1568	98.64	"23 6 1 32 5 24 33 4"
gaussian	IFM	96183	390323	6420	1095	98.48	"23 6 1 32 5 24 33 4 3"

As seen in Table 11, the Gaussian copula family has obtained the best success rate with 98.68% by using the "23 6 1 32 5 24 33 4 3 2 7 9 10 11" features and the IFM method. Gaussian copula family, IFM method and "23 6 1 32 5 24 33 4 3 2 7 9 10 11" features should be preferred for 100% of the KDD10 dataset.

3.2. Application 2: KDD100

In Application 2; Attack detection has been performed using Gaussian, Gumbel, Clayton, Student's-t, Independent and Frank copula families and CML and IFM methods. As has been seen in Table 12, the "23 6 1 32 5 24 33 4 3 2 7 9 10 11" feature set, the "23 6 1 32 5 24 33 4" feature set and the feature set "23 6 1 32 5 24 33 4 3" has been used in this study. These feature sets in the KDD100 dataset have been selected and classification has been carried out using the six copula families mentioned above. The KDD100 dataset has been trained on an HP-Z840 workstation with 10 cores, 2 x Intel CPUs (Xeon E52687Wv3), 64GB Ram and Quadro P5000 GPU. The trainings have been conducted in the MATLAB environment. In the classification phase, the 10-fold cross-validation method has been used. Since the degree of freedom (v) of the Student's-t copula is too large, an error has occurred while calculating the performance measurement. Therefore, the performance of the Student's-t copula has not shown in Table 13 and Table 14, Table 15. In Table 12 has been shown the numbers and names of the features that achieve the best performance on the KDD100 dataset.

Table 12. Numbers and names of features that achieve the best performance on the KDD100 dataset

Feature Number	Feature Name
23	count
6	dst_bytes
1	duration
32	dst_host_count
5	src_bytes
24	srv_count
33	dst_host_srv_count
4	flag

In Table 13, the performance rates of copula families on the KDD100 dataset have been shown by using the "23 6 1 32 5 24 33 4" features.

Table 13. Performance rates of copula families on KDD100 dataset

Copula Family	Method	TP	TN	FP	FN	Accuracy (%)	Features Used
independent	IFM	932928	3905436	20214	39853	98.77	"23 6 1 32 5 24 33 4"
gaussian	**IFM**	**972741**	**3896885**	**28765**	**40**	**99.41**	**"23 6 1 32 5 24 33 4"**
clayton	IFM	872190	3909171	16479	100591	97.61	"23 6 1 32 5 24 33 4"
frank	IFM	906564	3906204	19446	66217	98.25	"23 6 1 32 5 24 33 4"
gumbel	IFM	933022	3905382	20268	39759	98.77	"23 6 1 32 5 24 33 4"

As seen in Table 13, the Gaussian copula family has obtained the best success rate with 99.41% by using the "23 6 1 32 5 24 33 4" features and the IFM method. The worst success rate has been 97.61% using the "23 6 1 32 5 24 33 4" features and the IFM method for the Clayton copula family. Gaussian copula family and IFM method should be preferred for the "23 6 1 32 5 24 33 4" features in the KDD100 dataset. In Table 14, the performance rates of copula families on the KDD100 dataset have been shown by using the "23 6 1 32 5 24 33 4 3" features.

Table 14. Performance rates of copula families on KDD100 dataset using "23 6 1 32 5 24 33 4 3" features

Copula Family	Method	TP	TN	FP	FN	Accuracy (%)	Features Used
independent	IFM	933635	3905622	20028	39146	98.79	"23 6 1 32 5 24 33 4 3"
gaussian	**IFM**	**972676**	**3894142**	**31508**	**105**	**99.35**	**"23 6 1 32 5 24 33 4 3"**
clayton	IFM	831992	3912384	13266	140789	96.86	"23 6 1 32 5 24 33 4 3"
frank	IFM	906560	3906189	19461	66221	98.25	"23 6 1 32 5 24 33 4 3"
gumbel	IFM	933740	3905665	19985	39041	98.80	"23 6 1 32 5 24 33 4 3"

As seen in Table 14, the Gaussian copula family has obtained the best success rate with 99.35% by using the "23 6 1 32 5 24 33 4 3" features and the IFM method. The worst success rate has been 96.86% using the "23 6 1 32 5 24 33 4 3" features and the IFM method for the Clayton copula family. Gaussian copula family and IFM method should be preferred for the "23 6 1 32 5 24 33 4 3" features in the KDD100 dataset. In Table 15, the performance rates of copula families on the KDD100 dataset have been shown using the "23 6 1 32 5 24 33 4 3 2 7 9 10 11" features.

