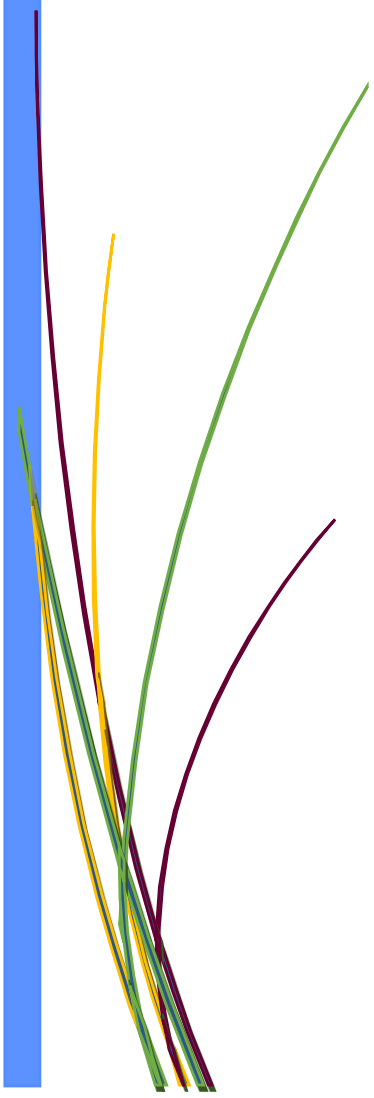
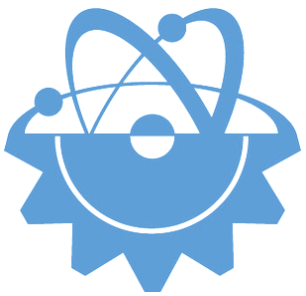
ISSN-Printed: 2536-5010
ISSN-Online: 2536-5134

Volume 14, No 1, 2024

EJT

EUROPEAN JOURNAL OF TECHNIC





INESEG

Email (for orders and customer services enquiries): info@ineseg.org, ejt@ineseg.org
Visit our home page on www.ineseg.org

EUROPEAN JOURNAL OF TECHNIQUE (EJT)
ISSN-Printed: 2536-5010
ISSN-Online: 2536-5134

Scope: European Journal of Technique (EJT) established in 2010. It is a peer –reviewed international journal to be of interest and use to all those concerned with research in various fields of, or closely related to, Engineering disciplines. European Journal of Technique (EJT) aims to provide a highly readable and valuable addition to the literature which will serve as an indispensable reference tool for years to come. The coverage of the journal includes all new theoretical and experimental findings in the fields of Engineering or any closely related fields. The journal also encourages the submission of critical review articles covering advances in recent research of such fields as well as technical notes.

The scopes include:

- Mechanical Engineering
- Textile Engineering
- Electrical-Electronics Engineering
- Computer and Informatics Engineering
- Civil and Architecture Engineering
- Mining Engineering
- Chemical Engineering
- Metallurgical and Materials Engineering
- Environmental Engineering
- Food Engineering
- Geological Engineering
- Industrial Engineering
- Renewable Energy

EDITORIAL BOARD MEMBERS

Editor-in-Chief

- Musa YILMAZ

Publisher Of Journal

- Heybet KILIÇ

ETHICS and POLICIES

European Journal of Technique (EJT) is committed to following the Code of Conduct and Best Practice Guidelines of COPE (Committee on Publication Ethics). It is a duty of our editors to follow Cope Guidance for Editors and our peer-reviewers must follow COPE Ethical Guidelines for Peer Reviewers We expect all prospective authors to read and understand our Ethics Policy before submitting any manuscripts to our journals.

Please note that submitted manuscripts may be subject to checks using the iThenticate service, in conjunction with CrossCheck, in order to detect instances of overlapping and similar text.

The iThenticate software checks submissions against millions of published research papers, documents on the web, and other relevant sources. If plagiarism or misconduct is found, consequences are detailed in the policy.

The chief goal of our policy is threefold: to provide advice for our authors, to maintain the scholarly integrity of our journals and their content, and to detail the ethical responsibilities of EJT, our editors and authors.

We expect all authors to read and understand our ethics policy before submitting to any of our journals. This is in accordance with our commitment to the prevention of ethical misconduct, which we recognise to be a growing problem in academic and professional publications. It is important to note that most incidents of plagiarism, redundant publication, copyright infringement or similar occur because of a lack of understanding, and not through fraudulent intent. Our policy is one of prevention and not persecution.

If you have any questions, please contact the relevant editorial office, or European Journal of Technique (EJT)' ethics representative: ejtineseg@gmail.com

Download a PDF version of the Ethics and Policies [PDF,392KB].

Authors' Responsibilities

Authors should:

- Ensure that all researched work submitted is original, fully referenced and that all authors are represented accurately. The submission must be exclusive and not under consideration elsewhere.
- Provide accurate contact details for a designated corresponding author, who shall be deemed by the publisher and editor as fully responsible for the authorship of the paper and all communications concerning the ethical status and originality of the paper. This includes any queries or investigations that may arise, pre- or post publication.
- Openly disclose the source of all data and third party material, including previously unpublished work by the authors themselves. Anything that could compromise the originality of the submission should be expressly avoided and/or discussed with the editorial office in the first instance.
- Identify any third party material that they intend to include in their article, and obtain written permission for re-use in each instance from the relevant copyright holders. Such permissions should be submitted once the manuscript is accepted, or requires small changes to be accepted. For further guidance on seeking permission to use 3rd party material please see the Rights and Permissions section.
- Openly disclose any conflict of interest - for example, if publication were to benefit a company or services in which the author(s) has a vested interest.

- Expect to formally agree publication terms which defines the author and the publishers rights for the work. Visit our website for further information.
- Expect the editor to scan submissions using plagiarism detection software at iThenticate to check a paper's originality before sending out for review.
- Fully correspond and comply with the editor and publisher in any requests for source data, proof of authorship or originality in a timely manner, providing reasonable explanation for discrepancies or failures to disclose vital information.
- Fully co-operate with any consequent investigations if the editor and/or publisher are dissatisfied with the evidence available or the explanations provided.
- Expect transparency, efficiency and respect from the publisher and the editor during the submissions process.
- Remain in good communication with both the publisher and the editor.
- When necessary, submit corrigenda in a timely and responsible fashion.
- Co-operate fully with the publication of errata and with the retraction of articles found to be unethical, misleading or damaging.
- Remain in good communication with the editor(s), the publisher and any co-authors.

Editors' Responsibilities

Editors should:

- Read and understand COPE guidelines as well as EJT's ethics policy, and follow them during all editorial processes.
- Protect the reputation of their journal(s) and published work by only publishing content of the highest quality and relevance in a timely and responsible manner.
- Carry out thorough, objective and confidential peer review for original article submissions that pass the initial quality check and editorial assessment, in adherence with COPE guidelines and EJT' ethics policy.
- Detail and justify any article types which will not be peer reviewed (e.g. editorials, opinion pieces etc.).
- Provide a transparent review and publication process as far as is possible, with full respect and care paid to the author(s).
- Provide advice and give reasonable explanation and updates to authors during the submissions process and once a decision has been made.
- Allow authors the right to appeal any editorial decision.
- Only accept papers based on the original merit, quality and relevance of their content.
- Support authors in queries concerning the originality of their submissions and request the support of EJT if necessary.
- Advise the publisher of any third party material which has been included for which they do not believe sufficient permission has been cleared.

- Be ready and prepared to publish corrections, corrigenda, errata when necessary, as well as retract articles that (the editor and EJT) deem unethical, misleading or damaging.
- Remain in good communication with both the publisher and the author(s).

Reviewers' Responsibilities

Reviewers should:

- Adhere to EJT's policy of confidential peer review of their journals. This includes, but is not restricted to, keeping their identity hidden from authors and not externally distributing any work that is passed to them for their eyes only.
- Only accept invitations to review work that is relevant to their own expertise and speciality.
- Review submitted work in a responsible, impartial and timely manner.
- Report any suspected ethical misconduct as part of a thorough and honest review of the work.
- Avoid the use of unnecessarily inflammatory or offensive language in their appraisal of the work.
- Accept the commitment to review future versions of the work and provide 'follow up' advice to the editor(s), if requested.
- Seek advice from the editor if anything is unclear at the time of invitation.
- Remain in good communication with both the publisher and the editor.

EJT's Responsibilities

EJT will:

- Protect the reputation of our journals and published work by only publishing content of the highest quality and relevance in a timely and responsible manner.
- Provide detailed information concerning both our understanding of publication ethics and our implementation of the same. Emphasise a desire for prevention, not eventual detection, of ethical misconduct.
- Uphold our COPE membership (or of such similar organisations) and keep our editorial offices, publishing staff and society partners up-to-date with their guidelines and policies, adapting our own where appropriate (and publicising any update).
- When necessary, request proof of originality/accuracy from the corresponding author of any work submitted to any of our journals.
- Use plagiarism detection software when necessary for any submission to any journal at any stage of the submissions and publication process.
- Provide a transparent submissions and publication process, with full respect and care paid to the author. This includes detailed and dedicated instructions to authors for each journal, outlining referencing style, accepted article types and submission processes.
- Investigate thoroughly any suggestion of ethical misconduct detected during any stage of the submissions process. This can include, but is not restricted to, the following: plagiarism, redundant publication, fabrication or misuse of data and authorial disputes.

- When necessary, retract articles that we deem to be unethical, misleading or damaging.
- When necessary, publish errata, corrigenda and retractions in a timely and responsible fashion, detailing the decision online in an open access format and publishing in print as soon as possible.
- Remain in good communication with editors, authors, reviewers and society partners (where applicable).

Further reading

- Authorship of the paper: Authorship should be limited to those who have made a significant contribution to the conception, design, execution, or interpretation of the reported study.
- Originality and plagiarism: The authors should ensure that they have written entirely original works, and if the authors have used the work and/or words of others that this has been appropriately cited or quoted.
- Data access and retention: Authors may be asked to provide the raw data in connection with a paper for editorial review, and should be prepared to provide public access to such data.
- Multiple, redundant or concurrent publication: An author should not in general publish manuscripts describing essentially the same research in more than one journal or primary publication. EJT do not view the following uses of a work as prior publication: publication in the form of an abstract; publication as an academic thesis; publication as an electronic preprint. Information on prior publication is included within each EJT and its journal Guideline for Authors.
- Acknowledgement of sources: Proper acknowledgment.
- Disclosure and conflicts of interest: All submissions must include disclosure of all relationships that could be viewed as presenting a potential conflict of interest.
- Fundamental errors in published works: When an author discovers a significant error or inaccuracy in his/her own published work, it is the author's obligation to promptly notify the journal editor or publisher and cooperate with the editor to retract or correct the paper.
- Reporting standards: Authors of reports of original research should present an accurate account of the work performed as well as an objective discussion of its significance.
- Hazards and human or animal subjects: Statements of compliance are required if the work involves chemicals, procedures or equipment that have any unusual hazards inherent in their use, or if it involves the use of animal or human subjects.
- Use of patient images or case details: Studies on patients or volunteers require ethics committee approval and informed consent, which should be documented in the paper.

EJT has also accessed and learned from the existing policies of other publishers and leading experts as well as open access articles that detail and define ethical misconduct.

- 'Plagiarism and the law', Joss Saunders, Learned Publishing, 23:279-202: http://www.ingentaconnect.com/content/alpsp/lp/2010/00000023/00000004/art00002
- iThenticate Plagiarism Resources: http://www.ithenticate.com/resources/6-consequences-of-plagiarism

Publisher of Journal

Heybey Kilic Dicle University, Turkey / TU Delft, Netherland

Vol 14, Number 1, 2024
European Journal of Technique (EJT)
ISSN: 2536-5134

CONTENTS

European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.14, No.1, 2024

Research Article

Text Detection and Recognition in Natural Scenes by Mobile Robot

Erdal Alimovski[1*], Gökhan Erdemir[2] and Ahmet Emin Kuzucuoğlu[3]

[1*]Istanbul Sabahattin Zaim University, Computer Engineering Department, 34303, Istanbul, Turkiye. (e-mail: erdal.alimovski@izu.edu.tr).
[2] University of Tennessee at Chattanooga, Dep. of Engineering Management and Tech., Chattanooga, TN 37405, USA. (e-mail: gokhan-erdemir@utc.edu).
[3] Marmara University, Department of Electrical and Computer Engineering, 34722, Istanbul, Turkiye. (e-mail: kuzucuoglu@marmara.edu.tr).

ARTICLE INFO

Received: Dec., 19. 2023
Revised: Apr., 17. 2024
Accepted: Apr, 20. 2024

Keywords:
Natural scene
Mobile robot
Optical character recognition
Text detection
Text recognition

Corresponding author: *Gökhan Erdemir*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1407231

ABSTRACT

Detecting and identifying signboards on their route is crucial for all autonomous and semi-autonomous vehicles, such as delivery robots, UAVs, UGVs, etc. If autonomous systems interact more with their environments, they have the ability to improve their operational aspects. Extracting and comprehending textual information embedded in urban areas has recently grown in importance and popularity, especially for autonomous vehicles. Text detection and recognition in urban areas (e.g., store names and street nameplates, signs) is challenging due to the natural environment factors such as lighting, obstructions, weather conditions, and shooting angles, as well as large variability in scene characteristics in terms of text size, color, and background type. In this study, we proposed a three-stage text detection and recognition approach for outdoor applications of autonomous and semi-autonomous mobile robots. The first step of the proposed approach is to detect the text in urban areas using the "Efficient And Accurate Scene Text Detector (EAST)" algorithm. Easy, Tesseract, and Keras Optical Character Recognition (OCR) algorithms were applied to the detected text to perform a comparative analysis of character recognition methods. As the last step, we used the Sequence Matcher to the recognized text values to improve the method's impact on OCR algorithms in urban areas. Experiments were conducted on the university campus by an 8-wheeled mobile robot, and a video stream process was carried out through the camera mounted on the top of the mobile robot. The results demonstrate that the Efficient And Accurate Scene Text Detector (EAST) text detection algorithm combined with Keras OCR outperforms other algorithms and reaches an accuracy of 91.6%

1. INTRODUCTION

In recent years, Computer Vision (CV) technology has made remarkable progress in practice and has a wide range of applications across many different fields, such as robotics [1], automotive[2], healthcare[3], agriculture [4], etc. Mobile robots, which are increasingly used daily, have started to be used frequently, especially in work such as delivery in urban areas. Therefore, they can be operated in different places, like streets, homes, hospitals, and factories. As a result of the interaction of mobile robots with their working environments and the need to sense some environmental features or landmarks, CV-based algorithms have become common for indoor and outdoor applications. Environmental knowledge can help mobile robots position themselves, alter their movement, and make decisions without human intervention [5]. Therefore, extracting rich semantic information such as textual information, objects and structures, terrain features, and landmarks from the environment is necessary. For example, if the mobile robot operates in urban areas, it can benefit from textual information about street names, shop names, and signboards. In this way, it may navigate itself according to the obtained data.