Table 15. Performance rates of copula families on KDD100 dataset

Copula Family	Method	TP	TN	FP	FN	Accuracy (%)	Features Used
independent	IFM	922035	3907600	18050	50746	98.60	"23 6 1 32 5 24 33 4 3 2 7 9 10 11"
gaussian	**IFM**	**972757**	**3895289**	**30361**	**24**	**99.38**	**"23 6 1 32 5 24 33 4 3 2 7 9 10 11"**
clayton	IFM	910658	3908739	16911	62123	98.39	"23 6 1 32 5 24 33 4 3 2 7 9 10 11"
frank	IFM	658964	3909941	15709	313817	93.27	"23 6 1 32 5 24 33 4 3 2 7 9 10 11"
gumbel	IFM	926372	3907381	18269	46409	98.68	"23 6 1 32 5 24 33 4 3 2 7 9 10 11"

As seen in Table 15, the Gaussian copula family has obtained the best success rate with 99.38% by using the "23 6 1 32 5 24 33 4 3 2 7 9 10 11" features and the IFM method. The worst success rate has been 93.27%, and Frank copula family has been obtained by using "23 6 1 32 5 24 33 4" features and IFM method. Gaussian copula family and IFM method should be preferred for the "23 6 1 32 5 24 33 4 3 2 7 9 10 11" features in the KDD100 dataset.

4. Conclusion and Suggestions

In this study, attack detection was performed using copula-based classifiers. In Table 16 was shown the performance comparison of copula-based classifiers for different dataset amounts

Table 16. Performance comparison of copula-based classifiers for different dataset amounts

Best Copula Algorithm		Data Set Used	Accuracy (%)
Gumbel	IFM	KDD10 (1%)	98.73
Independent Clayton Gumbel Student's-t	IFM	KDD10 (5%)	99.12
Gaussian Frank Independent Gaussian Frank Clayton Gumbel Student's-t	IFM	KDD10 (10%)	99.08
Gumbel	CML	KDD10 (50%)	98.97
Gaussian	IFM	KDD10 (100%)	98.68
Gaussian	**IFM**	**KDD100**	**99.41**

As can be seen in Table 16, attack detection was made using copula-based classifiers according to various versions of the dataset. When the amount of data set is small, the Gumbel copula-based classifier achieves better performance, while the Gaussian copula-based classifier comes to the fore as the amount of data increases. In Table 17, the success rates of some studies on IDSs in the literature and the performance of the copula-based classifiers used in this study were compared.

Table 17. Comparing the proposed study's performance to earlier IDS studies published in the literature

Some Studies in the Literature	Method Used	Data Set Used	Accuracy (%)
A.Dastanpour et al.[10]	GA+ANN	KDD'99	100.00
J.Esmaily et al. [13]	ANN	KDD'99	99.71
This study	**Copula**	**KDD'99**	**99.41**
W.Wang et al. [11]	PCA	DARPA	98.80
Y.B.Bhavsar et al. [14]	SVM	NSL-KDD	98.57
Ş.Sağıroğlu et al. [4]	ANN	KDD'99	97.92
B.W.Masduki et al. [2]	SVM	KDD'99	96.08
G.Poojitha et al. [15]	ANN	KDD'99	94.93
S.Kumar et al. [12]	ANN	KDD'99	91.90
H.A.Sonawane et al. [8]	NN	KDD'99	90.20
M.Govindarajan et al. [9]	RBF+SVM	NSL-KDD	85.19
B.Huyot et al. [5]	Copula	DARPA	79.00

As can be seen in Table 17, attack detection was carried out using many different methods in IDSs. A.Dastanpour et al. [10] achieved 100% success by using 18 features of KDD'99 data set in their study. J.Esmaily et al. [13] achieved 99.71% success by using all of 41 features in their study. In this study, as seen in Table 17, 99.41% success was achieved by using fewer (8) features then others. When the results obtained from copula-based classifiers are compared with previous studies, quite remarkable results were obtained. Thus, it was shown that copula-based classifiers can be an alternative to machine learning classifiers. As a result; In this study, Independent, Clayton, Frank, Gaussian, Gumbel and Student's-t copula-based classifiers have been preferred, and the usability of these copula-based classifiers in intrusion detection systems were investigated. Classification was performed on KDD10 (10%) and KDD100 (full) datasets of KDD'99 using copula-based classifiers. The 10-fold cross-validation method has been used in the classification phase. While all copula classifiers achieved a good 99.12% performance on the KDD10 dataset, the Gaussian copula-based classifier achieved the best success rate of 99.41% on the KDD100 dataset. As can be seen in Table 17, copula-based classifiers achieved good values when compared to other methods.

In future studies, in addition to these copula families, attack detection performances will be examined by using different copula families. In addition, the usability of ANN and copula-based approaches will be investigated.

Contributions of the authors

Mehmet Burukanlı: Methodology, Conceptualisation, Validation, Data curation, Writing original draft. Musa Çıbuk: Methodology, Conceptualisation, Validation, Investigation, Supervision.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] M. Burukanlı, Ü. Budak, and M. Çıbuk, "Saldırı Tespit Sistemlerinde Makine Öğrenme Metotlarının Kullanımı," in *Uluslararası Bilim ve Mühendislik Sempozyumu*, 2019, pp. 1052–1057.

[2] B. W. Masduki, K. Ramli, F. A. Saputra, and D. Sugiarto, "Study on Implementation of Machine Learning Methods Combination for Improving Attacks Detection Accuracy on Intrusion Detection System (IDS)," in *2015 International Conference on Quality in Research (QiR)*, 2015, pp. 56–64.

[3] M. Burukanlı, "Copula fonksiyonlarını kullanarak bilgisayar ağlarında saldırı tespiti," M.S. thesis, Bitlis Eren Üniversitesi Lisansüstü Eğitim Enstitüsü, Bitlis, Turkey, 2020.

[4] Ş. Sağıroğlu, E. N. Yolaçan, and U. Yavanoğlu, "Zeki Saldırı Tespit Sistemi Tasarımı ve Gerçekleştirilmesi," *Journal of Faculty of Engineering and Architecture of Gazi University*, vol. 26, no. 2, pp. 325–340, 2011.

[5] B. Huyot, Y. Mabiala, and J.-F. Marcotorchino, "Online Unsupervised Anomaly Detection in Large Information Systems Using Copula Theory," in *2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems*, Nov. 2014, pp. 679–684, doi: 10.1109/CCIS.2014.7175820.

[6] R. Salinas-Gutiérrez, A. Hernández-Aguirre, M. J. J. Rivera-Meraz, and E. R. Villa-Diharce, "Using Gaussian Copulas in Supervised Probabilistic Classification," in *Soft Computing for Intelligent Control and Mobile Robotics*, C. Castillo, J. Kacprzyk, and W. Pedrycz, Eds., Springer-Verlag Berlin and Heidelberg GmbH & Co. KG, 2010, pp. 355–372.

[7] M. Scavnicky, "A study of Applying Copulas in Data Mining," M.S. thesis, Charles University in Prague Faculty of Mathematics and Physics, Prague, Czech Republic, 2013.

[8] H. A. Sonawane and T. M. Pattewar, "A Comparative Performance Evaluation of Intrusion Detection Based on Neural Network and PCA," in *2015 International Conference on Communications and Signal Processing (ICCSP)*, 2015, pp. 841–845, doi: 10.1109/ICCSP.2015.7322612.

[9] M. Govindarajan and R. M. Chandrasekaran, "Intrusion Detection using an Ensemble of Classification Methods," in *Lecture Notes in Engineering and Computer Science*, 2012, vol. 1, pp. 459–464.