Text detection and recognition in urban areas (e.g., streets, squares, university campuses) has drawn much interest and is essential in various computer vision-based applications. Performing detection and recognition of text in urban areas is difficult because of the variety of backgrounds, low resolution, font types, distortion, occlusions, etc. [6]. Text detection, a prerequisite of text recognition, is a critical phase in textual information extraction and understanding. The fundamental text detection component is designing features that differentiate text from the backgrounds. In traditional methods, features are manually intended to acquire the characteristics of scene text [7], [8], whereas features are obtained entirely from training data in deep learning-based techniques [9], [10]. Generally, traditional methods comprise sliding window-based and connected components-based methods. To detect text, sliding window-based techniques change a window at each position in an image [11]. Character candidates are initially extracted as a first step in connected components-based approaches, and then post-processing is done to remove non-text noise and connect

the candidate [12]. In [13], [14], authors performed (HOG) and Random [14] Ferns to identify the characters. Following through pictorial structure, a particular word's optimal configuration was found. Mishra et al. utilized sliding windows and a Conditional Random Field model with a combination of bottom-up and top-down cues to identify character candidates [15]. A part-based tree-structured model was created in the study [14] to identify the characters in cropped photos. Authors in [16] proposed a combination of multi-scale mid-level features called Strokelets as an alternative for character representation. However, the efficiency of traditional methods is constrained in problematic conditions where low resolution, multi-orientation, and perspective distortion exist.

Recently, convolutional neural networks (CNNs) have been widely applied to text detection tasks to overcome the limitations of traditional methods. The Rotation Region Proposal Network (RRPN) was introduced, offering new Faster-RCNN components [17]. The proposed framework aims to detect arbitrary-oriented text in various environments.

The framework is designed to produce inclined proposals with text orientation angle data, which later is utilized for bounding box regression. Therefore, to set the arbitrary orientation proposals, the RoI pooling layer is introduced [18]. In [19], authors proposed a model that employs a Fully Convolutional Network (FCN) and one-step Non-Maximum Suppression (NMS) structure for scene text detection. FCN model consists of three main parts: feature extraction, feature fusion, and multi-task learning. Localizing the scene text's quadrilateral boundaries is adequate due to direct regression. Results from experiments have demonstrated the efficiency of the proposed model. In another study [19], authors proposed a novel model to detect arbitrary-oriented texts called Rotation Region CNN, based on Faster-RCNN. First, the Region Proposal Network (RPN) generates axis-aligned bounding boxes. Axis-aligned bounding boxes are refined and inclined minimum area boxes are predicted using pooled features. Finally, to obtain the detection results, NMS is performed.

OCRs have demonstrated impressive performance in various commercial applications over the past few decades, focusing on natural scene texts [20]. The ability of OCR technology to recognize documents with a constant background color, the most basic fonts, and nicely aligned text is impressive. However, the performance in scene text recognition, such as bills, traffic signs, and shop names, is limited due to the complex backgrounds, distinct and distorted fonts, uneven illumination, and color variations [21]. Consequently, scene text recognition has grown in popularity as a field for CV applications in robots. Character-based and full word-based recognition are the two main categories of conventional text recognition methods used in urban areas. For scene text recognition, many research efforts have been carried out. For example, authors in [22] proposed detecting and recognizing scene text frameworks for multi-oriented texts. Forest classifiers demonstrated tasks of text detection and recognition. Comprehensive experiments show that the suggested algorithm performs better than the existing approaches. In [23], a model was presented to perform recognition of scene texts under different orientations, such as vertical texts, top-to-bottom and vice versa, horizontally stacked vertical texts. The proposed model comprises three main processes: text localization, segmentation, and recognition. For localization and segmentation, excluding the pre-processing and post-processing steps, they used the maximally stable extremal regions detector while recognizing the Tesseract algorithm was performed. The results show the proposed model's effectiveness for the vertical text recognition task. In natural scenes, Ebin Zacharias et al. mounted a camera in a vehicle to detect and recognize [24]. Gamma correction, skew correction, and canny edge detector were performed to get the region of the text in an image. Further, Tesseract 5 with Long-Short-Term Memory (LSTM) was utilized during the recognition phase. This applied approach achieved around 83% correct character recognition rate.

In this study, we propose an approach that automatically recognizes the text information in signboards on the robot's working areas during mobile robot movement. The EAST algorithm was performed to detect the text values. Once the textual pattern was detected, the detected patterns were transmitted to Easy, Keras, and Tesseract OCR algorithms to compare character recognition methods in urban areas. In addition, we implemented a word similarity calculator method called the Sequence Matcher (SM) to repair wrongly recognized words and to examine its effect on OCR algorithms in urban areas.

2. MATERIALS AND METHODS

The proposed approach to detect and recognize the textual patterns in the urban areas during the mobile robot's movement consists of three phases: data collection, mobile robot control, and applied algorithms. Data collection was performed on campus with an 8-wheeled mobile robot controlled remotely. After investigating the existing literature, the EAST algorithm was used for text detection. As a first step, EAST detected the textual patterns in the live video. After that, it was transmitted to Easy OCR, Keras OCR, and Tesseract OCR for text recognition. Thus, the performance of the open-source OCR algorithms in urban areas on a live video stream was compared and evaluated by comprehensive experiments in this study. Figure 1 shows the study flowchart.

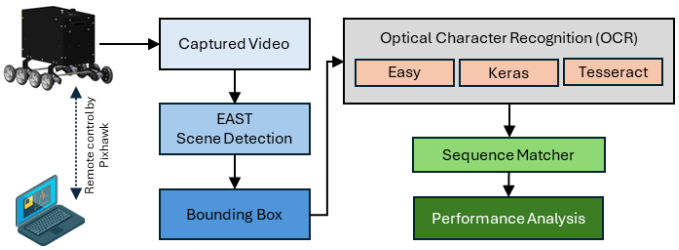


Figure 1. Flowchart of the proposed approach.

2.1. Data Collection

As mentioned in the introduction, more research is needed to recognize text in urban areas, such as streets, squares, etc., using open-source OCR algorithms. Therefore, this research captured live videos on campus utilizing a Zed2i camera [25] mounted at the top of an 8-wheeled mobile robot, where details will be clarified in the next section. In the conducted scenario, start and finish points were defined for the mobile robot. Thus, data were gathered in a specific area on campus where signboards, shops, and stationery are often used.

In addition, the fact that the signboard texts in that region were easily recognizable by the human eye played an important role in choosing that region. Thus, in the testing phase, we could compare the recognized text by the algorithm with the

actual text. Figure 2 presents the satellite image of the region, which contains the locations of each signboard on the test location. Also, the following path of the mobile robot is presented in the same figure. Pixhawk controller[26] was used for trajectory tracking and control of the robot. The existing text on the test area signboards is indicated in Table I.

Figure 2. Area of data collection phase.

TABLE I
LABELS IN SIGNBOARDS

Signboard Nr	Signboard Name	Existing Label
1	Sabri Ülker Araştırma Merkezi	"Sabri", "Ülker", "Araştırma ", "Merkezi"
2	KIRTASİYE STATIONARY	“KIRTASİYE”, “STATIONARY”
3	TDV Book Store Kitabevi	"TDV","Book", "Store","Kitabevi"
4	Yesen Burger	"Yesen", "Burger"

As seen in Figure 3, each signboard contains different textual patterns with fonts, backgrounds, colors, and board types. For example, some shop names' characters have various colors and font types. All tests were performed under limited lighting conditions during cloudy weather. During experiments, the camera's resolution was 3840x1080 pixels, with 30 FPS, while the total video record duration was around 35 minutes.

Figure 3. Signboards samples in the test area.

2.2. Mobile Robot Platform

The 8-wheel drive mobile robot designed in [27] was used during this study's data collection and experiments. It has eight wheels that can climb the pavements. The robot's dimensions are listed as total weight: 20 kg, width: 648 mm, length: 640 mm, height: 571 mm, and tire diameter: 140 mm [27]. In the standard configuration of the robot, the camera was mounted on the top of the robot. For the default setup of the robot, the camera's position was out of angle concerning target views for detection. This situation could hurt the applied text detection and recognition algorithms. The platform was designed and mounted on the top of the mobile robot to bring the camera to the same angle as the target views. Thus, the camera was set 1.13 m above the ground after the modifications. The mobile robot, after the modification, is presented in Figure 4.

Figure 4. Mobile robot with vision system.

2.3. Camera and Controller

The Zed2i camera was used in this study to gather visual data from the working environment. Pixhawk controller was used to control and track the robot's trajectory. The Zed2i camera and Pixhawk controller have often been used in various robotic applications. While the Zed2i camera can obtain 3D images and depth information due to its stereo imaging technology, the Pixhawk controller provides significant convenience in autonomous driving in robotic systems such as unmanned aerial vehicles (UAV) and unmanned ground vehicles (UGV). Stereolabs developed a ZED2i camera for image acquisition in various applications such as virtual reality, augmented reality, robotics, industrial automation, and many others. It is suitable not only for indoor but also for outdoor applications. The camera has a robust and reliable IR sensor that can provide high-quality images even in low-lighting conditions. The ZED2i captures both RGB and depth images in high resolution (2K) with passive stereoscopic 3D technology based on a composite stereo image of the camera. The ZED2i camera is compatible with NVIDIA Jetson platforms and other ARM-based systems. This makes it appropriate for various applications, including mobile robots, UAVs, self-driving vehicles, and other intelligent devices.

Pixhawk is an open-source autonomous vehicle control system. It can provide a wide range of applications for robotic systems when integrated with ArduPilot software. Pixhawk comprises various modules, such as GPS, high-speed processors, and configurable interfaces. It can accomplish the

navigation and trajectory control tasks crucial for mobile robots. To get the actual position of the mobile robot and perform trajectory tracking reliably, the u-blox NEO M8N GPS Module was mounted to the Pixhawk board. The GPS module starts searching for satellites and establishes a connection after a few minutes of being connected to the Pixhawk. Then, it sends the received location data packages to the board. Thus, the remote-control station can wirelessly receive the location data thanks to 433 MHz Telemetry Radio modules. Mission Planner (MP) software was installed on the remote-control station. Telemetry modules established the connection between the Pixhawk board and the remote PC. Hence, the robot was controlled and tracked by MP.

2.4. Scene Text Detection

Scene detection algorithms use image processing and artificial intelligence techniques to detect text accurately and effectively in complex scenes. These algorithms are used in many applications to detect and recognize, such as signage, logos, traffic signs, and security systems. Towards the literature review of scene detection algorithms, the EAST algorithm used in this study surpasses other algorithms in obtaining fast and accurate results. Therefore, it was decided to be used.

Scene text detection techniques have been accomplished on various benchmarks. These techniques, especially those that employ deep neural network models, have limitations when dealing with complex scenes. Interactions between the modules in the algorithmic model impact the text detection models' performance. Thus, text detection performance can be improved by utilizing a basic model that optimizes the loss function. Therefore, using a simple and effective EAST algorithm, text regions can be detected more rapidly and accurately [28]. The EAST algorithm was proposed by Zhou in 2017 to surpass the limitations in scene detection. The EAST algorithm employs a single neural network to predict the text's words or lines by avoiding candidate aggregation and word segmentation.

The model can eventually predict the 1-channel score map and the 4-channel box map if box data are tagged as RBOX. The algorithm can ultimately estimate the 1-channel score map and the 8-channel BOX map if the box data are annotated as QUAD.

2.5. Character Recognition

As a first step, the EAST algorithm detected the text regions and taken into a bounding box. Then, the detected pattern was transmitted to three open-source OCR algorithms. These are Easy OCR, Keras OCR, and Tesseract OCR. In this study, the performances of these character recognition algorithms were compared.

2.5.1. Easy OCR

Easy OCR is a cutting-edge OCR library that performs efficiently and supports over 80 languages. It was developed using Python and PyTorch frameworks. It executes detection using the CRAFT [29] method, a scene text detection model based on neural networks.

As mentioned previously, the EAST algorithm was performed for scene detection in this study. Therefore, both EAST and CRAFT algorithms were used for text detection before feeding the detected texts to the recognition phase. For recognition, Easy OCR uses a Convolutional Recurrent Neural Network (CRNN) [30], which is comprised of three parts: ResNet for feature extraction [31], LSTM sequence labeling [32], and Connectionist Temporal Classification (CTC) decoding [33]. The Easy OCR engine is among the best because of the pre-processing procedures included in this pipeline.

This OCR engine can identify and detect more than 80 languages. We employed the algorithm in Turkish configuration since our dataset contains Turkish texts. When the text is detected and taken into the bounding box, we feed it to the OCR engine to recognize the texts in the environment.

2.5.2. Keras OCR

Keras OCR is a deep learning-based OCR method in the Keras library [34]. Keras OCR offers a convenient interface for developing OCR models, which can identify text in various formats, including handwritten, printed, and even noisy or damaged text. Keras OCR has many typical applications, such as digitizing article texts, scanning, processing financial documents, and detecting texts in traffic signs, streets, and shops. Keras OCR is composed of two neural network architectures: CRAFT and CRNN. By investigating each character region and the affinities between characters, Keras OCR employs CRAFT to detect the text areas. Whereas for the text recognition phase, the original CRNN model was used. Due to the high accuracy of EAST text detection and some noisy detections of CRAFT during the implementation, we decided to disable the detection part of Keras OCR and utilize it just for character recognition.