[10] A. Dastanpour, S. Ibrahim, R. Mashinchi, and A. Selamat, "Comparison of Genetic Algorithm Optimization on Artificial Neural Network and Support Vector Machine in Intrusion Detection System," in *2014 IEEE Conference on Open Systems (ICOS)*, Oct. 2014, pp. 72–77, doi: 10.1109/ICOS.2014.7042412.

[11] W. Wang and R. Battiti, "Identifying Intrusions in Computer Networks with Principal Component Analysis," in *First International Conference on Availability, Reliability and Security (ARES'06)*, 2006, pp. 270–279, doi: 10.1109/ARES.2006.73.

[12] S. Kumar and A. Yadav, "Increasing Performance Of Intrusion Detection System Using Neural Network," in *2014 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, 2014, pp. 546–550.

[13] J. Esmaily, R. Moradinezhad, and J. Ghasemi, "Intrusion Detection System Based on Multi-Layer Perceptron Neural Networks and Decision Tree," in *2015 7th Conference on Information and Knowledge Technology (IKT)*, May 2015, pp. 1–5, doi: 10.1109/IKT.2015.7288736.

[14] Y. B. Bhavsar and K. C. Waghmare, "Intrusion Detection System using Data Mining Technique: Support Vector Machine," *International Journal of Emerging Technologies and Advanced Engineering*, vol. 3, no. 3, pp. 581–586, 2013.

[15] G. Poojitha, K. N. Kumar, and P. J. Reddy, "Intrusion Detection using Artificial Neural Network," in *2010 Second International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Jul. 2010, pp. 1–7, doi: 10.1109/ICCCNT.2010.5592568.

[16] S. Sathe, "A Novel Bayesian Classifier using Copula Functions," *arXiv Preprint cs/0611150*, 2006.

[17] D. Qian et al., "Drowsiness Detection by Bayesian-Copula Discriminant Classifier Based on EEG Signals during Daytime Short Nap," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 743–754, 2017, doi: 10.1109/TBME.2016.2574812.

[18] L. Slechan and J. Górecki, "On the Accuracy of Copula-Based Bayesian Classifiers: An Experimental Comparison with Neural Networks," in *Computational Collective Intelligence*, M. Nunez, N. T. Nguyen, D. Camacho, and B. Trawinski, Eds., Springer International, Madrid, 2015, pp. 485–493.

[19] Y. Chen, "A Copula-Based Supervised Learning Classification for Continuous and Discrete Data," *Journal of Data Science*, vol. 13, pp. 769–790, 2014.

[20] N. Hammami, M. Bedda, and N. Farah, "Probabilistic Classification Based on Gaussian Copula for Speech Recognition: Application to Spoken Arabic Digits," in *Signal Processing - Algorithms, Architectures, Arrangements, and Applications Conference Proceedings (SPA)*, 2013, pp. 312–317.

[21] Y. He, J. Deng, and H. Li, "Short-Term Power Load Forecasting with Deep Belief Network and Copula Models," in *2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Aug. 2017, vol. 1, pp. 191–194, doi: 10.1109/IHMSC.2017.50.

[22] R. B. Nelsen, *An Introduction to Copulas*. Springer Science+Business Media, Inc., 2006.

[23] J. Lu, W. Tian, and P. Zhang, "The Archimedean Copulas Measure of the Risk Characteristic for the Tail Dependent Asset Returns," in *2008 International Conference on Management Science and Engineering 15th Annual Conference Proceedings*, Sep. 2008, pp. 173–181, doi: 10.1109/ICMSE.2008.4668912.

[24] P. Embrechts, F. Lindskog, and A. McNeil, "Modelling Dependence with Copulas and Applications to Risk Management," in *Handbook of Heavy Tailed Distributions in Finance*, S. T. Rachev, Ed., Elsevier, Amsterdam, 2003, pp. 329–384.

[25] T. Schmidt, "Coping with Copulas," in *Copulas: From Theory to Application in Finance*, J. Rank, Ed., Risk Books Publishing, Berkeley, 2006, pp. 3–34.

[26] E. Bouyé, V. Durrleman, A. Nikeghbali, G. Riboulet, and T. Roncalli, *Copulas for Finance-A Reading Guide and Some Applications*, SSRN Electronic Journal, 2000.