2.5.3. Tesseract OCR

Tesseract is a well-known open-source OCR engine that Hewlett-Packard first developed and later supported by Google. Tesseract has been adapted for more than 140 different languages [35]. Since the 4.0 version, a new engine built on Long Short-Term Memory (LSTM) was developed. Compared to prior versions of Tesseract, LSTM, a particular type of RNN, offers significantly improved accuracy. Additionally, the "pytesseract" Python library exists and offers quick access to this engine in Python. The image was initially transformed into a binary image using an adaptive threshold. Following, character outlines were extracted by applying connected component analysis. Next, to organize the outlines into words, methods for character split and character association are performed. The two-password recognition process is ultimately carried out using clustering and classification approaches. To make its final determination regarding the recognized word, Tesseract consults both the language and user-defined dictionaries. Thus, the word with the smallest distance is given as an output.

2.6. Sequence Matcher

Sequence Matcher (SM)[36] is a class of ‘difflib’ modules used to compare the similarity of two given strings. The Ratcliff/Obershelp algorithm [37] is run in the background. After comparing the two strings, the algorithm returns a score between 0 and 1. After comparing two strings, if the obtained

score is greater than 0.7, it will be regarded as a keyword and stored as an actual word or label. The equation of the algorithm is as follows:

$$D_{ro} = \frac{2 * K_m}{|S1||S2|} \quad (1)$$

where K_m represents the number of the same characters in sequence, whereas $|S1|$ and $|S2|$ give the corresponding length for each of these two strings.

3. EXPERIMENTAL STUDY

In this study, once the video data was gathered with the mobile robot, the experiments were carried out on the computer with an Intel i5 Central Processing Unit (CPU), 16 GB random access memory (RAM), and a single Graphics Processing Unit (GPU) NVIDIA GeForce GTX 1680.

The EAST algorithm was first performed during the experiments to detect the target scene texts. Thus, textual patterns in signboards, shop boards, and others have been taken into a bounding box. Secondly, the detected textual pattern was inputted into Easy OCR, Keras OCR, and Tesseract OCR algorithms for the character recognition step. To evaluate the performance of the OCR algorithms, we define the accuracy metric, the number of correctly recognized words divided by the total number of existing words multiplied by 100\%. In addition, an analysis was conducted regarding how many times the existing words were recognized correctly by each algorithm and recognition time. Besides, the effect of the SM algorithm over OCR algorithms was investigated.

Figure 5. A sample of textual pattern detection by the EAST algorithm in cloudy weather.

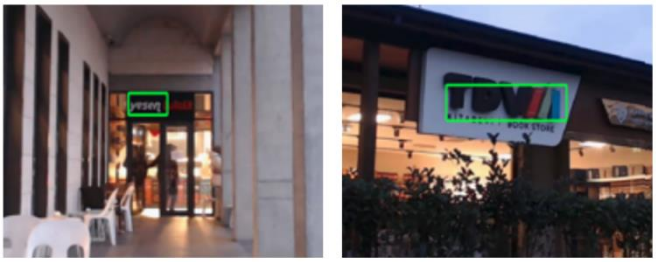


Figure 6. A sample of textual pattern detection by the EAST algorithm under limited lighting conditions.

In most cases, the EAST algorithm accurately detected the textual patterns in signboards, except in cases where the camera's distance was far away from the possible target texts. Its performance in cloudy weather and under limited lighting conditions is highly accurate, improving OCR algorithms' performance and correctness. In Figure 5, several boards with text that are accurately detected are presented. During the experiments, it was observed that when the angle of the mobile robot's camera was on the side position, the EAST algorithm was limited in detection. Therefore, as shown in Figure 6, the alignments of the bounding box vary.

When Table II is examined, it is clear that the Keras OCR algorithm outperformed other algorithms in accuracy. In addition, the average recognition time of the Keras OCR algorithm was the shortest compared to the other OCR algorithms. The algorithm recognized all the labels in the scene except the label "araştırma" even in bad conditions, such as low illumination and blurriness on the video stream caused by the vibrations of the robot. In contrast, the ability of the Easy OCR and Tesseract OCR to recognize the labels in urban areas was limited. Thus, both algorithms could not recognize several labels such as "TDV" and "Burger".

TABLE II
PERFORMANCE ANALYSIS OF OCR ALGORITHMS.

Method name	Average Recognition Time (second)	Accuracy (%)
Easy OCR	0.06	83.3
Keras OCR	0.04	91.6
Tesseract OCR	0.08	75

On the other hand, we measured the performance of the algorithms in terms of the number of repeats of each label and the recognition time of each label, as listed in Table III. Notice that recognition time means the time passed between text detection and recognition.

Table III demonstrates that the Keras OCR algorithm recognized each label at least two or three times more than Easy OCR and Tesseract OCR. Besides this, recognition time is less than other algorithms for each label. Some labels, such as *Sabri*, *Kırtasiye*, and *Yesen*, were recognized many times, while labels such as *Stationary*, *TDV*, *Store*, and *Kitabevi* were less frequently identified. In some cases where the mobile robot was moved on a smooth ground surface, the vibration was lesser. Thus, the video quality was good, so the algorithms performed better due to the camera position alignment and type of surfaces. On the other hand, as the mobile robot was moved across the grass, the vibrations increased, and the video quality became blurry, making it harder for the algorithms to recognize the labels.

During the experiments, it was observed that character recognition algorithms misidentified some labels by missing one or two characters of the actual labels. For example, algorithms recognized the actual labels that are *sabri*, *ülker*, *merkezi*, *Book*, and *Store* as ***abri***, **ülke**, ***merkez***, ***boo***, and ***tore***, respectively. It is caused by the blurriness of the video stream and the camera's position. It was worth noting that, during the movement of the mobile robot, some vibrations occurred due to the ground type, affecting the quality of the video stream.

We applied the SM method to address the shortcomings in the previous step, which calculates the similarity between two words. The SM supported the output of each OCR algorithm, and it figured out the word similarity between the recognized label and actual labels in a text file, which was previously stated. Thus, it suggested the most appropriate words. In our case, we set a threshold of 70%, so it only suggested labels higher than that percentage. Table IV demonstrates the performance results of each algorithm after applying the word similarity calculation method. Compared with Table III, it was

observed that each algorithm's recognition performance increased for all the labels.

TABLE III
PERFORMANCE COMPARISON OF OCR ALGORITHMS

	Easy OCR		Keras OCR		Tesseract OCR	
Recognized Labels	Rec. Number	Average Time (s)	Rec. Number	Average Time (s)	Rec. Number	Average Time (s)
Sabri	10	0.05	16	0.04	3	0.08
Ülker	7	0.08	4	0.04	2	0.07
Araştırma	-	-	-	-	-	-
Merkezi	7	0.06	9	0.03	1	0.08
Kırtasiye	63	0.08	69	0.04	19	0.08
Stationery	1	0.07	4	0.03	1	0.07
TDV	-	-	4	0.03	-	-
Book	1	0.09	2	0.04	1	0.08
Store	1	0.06	2	0.05	1	0.09
Kitabevi	1	0.07	3	0.04	1	0.09
Yesen	16	0.10	160	0.05	75	0.08
Burger	2	0.08	71	0.05	-	-

TABLE IV
PERFORMANCE COMPARISON OF OCR ALGORITHMS WITH THE SEQUENCE MATCHER (SM) METHOD

	Easy OCR & SM	**Keras OCR & SM**	**Tesseract OCR & SM**
Recognized Labels	Recognition Nr.	Recognition Nr.	Recognition Nr.
Sabri	28	39	11
Ülker	19	16	7
Araştırma	6	7	1
Merkezi	20	27	4
Kırtasiye	89	99	35
Stationery	8	14	7
TDV	5	13	3
Book	7	9	2
Store	4	8	5
Kitabevi	8	10	4
Yesen	31	246	115
Burger	6	127	4

Table IV demonstrates the performance results of each algorithm after applying the word similarity calculation method. The experiments show that applying word similarity methods positively affects character recognition algorithms, especially in limited labels.

After examining the experiments, it can be concluded that Keras OCR combined with the EAST algorithm performs well in urban areas scenarios. However, Easy OCR and Tesseract OCR have some recognition and time limitations.

4. CONCLUSIONS

Extracting textual information in urban areas is vital because it includes crucial environmental information for robot navigation. The dataset was collected on the university campus with a camera mounted at the top of the mobile robot. As a first step, the EAST algorithm detected textual patterns in natural scenes. Secondly, the detected textual patterns were transmitted to three OCR algorithms, Easy OCR, Keras OCR, and Tesseract OCR, for the character recognition step. As a result, the performance of these OCR algorithms was tested, analyzed, and compared. Performance analysis of each OCR algorithm was conducted based on correctly recognized labels and recognition time. Additionally, we added the Sequence Matcher method to OCR algorithms to analyze its effect on OCR algorithms in urban areas. While capturing video from the camera during the robot's movement, the viewing angle was not always sufficient to read the entire text on the signboards; thus, some labels were recognized as missing one or two characters. However, by applying a fusion of OCR algorithms and the Sequence Matcher method, this problem has been eliminated, and the number of label recognition has increased significantly. When all experiments are investigated, the obtained results show that the Keras OCR algorithm achieved the highest results compared to the Easy OCR and Tesseract OCR algorithms in terms of accuracy, recognition number of labels with and without the Sequence Matcher method, and recognition time.

REFERENCES

[1] D. Sankowski and J. Nowakowski, *Computer Vision in Robotics and Industrial Applications*, vol. 3. WORLD SCIENTIFIC, 2014. doi: 10.1142/9090.

[2] M. Nagy and G. Lăzăroiu, "Computer Vision Algorithms, Remote Sensing Data Fusion Techniques, and Mapping and Navigation Tools in the Industry 4.0-Based Slovak Automotive Sector," *Mathematics*, vol. 10, no. 19, 2022, doi: 10.3390/math10193543.

[3] M. Bicakci, O. Ayyildiz, Z. Aydin, A. Basturk, S. Karacavus, and B. Yilmaz, "Metabolic Imaging Based Sub-Classification of Lung Cancer," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3040155.

[4] Z. Lin *et al.*, "A unified matrix-based convolutional neural network for fine-grained image classification of wheat leaf diseases," *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2891739.

[5] S. Cebollada, L. Payá, M. Flores, A. Peidró, and O. Reinoso, "A state-of-the-art review on mobile robotics tasks using artificial intelligence and visual data," *Expert Systems with Applications*, vol. 167. 2021. doi: 10.1016/j.eswa.2020.114195.

[6] M. Yousef, K. F. Hussain, and U. S. Mohammed, "Accurate, data-efficient, unconstrained text recognition with convolutional neural networks," *Pattern Recognit*, vol. 108, 2020, doi: 10.1016/j.patcog.2020.107482.

[7] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010. doi: 10.1109/CVPR.2010.5540041.

[8] C. Yao, X. Bai, W. Liu, Y. Ma, and Z. Tu, "Detecting texts of arbitrary orientations in natural images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012. doi: 10.1109/CVPR.2012.6247787.

[9] Y. Netzer and T. Wang, "Reading digits in natural images with unsupervised feature learning," *Nips*, 2011.

[10] H. Badri, H. Yahia, and K. Daoudi, "Fast and accurate texture recognition with multilayer convolution and multifractal analysis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014. doi: 10.1007/978-3-319-10590-1_33.

[11] L. Neumann and J. Matas, "Scene text localization and recognition with oriented stroke detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013. doi: 10.1109/ICCV.2013.19.

[12] S. Zhang, M. Lin, T. Chen, L. Jin, and L. Lin, "Character proposal network for robust text extraction," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2016. doi: 10.1109/ICASSP.2016.7472154.

[13] K. Wang, B. Babenko, and S. Belongie, "End-to-end scene text recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2011. doi: 10.1109/ICCV.2011.6126402.

[14] C. Shi, C. Wang, B. Xiao, S. Gao, and J. Hu, "End-to-end scene text recognition using tree-structured models," *Pattern Recognit*, vol. 47, no. 9, 2014, doi: 10.1016/j.patcog.2014.03.023.

[15] A. Mishra, K. Alahari, and C. V. Jawahar, "Top-down and bottom-up cues for scene text recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012. doi: 10.1109/CVPR.2012.6247990.

[16] C. Yao, X. Bai, B. Shi, and W. Liu, "Strokelets: A learned multi-scale representation for scene text recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014. doi: 10.1109/CVPR.2014.515.

[17] Y. Xin, D. Chen, C. Zeng, W. Zhang, Y. Wang, and R. C. C. Cheung, "High Throughput Hardware/Software Heterogeneous System for RRPN-Based Scene Text Detection," *IEEE Transactions on Computers*, vol. 71, no. 7, 2022, doi: 10.1109/TC.2021.3092195.

[18] J. Ma *et al.*, "Arbitrary-oriented scene text detection via rotation proposals," *IEEE Trans Multimedia*, vol. 20, no. 11, 2018, doi: 10.1109/TMM.2018.2818020.

[19] W. He, X. Y. Zhang, F. Yin, and C. L. Liu, "Deep Direct Regression for Multi-oriented Scene Text Detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017. doi: 10.1109/ICCV.2017.87.

[20] J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," *IEEE Access*, vol. 8. 2020. doi: 10.1109/ACCESS.2020.3012542.

[21] L. Q. Zuo, H. M. Sun, Q. C. Mao, R. Qi, and R. S. Jia, "Natural Scene Text Recognition Based on Encoder-Decoder Framework," *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2916616.

[22] C. Yao, X. Bai, and W. Liu, "A unified framework for multioriented text detection and recognition," *IEEE Transactions on Image Processing*, vol. 23, no. 11, 2014, doi: 10.1109/TIP.2014.2353813.

[23] O. Y. Ling, L. B. Theng, A. Chai, and C. McCarthy, "A model for automatic recognition of vertical texts in natural scene images," in *Proceedings - 8th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2018*, 2019. doi: 10.1109/ICCSCE.2018.8685019.

[24] E. Zacharias, M. Teuchler, and B. Bernier, "Image Processing Based Scene-Text Detection and Recognition with Tesseract," Apr. 2020, Accessed: Mar. 31, 2023. [Online]. Available: https://keras-ocr.readthedocs.io/en/latest/

[25] "Zed2i Documentation." Accessed: May 01, 2023. [Online]. Available: https://www.stereolabs.com/docs/

[26] "Pixhawk Documentation." Accessed: May 01, 2023. [Online]. Available: https://pixhawk.org/

[27] O. M. T. Kaya and G. Erdemir, "Design of an Eight-Wheeled Mobile Delivery Robot and Its Climbing Simulations," in *Conference Proceedings - IEEE SOUTHEASTCON*, 2023. doi: 10.1109/SoutheastCon51012.2023.10115114.