[27] A. Surana and A. Pinto, "Analysis of Stochastic Automata Networks Using Copula Functions," in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2010, pp. 1699–1706, doi: 10.1109/ALLERTON.2010.5707121.

[28] B. Z. Karagül, "Hayat Dışı Sigortalarda Doğrusal Olmayan Bağımlılığın Kopulalar ile Dinamik Finansal Analizi," M.S. thesis, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, Turkey, 2013.

[29] G. Yapakçı, "Kopulalar Teorisinin Finansta Uygulaması," M.S. thesis, Ege Üniversitesi Fen Bilimleri Enstitüsü, İzmir, Turkey, 2007.

[30] S. Aslan, S. Çelebioğlu, and F. Öztürk, "İki Boyutlu Arşimedyen Kopulalarda İstatistiksel Sonuç Çıkarımı ve Bir Uygulama," *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, vol. 14, no. 2, pp. 1–18, 2012.

[31] L. Andersen and J. Sidenius, "Extensions to the Gaussian Copula: Random Recovery and Random Factor Loadings," *Journal of Credit Risk*, vol. 1, no. 1, pp. 29–70, 2005, doi: 10.21314/jcr.2005.003.

[32] D. Çatal and R. S. Albayrak, "Riske Maruz Değer Hesabında Karışım Kopula Kullanımı: Dolar-Euro Portföyü," *Yaşar Üniversitesi E-Dergi*, vol. 8, no. 31, pp. 5187–5202, 2013.

[33] M. Mehdizadeh, R. Ghazi, and M. Ghayeni, "Power System Security Assessment with High Wind Penetration Using the Farms Models Based on Their Correlation," *IET Renewable Power Generation*, vol. 12, no. 8, pp. 893–900, 2018, doi: 10.1049/iet-rpg.2017.0386.

[34] P. Hájek and R. Mesiar, "On Copulas, Quasicopulas and Fuzzy Logic," *Soft Computing*, vol. 12, no. 12, pp. 1239–1243, 2008, doi: 10.1007/s00500-008-0286-z.

[35] J. Yan, "Enjoy the Joy of Copulas: With a Package copula," *Journal of Statistical Software*, vol. 21, no. 4, pp. 1–21, 2007, doi: 10.18637/jss.v021.i04.

[36] H. He and P. K. Varshney, "A Coalitional Game for Distributed Inference in Sensor Networks with Dependent Observations," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1854–1866, 2016, doi: 10.1109/TSP.2015.2508781.

[37] G. Van Der Wulp, "Using Copulas in Risk Management," M.S. thesis, Tilburg University Department of Econometrics, Tilburg, Netherlands, 2003.

[38] M. Sklar, "Fonctions de Répartition à n Dimensions et Leurs Marges," *Publications de l'Institut Statistique de l'Université de Paris*, vol. 8, pp. 229–231, 1959.

[39] A. Alhan, "Bağımsızlık Kapulasını İçeren Kapula Aileleri, Kapula Tahmin Yöntemleri ve İstanbul Menkul Kıymetler Borsasında Sektörler Arası Bağımlılık Yapısı," Ph.D. dissertation, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara, Turkey, 2008.

[40] A. M. Karataş, "Modeling of Daily Maximum and Minimum Temperature Changes in Bitlis Province Using Copula Method," *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, vol. 7, no. 2, pp. 268–275, 2018.

[41] J. Lu, W.-J. Tian, and P. Zhang, "The Extreme Value Copulas Analysis of the Risk Dependence for the Foreign Exchange Data," in *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCom)*, Oct. 2008, pp. 1–6, doi: 10.1109/WiCom.2008.2405.

[42] P. Mou, F. Tao, C. Jia, and W. Ma, "A Copula-Based Function Model in Fuzzy Reliability Analysis on the Planetary Steering Gear," in *2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE)*, Jul. 2013, pp. 375–378, doi: 10.1109/QR2MSE.2013.6625605.

[43] A. M. Karakaş, "Modelling Temperature Measurement Data by Using Copula Functions," *Bitlis Eren Üniversitesi Fen Bilimleri ve Teknoloji Dergisi*, vol. 7, no. 1, pp. 27–32, 2017.