[28] X. Zhou *et al.*, "EAST: An efficient and accurate scene text detector," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017. doi: 10.1109/CVPR.2017.283.

[29] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, "Character region awareness for text detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019. doi: 10.1109/CVPR.2019.00959.

[30] B. Shi, X. Bai, and C. Yao, "An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition," *IEEE Trans Pattern Anal Mach Intell*, vol. 39, no. 11, 2017, doi: 10.1109/TPAMI.2016.2646371.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016. doi: 10.1109/CVPR.2016.90.

[32] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[33] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks," in *ACM International Conference Proceeding Series*, 2006. doi: 10.1145/1143844.1143891.

[34] "Keras OCR Documentation." Accessed: Apr. 09, 2023. [Online]. Available: https://keras-ocr.readthedocs.io/

[35] R. Smith, "An overview of the tesseract OCR engine," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2007. doi: 10.1109/ICDAR.2007.4376991.

[36] "Difflib Library Documentation." Accessed: Apr. 09, 2023. [Online]. Available: https://docs.python.org/3/library/difflib.html

[37] J. W. Ratcliff and D. Metzener, "Pattern matching: The gestalt approach," *Dr. Dobb's Journal*, vol. 13, 1988.

BIOGRAPHIES

Erdal Alimovski received his B.Sc. degrees in Computer Engineering from St. Clement of Ohrid University-Bitola, North Macedonia in 2014, and the M.Sc. and Ph.D. degrees from Istanbul Sabahattin Zaim University, Istanbul, in 2020 and 2024, respectively. His research interests include computer vision, image processing, deep learning, machine learning, and robotics.

Gökhan Erdemir received the B.Sc., M.Sc., and Ph.D. degrees from Marmara University, Turkey. He was a Research Scholar with the Robotics and Automation Laboratory, Michigan State University, East Lansing, MI, USA, and the Health Management and Research Center, University of Michigan, Ann Arbor, MI, USA. He is currently an Associate Professor with the Engineering Management and Technology Department, The University of Tennessee at Chattanooga (UTC). His current research interests include control theory, robotics, industrial automation, AGVs, and engineering education.

Ahmet Emin Kuzucuoğlu received the B.Sc. degree from Electronics and Telecommunication Engineering Department, Istanbul Technical University, Turkey, in 1985, and the M.Sc. and Ph.D. degrees from Marmara University, Istanbul, Turkey, in 1994 and 2000, respectively. He is currently an Associate Professor with Marmara University. He has been in England and the United States of America in 1987 as part of YÖK-World Bank Vocational School Project. He has been in Lithuania in July-2006 as part of EU Leonardo da Vinci Type A mobility project. He is an Associate Professor with the Department of Electrical-Electronics Engineering. His current research interests include industrial automation, robotics, AI, control theory and applications.

European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.14, No.1, 2024

Research Article

Ballistic impact response of hybrid composite plates

Ferhat Ateş[1], **Gurbet Örçen**[*2]

[1] Dicle University, Institute of Science, Sur, Diyarbakır, Turkey. (e-mail: ferhatatesmech@gmail.com).
[2*] Dicle University, Mechanical Engineering Department, 21280, Sur, Diyarbakır, Turkey. (e-mail: gurbetorcenr@dicle.edu.tr).

ARTICLE INFO

Received: Dec., 18. 2023
Revised: Jan., 12. 2024
Accepted: Jan, 12. 2024

Keywords:
Para-Aramid
Ballistic impact
Hybrid composite
Failure analysis

Corresponding author: *Gurbet Örçen*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1406586

ABSTRACT

The behaviors exhibited by materials used in military and civilian daily applications in response to ballistic impacts constitute an important field of study. These behaviors have been investigated experimentally on composite plates with different configurations. For this purpose, Para-Aramid (Kevlar)/epoxy and Para-Aramid (Kevlar)/Glass hybrid composite plates with two different configurations were produced. Specimens obtained from these plates were subjected to tests at velocities corresponding to armor levels IIIA, IIA, and II. As a result of the tests, the effects of the velocities on the target specimens were compared. At the same time, the deformations on the plates were also examined. The study, conducted at three different velocities, yielded positive results from the Para-Aramid (Kevlar)/epoxy composite in the level IIA test, while negative results were obtained from the two hybrid plates. In the levels II and IIIA velocity tests, deformations were observed in the composite plates. Matrix cracks and fiber breakages were detected in the deformations obtained.

1. INTRODUCTION

Materials are exposed to different loading conditions in their areas of use. These loading types are factors that affect the life, use, and safety of the material. Impact loading in the ballistic field is one of the most important conditions under which materials will be affected. In this sense, the resistance of the material used against impact and puncture is an important parameter to seek.

The use of composite materials in ballistics is becoming widespread. The analysis of ballistic impact response and failure mechanisms in layered composites is complex. Since some variable parameters, such as composite composition, fibers, and volume, are affected, the level of complexity increases with the impact velocity. For this reason, a good understanding of the properties of composite materials is required for structural materials [1]. Again, the use of polymer matrix composites and hybrid composites is among the techniques used to increase the protection capabilities of composite materials in the ballistic field. It is critical to comprehend how hybrid composites behave when subjected to ballistic impact loads. Factors affecting this loading response include fiber and matrix type, plate thickness, arranging sequence, geometry, and boundary conditions. Ballistic impact response additionally depends on the dimensions, form, and kinetic energy of the projectile [2].

Experimental [3-13,15] and numerical [14] studies have been conducted on the use of composite materials in ballistics. Signetti et al. [3] developed an analytical model and conducted extensive experimental ballistic tests to investigate the ballistic behavior of multi-layer composite protective armor subjected to high-velocity impacts of projectiles with random incidence angles, form, dimensions, and friction characteristics. They concluded that graduated multi-layer configurations provide higher toughness as the projectile enters the layers with decreasing fracture strength. Kędzierski et al. [4] performed experimental and numerical investigations of the strength of ten fabrics made from high-strength fibers. They investigated materials in a variety of sizes, such as multiaxial fabrics, unidirectional plates, and plain woven fabrics. They found that the normalized results obtained experimental and numerical revealed that most of the fabrics analyzed provided similar levels of protection, but the best strength was provided by multidirectional fabrics and the least strength was provided by plain woven fabrics. Deliktaş et al. [5] developed a numerical model for predicting ballistic impact performance. They estimated parameters influencing the ballistic performance of composite materials using the model. They found that projectiles with different geometries and the orientation of the plate had significant effects on the ballistic results. Tarım et al. [6] experimentally examined the ballistic impact performances of some polymer composites. In one experimental group, they produced glass fiber prepregs with 6, 12, 18, 22, 28, and 36 layers. In the second group, they

bonded the Al layer on both sides of 6, 12, 18, 22, 28, and 36-layer glass fibers. In ballistics tests, they found that thinner composite plates had higher flexibility than thicker ones. Furthermore, they discovered that while perforation did not occur in the 28th and 36th layers of the composite plate, full perforations occurred in the 18th and 22nd levels. Choudhury et al. [7] examined the effect of temperature changes on the mechanical properties of glass fiber reinforced polymer (GFRP) composites and their capacity to halt projectiles during ballistic impact. They indicated that plain woven GFRP was affected by cylindrical projectiles at four different temperatures. Zhikharev et al. [8] experimentally and numerically investigated the ballistic impact response of 2 mm-thick composite plates subjected to a uniaxial tension preload. They determined that the ballistic limit is decreased by the single axial tension preload. Naik and Shrirao [9] compared the ballistic impact behavior of E-glass/epoxy and T300 carbon/epoxy composites. They reported that the ballistic limit of E-glass/epoxy was higher than T300 carbon/epoxy for the same ballistic impact conditions. Reddy et al. [10] used a mild steel core projectile to study the ballistic performance of an E-glass/phenolic composite as a function of plate thickness and projectile impact velocity. They observed a non-linear connection between plate thickness and energy absorbing. Ansari et al. [11] investigated the [0°/90°] ballistic performance of glass fiber-reinforced polymer (GFRP) composites against 19 mm diameter steel nose cone projectiles. They also reproduced ballistic tests by performing three-dimensional finite element simulations. They determined that as the aspect ratio of the composite target increased, so did its ballistic performance, with the effect of the aspect ratio being more obvious in thicker targets. Jenq et al. [12] investigated the ballistic limit of flat woven glass/epoxy composite plates struck by a hard projectile weighing 14.9 g and having a 5 mm tip radius. They investigated the progressive failure modes of the targets. They performed quasi-static penetration tests to characterize the penetration process. They indicated that the rhombic delamination of impact-failure specimens was greater than that of quasi-statically penetrated specimens. Liu et al. [13] conducted impact tests using a 14.5 mm ballistic gun with a cylindrical projectile to examined the ballistic resistance of polyurea-coated carbon fiber-reinforced plastic (CFRP) composites. They found that as the initial velocity of impact increased, the deformation of the projectiles increased. Sorrentino et al. [14] evaluated an analytical model to determine both the performance of composite armor made of Kevlar 29 fabrics impregnated with thermosetting resin and the ballistic limit velocity. In their studies, they concluded that the ballistic test results showed that the technological process was appropriate and that the materials used to produce the armor provided good ballistic performances, showing ballistic limits about 8% higher than predicted by the analytical model. Naik et al [15] presented a review of the latest technology in the field of ballistics for kevlar composites. Research on Kevlar fiber-reinforced composites has been conducted, focusing on structural properties and characterization, development, applications in various engineering and related sectors, ballistic post-impact degradation mechanisms, and methods used in analysis.

In the literature, hybrid composites [16-26] made by combining many composite materials, such as Kevlar, glass, and carbon, have been studied. In an effort to increase ballistic impact strength, Peng et al. [16] investigated hybrid layered constructions made by combining compressed wood, Dyneema, ceramics, and Kevlar. They stated to have given the information needed to verify computer models that would help with further conceptual designs on hybrid ballistic panels and the development of lightweight composite armor. The high-velocity projectile impact failure of a hybrid composite consisting of carbon, glass, and Kevlar composites was numerically investigated by Karthick et al. [17]. They stated that carbon fiber composites are not as strong as Kevlar in terms of impact strength, but that by hybridizing carbon composites with Kevlar and glass, the impact strength and ballistic limit velocity can be raised. Yanen et al. [18] experimentally investigated the usability of layered hybrid composites in individual armor materials. They conducted ballistic tests of Glass Fiber/Aramid Fiber/Carbon Fiber layered composite plates with different fiber-reinforcement angles, different layer numbers, and different thicknesses. In their study, they stated that the ballistic resistance of fabrics with twill weaving structures is better. Bitlisli et al. [19] found that in composite armor materials for armored vehicles, ultra-high molecular weight polyethylene (UHMWPE), Aramid, Carbon, Hybrid and Glass Fiber types used as reinforcement materials, and PE film layers without using any resin as matrix material were placed layer by layer and then combined under pressure and temperature effects under press to form a layered structure. They fired shots in accordance with NIJ standards used in ballistic applications. In their study, they determined that the most suitable process as a production method is the thinnest and lightest of the bulletproof plates. Alarçin [20] used a hydraulic press to compress UHMPW-PE and carbon fiber materials in various layers at a specific pressure and temperature to create hybrid composite plates. He determined that by adding 1 layer, 3 layers, and 6 layers of carbon fiber prepreg to 25 layers of UHMPW-PE composite, the ballistic strength enhanced by 8.06%, 14.6%, and 22.6%. Sah et al. [21] conducted research to perform a numerical analysis of the ballistic effect on different hybrid composite combinations. They used different stackings of Kevlar 3D, Basalt 3D, and Hybrid-3D layers (combination of Kevlar 3D and Basalt 3D fabric layers). In the six hybrid composite panels they modeled (K3B3, B3K3, B3H3, H3B3, K3H3, and H3K3), they used a 9 mm full metal jacket (FMJ) projectile with an impact velocity range of 240 m/s to 350 m/s. They concluded that the B3K3 panel performed the best among all panels (K3B3, B3K3, B3H3, H3B3, K3H3, H3K3) with a ballistic limit of 332 m/s, while H3B3 performed badly with a ballistic limit of 251 m/s. Reddy et al. [22] studied the development of hybrid composite plates for armor applications to take advantage of the advantages of carbon and E-glass fibers. For this purpose, they produced three hybrid composite plates based on carbon and E-glass with epoxy resin matrix in weight ratios of 75:25, 50:50, and 25:75 and carbon and E-glass composites for comparison. They used a 7.62 mm mild steel projectile in their study. In terms of energy absorption, they discovered that composite plates with a 50:50 (CE 50-50) carbon to E-glass ratio performed best. Using hydrocode simulations, Kumar et al. [23] examined the influence of hybridization in hybrid composite armor under ballistic impact. They stated that hybridization improves the relative ballistic performance of hybrid composite armor. They determined that the hybrid layer's stacking sequence had an impact on ballistic performance metrics like energy absorption, residual velocity, and ballistic limit. Yavaş et al. [24] investigated the effects of ply number on the ballistic performance of Kevlar49/UHMWPEHB26 (ultra-high molecular weight polyethylene) layered hybrid composite. They conducted ballistic tests according to NIJ Standard-0101.04 level III standards. They stated that as the total number of plies in the composite specimens decreased, the impact depth increased. Randjbaran et al. [25] experimentally

investigated the influences of stacking order on the ballistic energy absorption of hybrid composites obtained with the use of glass, carbon, and kevlar woven fabrics. They calculated the velocity and absorbed energy. According to the authors, glass and carbon combined in the middle layers provide more efficiency. Yahaya et al. [26] used two distinct procedures to create woven kenaf-Kevlar hybrid composites, altering the woven kenaf content between 5.40 and 14.99 by volume. The hybrid composites were subjected to ballistic measurement experiments using fragment simulated projectiles at varying impact and residual velocities. In comparison to other hybrid composites, they discovered that the hybrid composites exhibited better ballistic performance.

In this study, the ballistic test performances of hybrid composites formed by the sequence Para-Aramid (Kevlar), which has a good strength property [27], with glass fiber-reinforced layers were experimentally investigated. Glass fiber reinforced fabrics were positioned in the middle levels of the study, while Para-Aramid (Kevlar) fabric was used in the upper and lower layers. In this sense, plates produced by three different groups were subjected to the test. Ballistic tests were performed on three different plates at three different velocities. The results obtained were evaluated and discussed.