[44] S. Jadhav and R. Daruwala, "3-D Modeling of Statistical Dependencies Using Copulas for Wireless Sensor Network," in *Proceedings of the 2016 IEEE International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2016, pp. 1886–1889, doi: 10.1109/WiSPNET.2016.7566469.

[45] C. D. Tran, O. O. Rudovic, and V. Pavlovic, "Unsupervised Domain Adaptation with Copula Models," in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2017, pp. 1–6.

[46] C. Romano, "Calibrating and Simulating Copula Functions: An Application to the Italian Stock Market," *Risk Management Functional Capital*, vol. 180, pp. 1–26, 2002.

[47] Anonim, "Probability Distributions," 2017. [Online]. Available: http://www.nematrian.com/Pages/ProbabilityDistributionsCombined.pdf. [Accessed: Apr. 13, 2020].

[48] E. E. Sezgin, "Finansal Bağımlılık Analizi: Vine ve CD Vine Copula Yaklaşımları," M.S. thesis, Bitlis Eren Üniversitesi ve Fırat Üniversitesi Fen Bilimleri Enstitüsü, Bitlis, Turkey, 2019.

[49] S. Arslan, "Arşimedyen Kapulalar Üzerine Bir Çalışma," Ph.D. dissertation, Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara, Turkey, 2013.

[50] S. Çelebioğlu, "Arşimedyen Kapulalar ve Bir Uygulama," *Selçuk Üniversitesi Fen Fakültesi Fen Dergisi*, vol. 22, no. 1, pp. 43–52, 2003.

[51] S. S. Galiani, "Copula Functions and Their Application in Pricing and Risk Managing Multiname Credit Derivative Products," M.S. thesis, Department of Mathematics, King's College London, London, UK, 2003.

[52] H. Manner, *Estimation and Model Selection of Copulas with an Application to Exchange Rates*. Maastricht: Maastricht Research School of Economics of Technology and Organizations (METEOR) Press, 2007.

[53] Y. Dong, S. Zhang, G. Fan, L. Zhang, L. Yi, and M. Lin, "Application of Copula Function in the Reliability Analysis of the Electrical System and the Power Device of Certain-Type Armored Vehicle," in *CSAE 2012 - Proceedings of the 2012 IEEE International Conference on Computer Science and Automation Engineering*, 2012, vol. 1, pp. 386–389, doi: 10.1109/CSAE.2012.6272621.

[54] A. Setiawan, Soeheri, E. Panggabean, M. A. Elhias, F. Ikorasaki, and B. Riski, "Efficiency of Bayes Theorem in Detecting Early Symptoms of Avian Diseases," in *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, Aug. 2018, pp. 1–5, doi: 10.1109/CITSM.2018.8674273.

[55] N. S. B. Sembiring, E. Ginting, M. Fauzi, Yudi, F. Tambunan, and E. V. Haryanto, "An Expert System to Diagnose Herpes Zoster Disease Using Bayes Theorem," in *2019 7th International Conference on Cyber and IT Service Management (CITSM)*, Nov. 2019, pp. 1–3, doi: 10.1109/CITSM47753.2019.8965381.

[56] A. H. Jahromi and M. Taheri, "A Non-Parametric Mixture of Gaussian Naive Bayes Classifiers Based on Local Independent Features," in *2017 Artificial Intelligence and Signal Processing Conference (AISP)*, Oct. 2017, vol. 2018, pp. 209–212, doi: 10.1109/AISP.2017.8324083.

[57] K. P. Murphy, "Naive Bayes Classifiers," 2006. [Online]. Available: https://www.ic.unicamp.br/~rocha/teaching/2011s1/mc906/aulas/naive-bayes.pdf. [Accessed: Apr. 02, 2020].

[58] K. Netti and Y. Radhika, "A Novel Method for Minimizing Loss of Accuracy in Naive Bayes Classifier," in *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Dec. 2015, pp. 1–4, doi: 10.1109/ICCIC.2015.7435801.

[59] F.-J. Yang, "An Implementation of Naive Bayes Classifier," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, Dec. 2018, pp. 301–306, doi: 10.1109/CSCI46756.2018.00065.

[60] Anonim, "Maximum A Posteriori Estimation," 2020. [Online]. Available: https://en.wikipedia.org/wiki/Maximum_a_posteriori_estimation. [Accessed: May 04, 2020].