2. MATERIAL AND METHOD

2.1. Preparation of composite plates

In this study, three different composite plates were produced. (Figure 1) All three productions were carried out in 11 layers. Hybrid composite production was carried out, with glass fiber reinforced fabric in the single-layer and 3-layer intermediate layers, and Para-Aramid 100 % (Kevlar) (it is abbreviated as K) in the lower and upper layers. In this sense, in the first group, kevlar/epoxy production was carried out. It was produced using 3Kevlar/1Glass/3Kevlar/1Glass/3Kevlar (3K/1G/3K/1G/3K) in Group II and 4Kevlar/3Glass/4Kevlar (4K/3G/4K) in Group III. Geometric drawings of these are given in Figure 1.

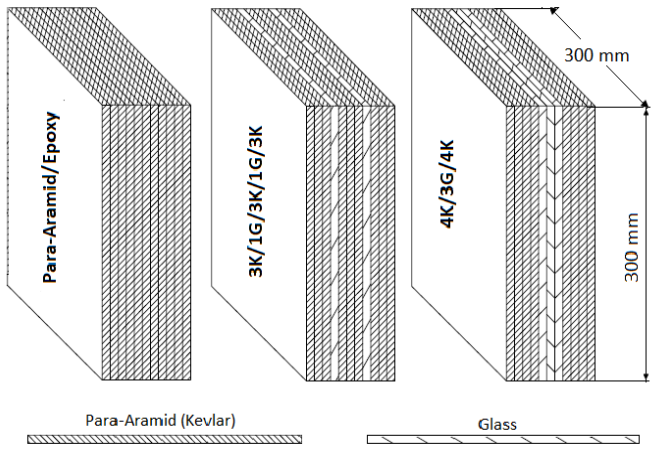


Figure 1. Demonstration of the sequence of layers

All three plates were produced in 500 mm x 500 mm dimensions using the hand lay-up method. Production was carried out at Fibermak Engineering Machinery Mold Composite San. Tic. Ltd. (İzmir/Turkey). The Para-Aramid 100 % (Kevlar) fabric used in the study was obtained from Kipaş Holding Company (Kahramanmaraş/Turkey). The fabric is woven and warp & weft type 3140 DTEX Flament with a Kevlar 100% structure. This Kevlar fabric is 0.4 mm thick, and 1 m^2 is 396.43 gr. F-RES 21-type epoxy resin (Fibermak, Turkey) with a density of 1.15 gr/cm^3 was used at 60%. F-HARD 22 type (Fibermak, Turkey) with a density of 1 g/cm^3 and 40% by volume was used as a hardener. Glass fiber-reinforced twill fabrics (Fibermak/Turkey) at 390 g/m^2 were used in the production of hybrid composites.

First of all, the 101 cm wide Kevlar fabric was prepared by cutting it in 11 layers with dimensions of 500 mm x 500 mm (Figure 2 a). Epoxy resin was applied to each fabric layer by hand lay-up method (Figure 2b). Then, after the gelation period of the fabrics, the semi-finished fabrics were stacked on top of each other, wrapped with fireproof film, and placed in the hydraulic hot press (Figure 2c). First, the fabrics were exposed to a pressure of 10 bars. After then, the temperature was increased to 125 °C from room temperature. For an hour, the plate was maintained in these conditions. The plates were allowed to reach room temperature when the heating element was shut off. The plates with 3K/1G/3K/1G/3K (Figure 2 d,e) in Group II and 4K/3G/4K (Figure 2f) in Group III were obtained by going through the same production steps.

The 11-layer and 4.5 mm thick Kevlar/epoxy composite plate in Group I (Figure 2h) and the plates in the other groups were cut with the help of a CNC machine in the dimensions of 300 mm x 300 mm (Figure 2g), which are the target plate dimensions for the experimental processes (Figure 2h).

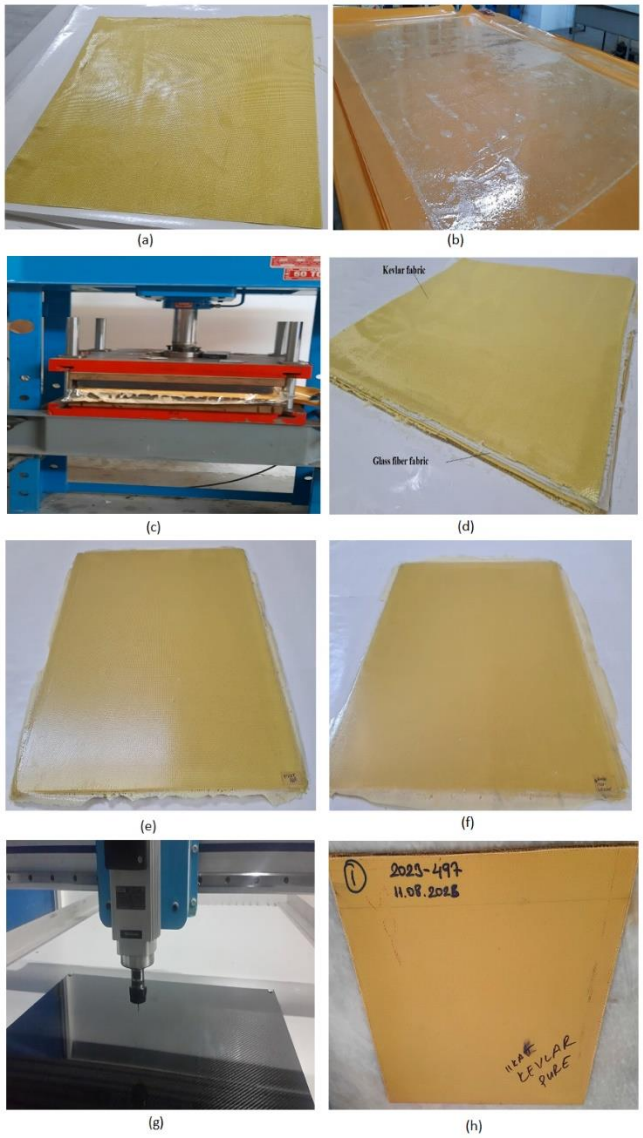


Figure 2. Production of composite plates

2.2. Ballistic Tests

In the Afyon 8th Main Maintenance Factory Directorate, ballistic impact tests were performed. NIJ Standard-0101.04 (The National Institute of Justice Standard) was used for the test setup. The properties of the 11-layer target plates are given in Table 1.

Before the test, the target plates were numbered according to with the standard (Figure 3a, b, c) in order to see the impact failure of the projectile impact on the target plates and the effect of boundary limits. They were then fixed to the setup for firing (Figure 3d). A schematic representation of the ballistic test setup according to NIJ Standard-0101.04 [28] is shown in Figure 3e.

A ballistic impact test apparatus operated by a stepped gas gun was used in the experimental study. A full metal jacket projectile (FMJ RN) was hit at the target specimen. According to 9×19 mm MP-5, FMJs are made in accordance with their definition by MKE (Mechanical and Chemical Industry). The masses, types, and velocities of the FMJ RN projectiles used in the experimental study were selected in accordance with the standard [28] and are indicated in Table 2. In addition, FMJ projectiles used for three different velocities in the study were fired at the target plate from a distance of 5 m. The tests were conducted at a temperature of 21.9 °C and a relative humidity of 48.0%.

The ballistic test starts with the firing of the projectile at a certain velocity, and when the projectile creates a hole in the target specimen, it is considered to have caused perforation and deformation. If the projectile does not create a hole in the target specimen, partial penetration is considered.

TABLE I. TARGET PLATE PROPERTİES

	Description	Weight (g)	Thickness (mm)
I.Group	Kevlar/epoxy	520.5	4.5
II.Group	3K/1G/3K/1G/3K	531.2	4.5
III.Group	4K/3G/4K	518.0	4.5

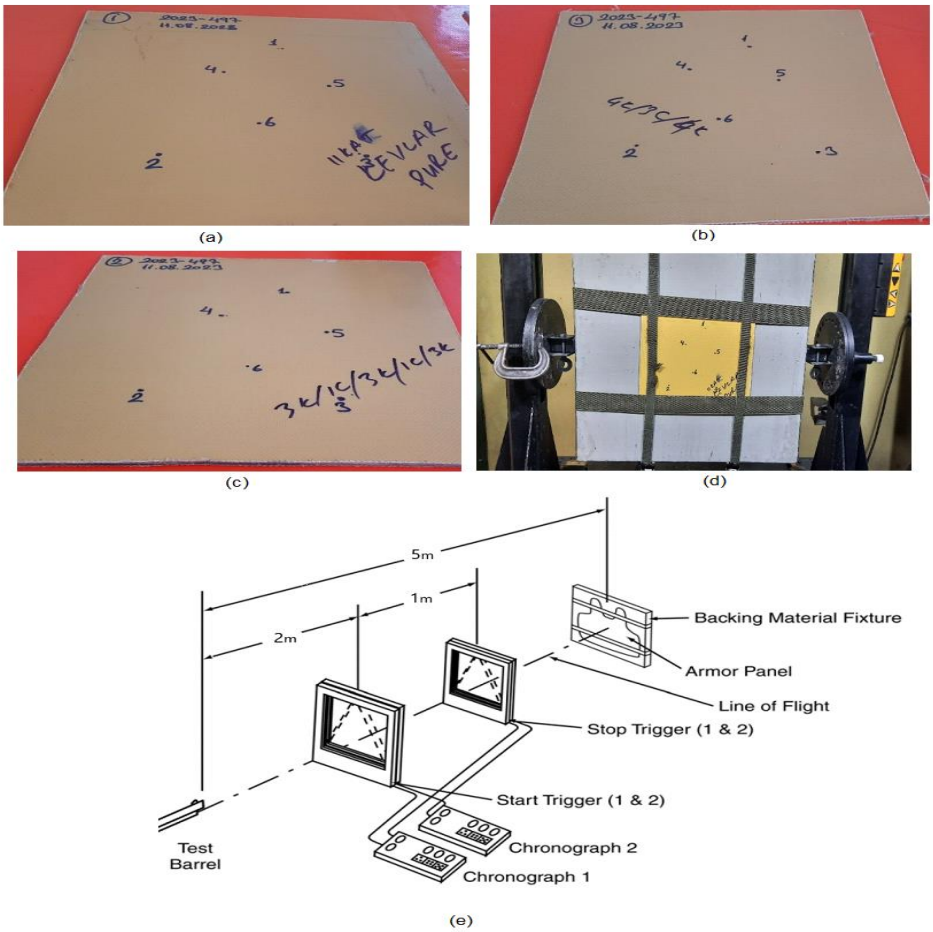


Figure 3. Ballistic testing procedures

TABLE II. EXPERIMENTAL TEST PARAMETERS

Armor-Type	Projectile-Type	Projectile-Weight	Projectile-Velocity
IIIA	9 mm FMJ RN	8.2 gr	436 ± 9.1 m/s
II	9 mm FMJ RN	8.0 gr	367 ± 9.1 m/s
IIA	9 mm FMJ RN	8.0 gr	341 ± 9.1 m/s

The aim of ballistic tests is to prevent or minimize this effect on the materials used. For this purpose, tests were carried out in three groups. In Group I, Kevlar with epoxy resin; in Group II, 3K/1G/3K/1G/3K; and in Group III, target composite plates with a 4K/3G/4K array were subjected to ballistic tests. The tests were conducted at the velocities specified in levels IIIA, IIA, and II (Table 2). These speeds are chosen according to whether the target plate is perforated or not. First, the target was shot at the velocity in level IIIA, then at the velocity in level II, and then at the velocity in level IIA, according to the result obtained. Since the high velocity impact was realized, the failure of the target plate was evaluated.

3. RESULT AND DISCUSSION

3.1. Ballistic results of hybrid composites

Kevlar/Epoxy, 3K/1G/3K/1G/3K, and 4K/3G/4K target composite plates were examined after the test. During the tests, the failure was analyzed to see if the target was completely perforated from behind. Group I; the average velocity values were realized in the range of 431 m/s in the shots fired at the IIIA level in the Kevlar/epoxy composite plate and perforation occurred in the plate (Figure 4). Then the shooting at level II was started. The shots here were fired at an average range of 362 m/s. Perforation was also detected as a result of these shots. The last shot was fired with a speed of IIA level. This firing rate was realized at an average speed of 344 m/s. As a result of the test, it was determined that no perforation occurred in the target plate. It was determined that the depth of depression was 12.11 mm. According to NIJ Standard-0101.04, the depth of depression should be a maximum of 44 mm. Therefore, it was determined that the result obtained in the test performed at level IIA velocity was appropriate according to the standard value.

In group II, the plate with the 3K/1G/3K/1G/1G/3K sequence was first tested at the velocity at level IIIA. In the measurements made, an average velocity of 434 m/s was realized. As a result of the test, it was determined that perforation occurred. Perforation was also obtained in the shots fired at level II, and tests were started at velocity level IIA. It was determined that the measured velocity range in this test was 344 m/s. As a result of this test, perforation occurred in the plate (Figure 5).

The composite plate with the 4K/3G/4K sequence in group III was subjected to tests at the velocities of all three levels. It was determined that the shooting velocities at level IIIA were 434 m/s, level II was 363 m/s, and level IIA was 348 m/s. In these test results, perforation occurred in the plate (Figure 6). The results of the tests are given in Table 3.

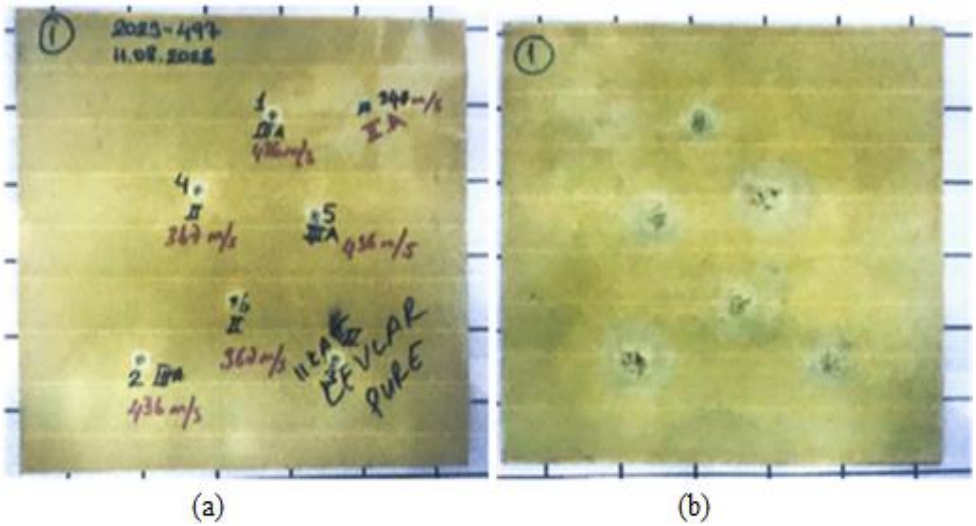
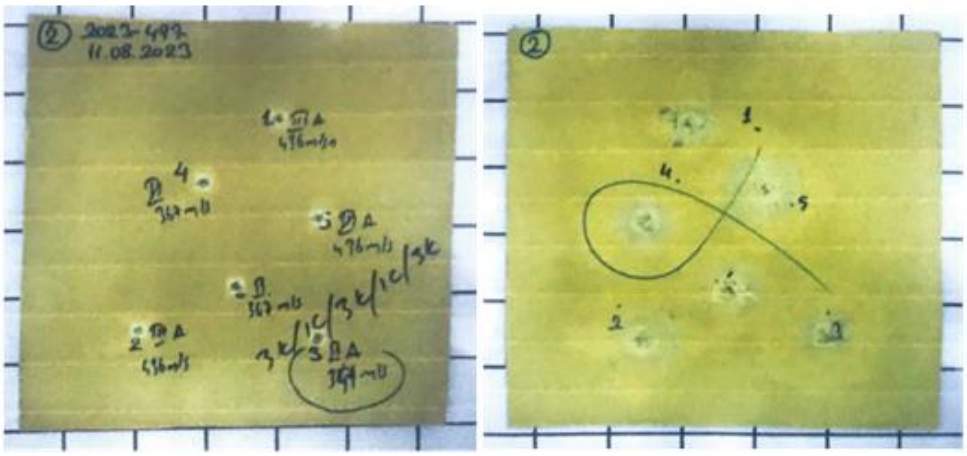
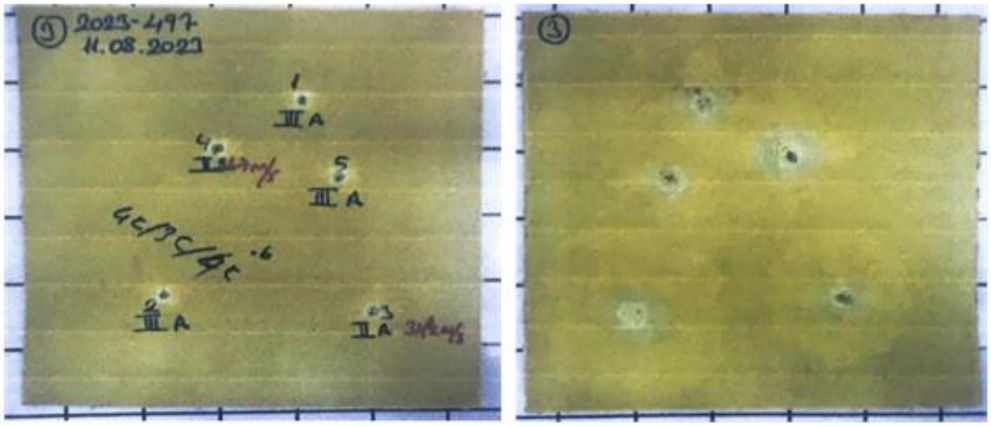


Figure 4. Shotted by projectiles Kevlar/epoxy plate a) Front side b) Back side



(a) (b)

Figure 5. Shotted by projectiles 3K/1G/3K/1G/1G/3K plate a) Front side b) Back side



(a) (b)

Figure 6. Shotted by projectiles 4K/3G/4K plate a) Front side b) Back side

TABLE II. BALLISTIC TEST RESULTS OF HYBRID COMPOSITES

Plate	Level of Velocity	Failure	Deflection
Kevlar/epoxy	IIIA	Perforated	-
	II	Perforated	-
	IIA	Not Perforated	12.11 mm
3K/1G/3K/1G/3K	IIIA	Perforated	-
	II	Perforated	-
	IIA	Perforated	-
4K/3G/4K	IIIA	Perforated	-
	II	Perforated	-
	IIA	Perforated	-

3.2. Failure Modes

Apart from the level IIA test on the Kevlar/epoxy plate, perforation occurred in the plates as a result of tests performed at other velocities and levels. During ballistic impact, the energy transfer from the fired projectile to the target depends on the density, tensile strength, hardness, toughness, fracture strength, strain-strength relationship, and fracture toughness of Kevlar/epoxy [29]. In this study, it was determined that fiber-matrix failure, matrix cracking, and fiber breakages occurred during ballistic impact as a result of shots at all three

levels. At the same time, delamination occurred because of the shear stress occurring between the layers [29].

In this study, ballistic resistance was observed by examining the failure areas on the back side of the target plates according to the shooting levels. After the shots at IIIA and IIA level, it was observed that the affected failure area (matrix cracking) of the hybrid composite plate with 3K/1G/3K/3K/1G/3K sequence was less than the affected area of the other hybrid plate. However, it was observed that the fiber fractures in the hybrid plate with 3K/1G/3K/1G/3K sequence were more noticeable than the fiber fractures in the other hybrid plate. In the Kevlar/epoxy plate, it was observed that the failure areas (fiber fracture + matrix cracks) affected after the shots were more. Therefore, this is thought to be due to the location and thickness of the glass/epoxy layers in the hybrid sequence. It is observed that there are matrix cracks around the point where the projectile hit when examining the Kevlar/epoxy plate, which was not penetrated at IIA level (Figure 7b). In all three plates, due to post-impact stress wave propagation [17], it was found that the failure areas affected on the back surface were more than the failure areas on the front surface. Some results of the plates are given as examples in Figures 7-9.

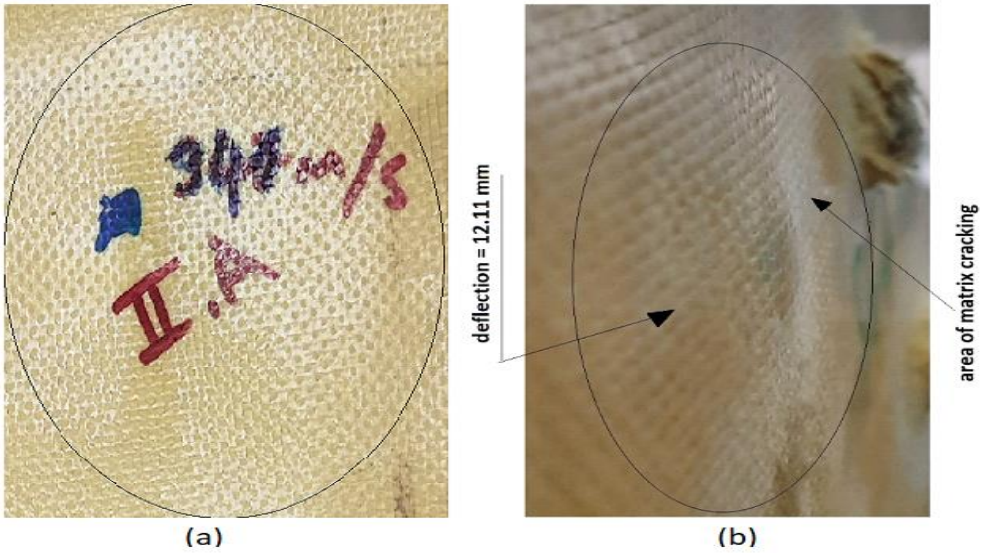


Figure 7. IIA level failure image of Kevlar/epoxy plate a) Front side b) Back side

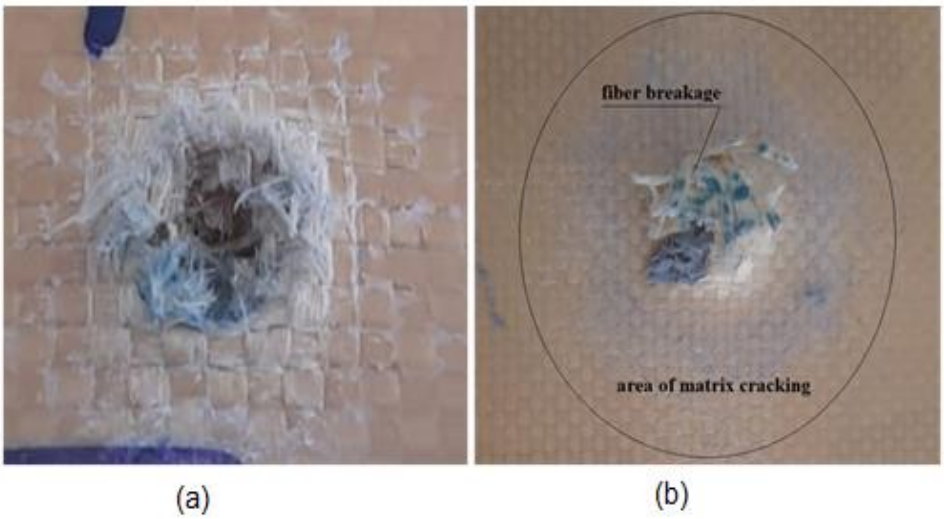


Figure 8. IIIA level failure image of 4K/3G/4K sequence plate a) Front side b) Back side

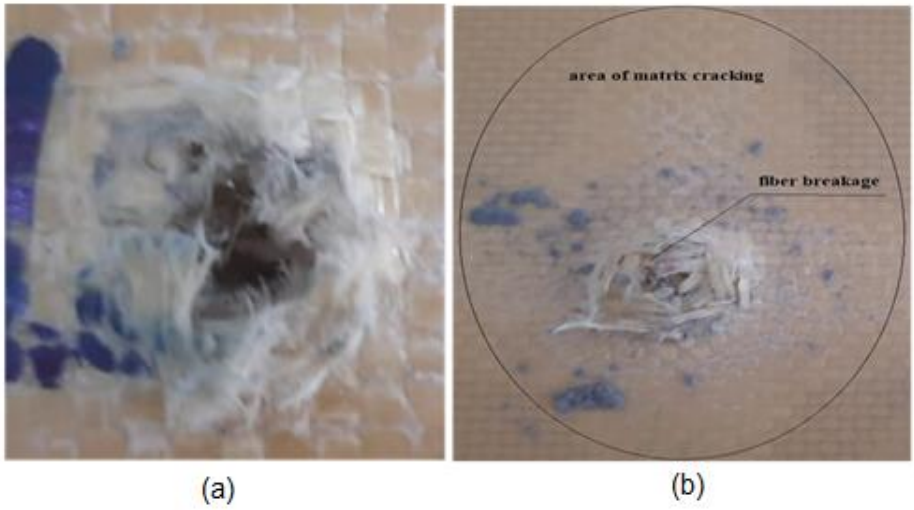


Figure 9. IIIA level failure image of 3K/1G/3K/1G/3K sequence plate a) Front face b) Back face

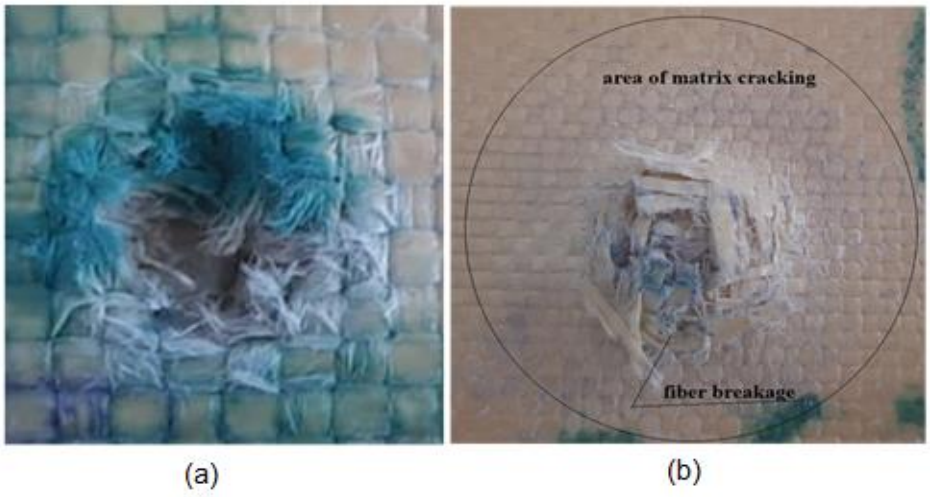


Figure 10. IIIA level failure image of Kevlar/epoxy plate a) front side b) back side

4. CONCLUSION

In this study, 11-layer composite plates were produced, in which Kevlar layers were placed in the lower and upper layers and glass fiber reinforced layers were placed in the intermediate layers. All production was carried out using the hand lay-up method. Ballistic tests were carried out on the produced Kevlar/epoxy, 3K/1G/3K/1G/3K, and 4K/3G/4K hybrid composite plates. Tests were carried out at three different levels with FMJ RN 9x19 mm-diameter projectiles on all three plates. After the test, the ballistic behavior of the plates was determined. The results obtained in the study are briefly stated below:

- The Kevlar/epoxy composite plates were not perforated in the shot made at the IIA level.
- Perforation occurred on all three plates in the shots made at levels IIIA and II.
- Matrix cracks and fiber breakages occurred in all perforated plates.

This work can be improved by changing the number of Glass and Kevlar layers used, their arrangement, and the types of fabric used. We believe these findings will further research in this area.

ACKNOWLEDGEMENT

This study was carried out with the contributions of the project numbered MÜHENDİSLİK.23.005, supported by DÜBAP. We would also like to thank Kipaş Holding Company (Kahramanmaraş, Turkey) for their contribution to the study.