[61] H. Bogunovic, J. M. Pozo, R. Cardenes, L. S. Roman, and A. F. Frangi, "Anatomical Labeling of the Circle of Willis Using Maximum A Posteriori Probability Estimation," *IEEE Transactions on Medical Imaging*, vol. 32, no. 9, pp. 1587–1599, 2013.

[62] F. Peng, D. Schuurmans, and S. Wang, "Augmenting Naive Bayes Classifiers with Statistical Language Models," *Information Retrieval Boston*, vol. 7, no. 3–4, pp. 317–345, 2004.

[63] M. J. Islam, Q. M. J. Wu, M. Ahmadi, and M. A. Sid Ahmed, "Investigating the Performance of Naive Bayes Classifiers and K-Nearest Neighbor Classifiers," in *2007 International Conference on Convergence Information Technology (ICCIT)*, Nov. 2007, pp. 1541–1546, doi: 10.1109/ICCIT.2007.148.

[64] J.-H. Lee, J.-H. Lee, S.-G. Sohn, J.-H. Ryu, and T.-M. Chung, "Effective Value of Decision Tree with KDD 99 Intrusion Detection Datasets for Intrusion Detection System," in *2008 10th International Conference on Advanced Communication Technology (ICACT)*, Feb. 2008, vol. 2, pp. 1170–1175, doi: 10.1109/ICACT.2008.4493974.

[65] Anonim, "Kddcup1999," 1999. [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/task.html. [Accessed: Apr. 09, 2020].

[66] B. Nethu, "Classification of Intrusion Detection Dataset Using Machine Learning Approaches," *International Journal of Electronics and Computer Science Engineering*, vol. 1, no. 3, pp. 1044–1051, 2012.

[67] D. H. Deshmukh, T. Ghorpade, and P. Padiya, "Intrusion Detection System by Improved Preprocessing Methods and Naive Bayes Classifier Using NSL-KDD 99 Dataset," in *2014 International Conference on Electronics and Communication Systems (ICECS)*, Feb. 2014, pp. 1–7, doi: 10.1109/ECS.2014.6892542.

[68] T. Janarthanan and S. Zargari, "Feature Selection in UNSW-NB15 and KDDCUP'99 Datasets," in *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, 2017, pp. 1881–1886.

[69] G. Meena and R. R. Choudhary, "A Review Paper on IDS Classification Using KDD 99 and NSL KDD Dataset in WEKA," in *2017 International Conference on Computer, Communications and Electronics (Comptelix)*, Jul. 2017, pp. 553–558, doi: 10.1109/COMPTELIX.2017.8004032.

[70] Y. Vural, "Kurumsal Bilgi Güvenliğinde Güvenlik Testleri ve Öneriler," *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, vol. 26, no. 1, pp. 89–103, 2011.

[71] Anonim, "The UCI KDD Archive Information and Computer Science University of California, Irvine," 1999. [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. [Accessed: Jan. 26, 2021].

[72] M. Burukanlı, M. Çıbuk, and Ü. Budak, "Saldırı Tespiti için Makine Öğrenme Yöntemlerinin Karşılaştırmalı Analizi," *BEÜ Fen Bilimleri Dergisi*, vol. 10, no. 2, pp. 613–624, 2021.

[73] H. Peng, F. Long, and C. Ding, "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005, doi: 10.1109/TPAMI.2005.159.

[74] M. Çıbuk, U. Budak, Y. Guo, M. C. Ince, and A. Sengur, "Efficient Deep Features Selections and Classification for Flower Species Recognition," *Measurement*, vol. 137, pp. 7–13, Apr. 2019, doi: 10.1016/j.measurement.2019.01.041.

[75] S. Wang, Y.-H. Zhang, J. Lu, W. Cui, J. Hu, and Y.-D. Cai, "Analysis and Identification of Aptamer-Compound Interactions with a Maximum Relevance Minimum Redundancy and Nearest Neighbor Algorithm," *BioMed Research International*, vol. 2016, pp. 1–12, 2016. [Online]. Available: http://downloads.hindawi.com/journals/bmri/2016/8351204.pdf. [Accessed: Apr. 29, 2020].