REFERENCES

[1] Loganathan, T.M., Sultan, M.T.H., Gobalakrishnan, M.K., Muthaiyah, G., Ballistic impact response of laminated hybrid composite materials, Mechanical and Physical Testing of Biocomposites, Fibre-Reinforced Composites and Hybrid Composite. Woodhead Publishing Series in Composites Science and Engineering, (2019), pp. 171-191. DOI:10.1016/B978-0-08-102292-4.00010-2

[2] Pandya, K.S., Pothnis, J.R., Ravikumar, G., Naik, N.K., Ballistic impact behavior of hybrid Composites, Materials and Design, 44, (2013), pp. 128–135. DOI:10.1016/j.matdes.2012.07.044

[3] Signetti, S., Ryu, S., Pugno, N.M., Impact mechanics of multilayer composite armors: analytical modeling, FEM numerical simulation, and ballistic experiments, Composite Structures, 297 (2022). DOI:10.1016/j.compstruct.2022.115916

[4] Kędzierski, P., Popławski, A., Gieleta, R., Morka, A., Sławiński, G., Experimental and numerical investigation of fabric impact behavior, Composites Part B: Engineering, 69 (2015), pp. 452-459. DOI:10.1016/j.compositesb.2014.10.028

[5] Deliktaş, B., Poyraz, S., Durmuş, A., Sonlu elamanlar analizi ile kompozit malzemelerin balistik performansının tahmini, XX. Ulusal Mekanik Kongresi, (2017),pp.754-762.

[6] Tarım, N., Fındık, F., Uzun, H., Ballistic impact performance of composite structures, Composite Structures, 56 (2002), pp. 13–20. DOI:10.1016/S0263-8223(01)00177-5

[7] Choudhury, S., Ramagiri, B., Shah, B.K., Yerramalli, C.S., Guha, A., Ballistic response of woven glass fabric-epoxy composites at low temperatures: Experimental investigation, Composites Part C: 8 (2022), 100263. DOI: 10.1016/j.jcomc.2022.100263

[8] Zhikharev, M.V., Sapozhnikov, S.B., Kudryavtsev, O.A., Zhikharev, V.M., Effect of tensile preloading on the ballistic properties of GFRP, Composites Part B: Engineering, 168 (2019), pp. 524-531. DOI: 10.1016/j.compositesb.2019.03.026

[9] Naik, N.K., Shrirao, P., Composite structures under ballistic impact, Composite Structures, 66 (2004), pp. 579–590. DOI: 10.1016/j.compstruct.2004.05.006

[10] Reddy, P.R.S., Reddy, T.S., Madhua, V., Gogia, A.K., Rao, K.V., Behavior of E-glass composite laminates under ballistic impact, Materials & Design, 84 (2015), pp. 79-86. DOI: 10.1016/j.matdes.2015.06.094

[11] Ansari, M.M., Chakrabarti, A., Iqbal, M.A., An experimental and finite element investigation of the ballistic performance of laminated GFRP composite target, Composites Part B, 125 (2017), pp. 211-226. DOI: 10.1016/j.compositesb.2017.05.079

[12] Jenq, S.T., Jing, H.S., Chung, C., Predicting the ballistic limit for plain woven glass/epoxy composite laminate, Int. J. Impact Eng 15 (1994), pp. 451-464. DOI: 10.1016/0734-743X(94)80028-8

[13] Liu, Q., Guo, B., Chen, P., Su, J., Arab, A., Ding, G., Yan, G., Jiang, H., Guo, F., Investigating ballistic resistance of CFRP/polyurea composite plates subjected to ballistic impact, Thin-Walled Structures, 166 (2021), 108111. DOI: 10.1016/j.tws.2021.108111

[14] Sorrentino, L., Bellini, C., Corrado, A., Polini, W., Aricò, İ.R., Ballistic performance evaluation of composite laminates in kevlar 29, Procedia Engineering, 88 (2015), pp. 255-262. DOI: 10.1016/j.proeng.2015.06.048

[15] Naik, S., Dandagwhalb, R.D., Loharkar, P.K., A review on various aspects of Kevlar composites used in ballistic applications, Materials Today: Proceedings, 21(2020), pp.1366–1374. DOI: 10.1016/j.matpr.2020.01.176

[16] Peng, L., Tan, M.T., Zhang, X., Han, G., Xiong, W., Al Teneiji, M., Guan, Z.W., Investigations of the ballistic response of hybrid composite laminated structures, Composite Structures, 282 (2022), 115019. DOI: 10.1016/j.compstruct.2021.115019

[17] Karthick, P., Ramajeyathilagam, K., Numerical study on ballistic impact behavior of hybrid composites, Materials Today: Proceedings, 59 (2022), pp. 995-1003. DOI: 10.1016/j.matpr.2022.02.270

[18] Yanen, C., Solmaz, M.Y., Production of Laminated Hybrid Composites As A Body Armor Material And Investigation of Ballistic Performance, El-Cezerî Journal of Science and Engineering, 3(2016), pp. 351-362. DOI: 10.31202/ecjse.264200

[19] Bitlisli, B., Yazıcı, M., Investigation of the Ballistic Performances of Composite Materials Used in Armored Vehicles, Uludağ University Faculty of Engineering Journal, 24 (2019), pp. 25-34. DOI: 10.17482/uumfd.494262

[20] Alarçin, S., Investigation of ballistic resistance of ultra high density polyethylene and carbon fiber hybrid composites, Technological Applied Sciences, 15 (2020), pp.29-40. DOI: 10.12739/NWSA.2020.15.3.2A0182

[21] Sah, A. K., Pathak, R. K., Patel, S., Design and analysis of hybrid composite panels under ballistic impact, Materials Today: Proceedings, 87 (2023), pp. 104-109. DOI:10.1016/j.matpr.2023.02.031

[22] Reddy, P. R. S., Reddy, T. S., Mogulanna, K., Srikanth, I., Madhu, V., Rao, K. V., Ballistic Impact Studies on Carbon and E-glass Fibre Based Hybrid Composite Laminates, Procedia Engineering, 173 (2017), pp. 293-298. DOI: 10.1016/j.proeng.2016.12.017

[23] Kumar, B. A., Lakshmi, V., Ahmad, S., The effect of hybridization on the ballistic impact behavior of hybrid composite armors, Composites Part B: Engineering, 76 (2015), pp. 300-319. DOI: 10.1016/j.compositesb.2015.03.012

[24] Yavaş, M. O., Avcı, A., Şımşır, M., Akdemir, A., Ballistic Performance of Kevlar49/ UHMWPEHB26 Hybrid Layered-Composite, International Journal of Engineering Research and Development, 7 (2015), pp.7-27. DOI: 10.29137/umagd.379789

[25] Randjbaran, E., Zahari, R., Jalil, N. A. A., Majid, D. L. A. A., Hybrid Composite Laminates Reinforced with Kevlar/Carbon/Glass Woven Fabrics for Ballistic Impact Testing, The Scientific World Journal, (2014), 413753. DOI: 10.1155/2014/413753

[26] Yahaya, R., Sapuan, S.M., Jawaid, M., Leman, Z., Zainudin, E.S., Measurement of ballistic impact properties of woven kenaf–aramid hybrid composites, Measurement, 77(2016), pp. 335-343. DOI: 10.1016/j.measurement.2015.09.016

[27] Rajesh, S., Ramnath, B. V., Elanchezhian, C., Abhijith, M., Riju, R. D., Kishan, K. K., Investigation of Tensile Behavior of Kevlar Composite, Materials Today: Proceedings 5(2018), pp. 1156–1161. DOI: 10.1016/j.matpr.2017.11.196

[28] National Institute of Justice, Ballistic Resistance of Personal Body Armor NIJ Standard–0101.04. Office of Law Enforcement Standards National Institute of Standards and Technology Gaithersburg, 2001, MD 20899–8102.

[29] Pekbey, Y., Aslantaş, K., Yumak, N., Ballistic impact response of Kevlar Composites with filled epoxy matrix, Steel and Composite Structures, 24 (2017), pp. 191-200. DOI: https://doi.org/10.12989/scs.2017.24.2.191

BIOGRAPHIES

Gurbet Örçen obtained her PhD degree in Graduate School of Science, Engineering Faculty of the Firat university in 2011. She joined the Faculty of Mechanical Engineering, Dicle University as a research assistant, where he is presently a assistant professor. In general, she works actively on the mechanics and damage analysis of composites.

Ferhat Ateş obtained his BSc degree in Mechanical Engineering from Kocaeli University in 2021. He continues his master's degree in Mechanical Engineering at Dicle University, Institute of Science and Technology.His research interests are composite materials, nanomaterials and ballistic performances of armor.

European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.14, No.1, 2024

Research Article

Predicting Smart City Traffic Models using Adaboost and Gradient Boosting Method

Özlem BEZEK GÜRE[1*]

[1*] Batman University, Department of Medical Documentation, Secretariat Program Health Services Vocational School, Batman, Turkey. (e-mail:ozlem.bezekgure@batman.edu.tr)

ARTICLE INFO

Received: Feb., 13. 2024
Revised: May., 08. 2024
Accepted: May, 23. 2024

Keywords:
Smart City
Vehicle Traffic Prediction
Machine Learning
Ensemble Method
Adaboost Regression

Corresponding author:
Özlem Bezek Güre

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1436180

ABSTRACT

In parallel with the population density in cities, noise, traffic congestion, parking problems and environmental pollution also increase. To address these problems, smart transportation and traffic systems have emerged, which benefit from internet technologies to offer solutions that concern nearly everyone. These systems generate a vast amount of data, often analyzed through machine learning methods. This study has utilized the Adaboost method and Gradient Boosting (GB) method from the ensemble methods family within the machine learning framework to predict a smart city's traffic model. This method is a combination of many weak learners randomly selected from the data set and created by applying machine learning algorithms to form a strong learner. Both methods have been applied on a smart city traffic models data set found in the Kaggle database. This data set consists of a total of 48,120 rows and 4 columns, including variables such as the number of vehicles, number of intersections, date and time, and ID number. New variables have been created from the date and time variable before starting to analyze the data. The analyses performed with the Adaboost and GB method were carried out in Orange, a free Python-based program. Performance indicators such as Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2) have been used in the study. A 10-fold cross-validation method was used to ensure the validity of the model and to avoid overfitting. In conclusion, it has been observed that the Adaboost method performs successful predictions with low error rates. The Adaboost method, which estimates with minimum error, is also recommended for applications in areas such as smart grid, smart hospital, and smart home, in addition to smart traffic prediction.

1. INTRODUCTION

Transportation is one facet of urban life that affects almost all individuals in society. Particularly, with the increase in city populations and the consequent increase in the number of vehicles, transportation has become one of the most significant problems [1]. Dense urban traffic can lead to fatal, injurious, or financially damaging accidents. In addition to these, it can cause time loss, environmental pollution, and noise pollution [2]. Addressing these issues requires innovative solutions that go beyond traditional traffic management. For these reasons, there is a need for a traffic monitoring and reduction system in smart cities [3]. These systems utilize real-time data and predictive algorithms to optimize traffic flow, reduce congestion, and improve safety.

In smart cities, large amounts of data are collected as a result of using internet-based technologies [4]. These big data are not merely numerical values but represent a rich tapestry of urban life that can be harnessed for actionable insights. They are used both to determine past information and to predict the future, enabling authorities to make informed and proactive decisions. It is possible to plan for the future by developing action plans based on the results predicted with the analysis of big data [5].

The hidden patterns, complex, and nonlinear relationships within these massive data sets are analyzed using methods such as artificial intelligence, deep learning, and machine learning [6]. These sophisticated tools extend the capabilities of data analysis, enabling nuanced understanding and precise predictions. Using these methods, it is possible to predict traffic flow, classify vehicle images, and adjust traffic signal timing [7].

In this study, the Adaboot and GB method, one of the Ensemble methods family under the umbrella of machine learning methods, is used to predict the traffic model in a smart

city. Upon reviewing the literature, although there are many studies related to traffic prediction in smart cities using machine learning methods, the use of this specific method (Adaboost and GB) has not been encountered. This research, therefore, presents a novel approach to addressing a pressing urban challenge, contributing to the ongoing discourse on smart cities and their potential to transform contemporary urban living.

The current study makes several contributions to the existing literature. The Adaboost and GB methods from the boosting family have been employed to predict the traffic model of smart cities. These methods, which aim to combine weak learners to produce a stronger learner, have been utilized to enhance predictive accuracy. The unique application of these methods in this context allows for a more dynamic adjustment to varying traffic conditions, potentially leading to more efficient urban mobility solutions. Furthermore, the adaptive nature of the boosting algorithms makes them particularly suitable for handling the non-linear and complex patterns often present in urban traffic data. The performance of these methods was assessed using various metrics, including mean square errors (MSE), root mean square errors (RMSE), mean absolute errors (MAE), and the determination coefficient (R^2). This evaluation helps in understanding the effectiveness of the methods in the context of smart city traffic modeling.

The rest of the paper is organized as follows: Section 2 discusses the literature related to traffic in smart cities. Section 3 presents the methods used in the study. Section 4 provides the analysis results, while Section 5 includes discussions and recommendations.

2. RELATED WORKS

Various machine learning methods have been employed in existing literature for traffic prediction within the context of smart cities. These studies have explored an array of algorithms and techniques, each providing unique insights into traffic prediction and modeling.

Oyewola, Dada, and Jibrin [8] utilized a dataset from the present study, exploring methods such as Bagging, K-Nearest Neighbors (KNN), Multivariate Adaptive Regression Spline (MARS), Bayesian Generalized Linear Model (BGLM), and Generalized Linear Model (GLM). They concluded that the GLM method yielded more accurate predictions with reduced error, highlighting the importance of model selection in achieving desired predictive outcomes. Furthermore, Ismaeel et al. [9] used the Recurrent Neural Networks (RNN) method in their study, which utilized the mentioned dataset for classification purposes. In contrast, a study by Mohammed and Kianfar [10] applied Neural Networks, Random Forest (RF), Gradient Boosting Machine (GBM), and GLM to anticipate traffic flow, finding a slightly superior predictive performance by the RF method. This study emphasizes the potential of ensemble methods in capturing complex patterns within traffic data. Navarro-Espinoza et al. [11] conducted a comprehensive study involving a variety of methods, including RNN, Multilayer Perceptron Neural Network (MLP), GBM, RF, and Stochastic Gradient. Their findings suggest that the MLP method shows more successful performance, reinforcing the adaptability of neural networks in handling non-linear relationships. Ramesh [12] used RF, Adaboost, and Logistic Regression methods, noting the more successful classification performance of logistic regression. The study underscores the effectiveness of Logistic Regression in binary classification problems within traffic prediction. Furthermore, Boukerche and Wang [13] implemented a hybrid method consisting of RNN and Graph Convolutional Network, an innovative approach that leverages the strengths of both techniques. An and Wu [14] added to this body of work by employing Neural Network methods, contributing to the ongoing exploration of neural architectures in traffic modeling. In the realm of incident classification, Devi, Alice, and Deepa [6] applied Support Vector Machine (SVM) and Logistic Regression methods to a similar dataset to classify incidents during heavy traffic times. Similarly, İbrahim and Hafez [15] achieved the best performance using the Decision Tree (DT) method among KNN, LR, SVM, Gaussian Naive Bayes (GNB), RF, and DT methods for classifying smart city traffic models. Lippmann et al. [16] used KNN, SVM, DT, and MLP methods and concluded that KNN and DT methods were more successful, offering valuable insights into the comparative performance of these popular algorithms. Saleem et al. [3] proposed a fusion-based intelligent traffic congestion control system, in conjunction with Artificial Neural Network and SVM methods. Their work represents an innovative approach to managing traffic congestion, stating that the proposed model had better classification performance. Other studies by Yıldırım, Birant, and Birant [17], Ozbayoğlu, Kucukayan, and Dogdu [18], and Niu et al. [19] have further contributed to the field by utilizing various machine learning algorithms to address different challenges within traffic prediction. Additionally, the literature includes studies that employ the Adaboost and Gradient Boosting methods used in the current study, illustrating the widespread adoption of these techniques in prediction and modeling.

Consequently, the related studies present a rich tapestry of approaches and methodologies in the domain of traffic prediction within smart cities. The collective insights from these works contribute to a better understanding of the complex dynamics of urban traffic and offer valuable guidance for future research and applications.

3. MATERIAL AND METHOD

In the present study, the smart city traffic models dataset, available in the Kaggle database, has been utilized [20]. The dataset, consists of traffic records from four junctions in a city between November 2015 and June 2017. Created to enhance city traffic management and increase the efficiency of services to citizens, the dataset aims to provide data that benefits future infrastructure planning [17]. Accordingly, the dataset, which includes variables such as ID number, date and time, vehicle count, and intersection count with 48,120 observation values, anticipates a robust traffic system for the city by preparing for heavy traffic. For this study, date and time information were reorganized as day, month, year, time slots (morning, afternoon, evening, and night), and weekdays/weekends. Predictions for Adaboost and GB were made using the free Python-based program Orange, and performance indicator values related to the predictions were obtained.

3.1.Adaboost Method

The Adaboost method is among the boosting algorithms. It is used to solve binary and multi-class classification problems as well as regression problems [21]. The method was developed by Freund and Schapire to enhance the performance

of different learning algorithms. Adaboost is a method that combines numerous weak learners created by randomly selecting from the dataset and applying machine learning algorithms to create a strong learner [22]. In the training phase, weights are assigned to each observation value. The assigned weights are used to learn each hypothesis [23]. The weights used to calculate the hypothesis error are recalculated at each iteration. Then, incorrect predictions are identified and higher weights are given to incorrectly predicted samples [22]. At each iteration, the prediction error is compared to a threshold that is used to increase or decrease the weight of the sample for the next iteration [24].

3.2.Gradient Boosting Method

The GB (Gradient Boosting) method is also among the machine learning methods used in classification and regression problems. The method was developed by Friedman in 1999 [25]. GB is an iterative method that combines a series of weak regression learners iteratively to create a single strong regression learner [26]. Affected by the presence of overfitting, the method is not sensitive to data types [27]. The GB method aims to find a cumulative model that minimizes the loss function. For this purpose, it uses the mean squared error [28]. In this method, a model is built incrementally by minimizing the expected values of a specific loss function. Increasing the number of trees in the model can lead to a small training error. To minimize the risks associated with prediction, it is necessary to optimally determine the number of iterations or trees [29]. The working principle of the method is simple: Initially, a decision tree is created from the dataset worked on. Then, the error amount between the prediction values of this decision tree and the output values is calculated. Subsequently, these new output values are used as residuals, also known as errors, for other samples. Thus, a new decision tree is created with these errors, and the process is repeated until the error created by the previously built tree is minimized [30]. In the model, errors are trained, giving more importance to observations that have been misclassified. Here, a gradient optimization process is applied to minimize the general error of strong learning [31].

The detailed analysis of traffic modeling in smart cities reveals significant patterns and insights. Utilizing various machine learning methods, with a focus on the Adaboost and GB method, the study uncovers essential trends related to vehicle distribution, intersection count, and prediction accuracy.

Performance Metrics

In this study, the following performance metrics were used: MSE, RMSE, MAE, R^2. R^2, which takes values between 0 and 1, is an indicator of the goodness of fit in regression equations. It shows the proportion of the variance in the dependent variable that is predictable from the independent variables.

4. RESULTS

Analyses were conducted using the smart city traffic models dataset available in the Kaggle database. The traffic condition is affected by factors such as the number of vehicles and the number of intersections. A graph showing the distribution of vehicle numbers over the years, based on the variables used in the study, is provided in Figure 1.

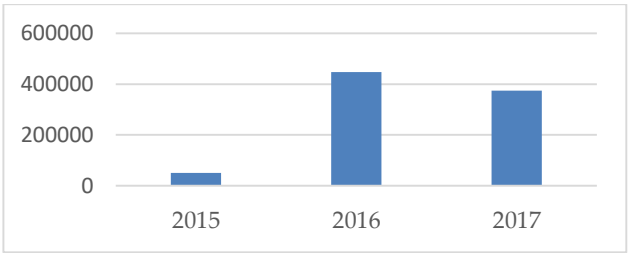


Figure 1. Vehicle count by year

Figure 1 illustrates the distribution of vehicles across different years, highlighting 2016 as the year with the highest number of vehicles and consequently the most traffic congestion. In contrast, 2015 is marked as the year with the lowest number of vehicles, leading to the least traffic congestion. These trends provide valuable insights into the fluctuations in traffic patterns over the observed period.

On the other hand Figure 2 shows a graph that illustrates the distribution of intersection counts over the years, which is another variable analyzed in the study.

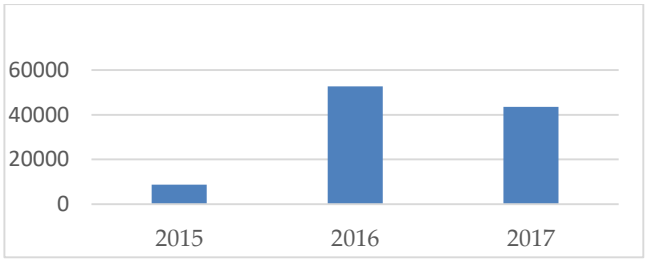


Figure 2. Number of junctions by year

Figure 2 shows a significant increase in the number of junctions between 2015 and 2016, followed by a decrease between 2016 and 2017.

Figure 3, 4, and 5 display the distributions of traffic density over time. Figure 3 presents a graph showing the number of vehicles by time for the year 2015.

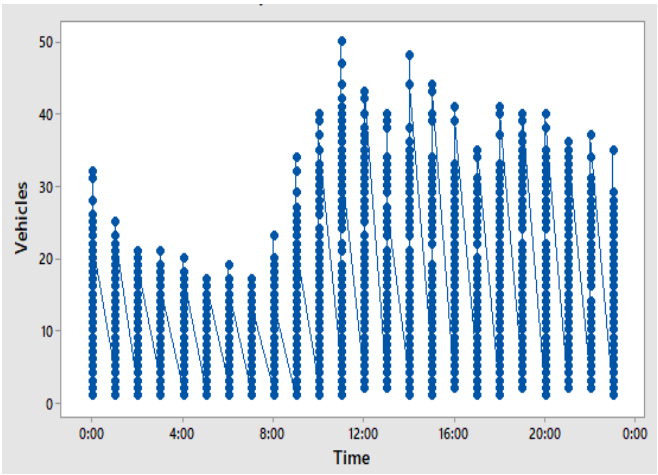


Figure 3. The distribution of vehicles over time in the year 2015

Figure 4 provides a graph showing the number of vehicles by time for the year 2016.

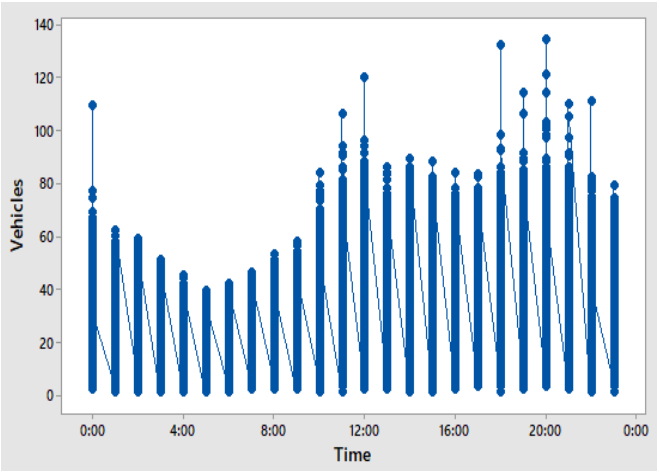


Figure 4. The distribution of vehicles over time in the year 2016

Figure 5 presents a graph showing the number of vehicles by time for the year 2017.

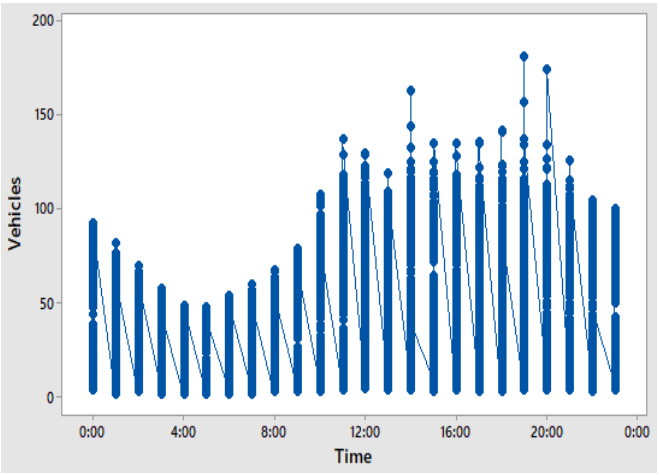


Figure 5. The distribution of vehicles over time in the year 2017

In the current study, the 10-fold cross-validation method is used for unbiased estimation. The cross-validation method is used to ensure the validity of the estimated model [32]. In the method, the data set is divided into 10 parts and one of them is used as test data and the other nine are used as training data. A different test data is used each time. Then, the overall error rate is calculated by averaging the error amounts of the 10 classes [33]. The performance indicators of the Adaboost and GB methods are given in Table 1.

TABLE 1.
PERFORMANCE METRICS

Method	MSE	RMSE	MAE	R^2
Adaboost	22.87	4.78	2.95	0.95
Gradient Boosting	42,61	6,53	4,10	0,90

Table 1 demonstrates that the Adaboost method achieves predictions with lower error amounts compared to the GB method. Additionally, the coefficient of determination measure shows that the independent variables used in the Adaboost method explain 95% of the variance in the dependent variable. This indicates a more successful prediction performance by the Adaboost method.

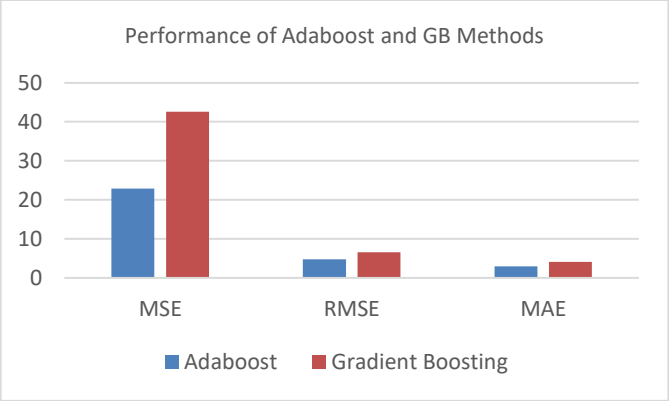


Figure 6. Performance of Adaboost and GB Methods

Based on Figure 6, it can be claimed that the AR method makes predictions with minimum error. As seen in Figure 6, the Adaboost method appears to make predictions with fewer errors compared to the GB method.

TABLE 2.
COMPARISON OF THE CURRENT STUDY WITH STUDIES IN THE LITERATURE

	Method	MSE	RMSE	MAE	R^2
The current study	Adaboost	22,87	4,78	2,95	0,95
	GB	42,61	6,53	4,10	0,90
Oyewola, Dada, and Jibrin [8]	Bagging	171,54	13,09	-	-
	KNN	85,30	9,23	-	-
	MARS	545,06	23,34	-	-
	BGLM	75,75	8,70	-	-
	GLM	75,34	8,68	-	-
Mohammed and Kianfar [10]	NN	-	8.63	5.95	0.93
	RF	-	5.56	3.57	**0.97**
	GBM	-	8.11	5.39	0.94
	GLM	-	9.94	6.74	0.92
Navarro-Espinoza et al. [11]	NN	-	9,80	7,24	0,95
	GB	-	9,66	7,12	0,94
	RF	-	9,57	7,05	0,94
	Stochastic Gradient	-	11,31	8,39	0,91
Zaman, Saha, and Abdelwahed [34]	Transformer	79.217	8.900	5.846	-
	LSTM	50.573	7.111	5.018	-
	BiLSTM	58.585	7.654	5.273	-
	Prophet	14208.68	119.20	102.81	-
Alekseeva et al. [35]	Bagging	-	2,99	1,66	50,8
	RF	-	3,38	2,19	34,2
	GB	-	**2,18**	1,43	60,2
	Bayesyen regression	-	2,25	1,49	49,7
	SVM	-	2,68	1,69	48,8
Tiwari [36]	LightGBM	-	4.14	2.49	-
	RF	-	3.95	2.36	-
	kNN	-	18.08	13.86	-
	XGBoost	-	18.25	14.12	-
Zheng et.al. [37]	CNN	-	24.52	13.01	
	Ensemble method	-	16.86	4.10	-
Savithramma, Sumathi & Sudhira [38]	SVM	8.30	2.88	0.92	-
	k-NN	9.14	3.02	1.20	-
	DT	16.07	4.00	2.00	-
	RF	9.00	3.00	1.36	-
	GB	**8.02**	2.83	**1.26**	-

5. CONCLUSION AND SUGGESTIONS

In this study, the Adaboost and GB method, a member of the Ensemble methods family within the scope of machine

learning techniques, has been employed to predict the traffic model in a smart city. An examination of the existing literature reveals a multitude of studies that have utilized various machine learning methods to forecast traffic in smart cities. Nonetheless, the combined use of Adaboost and GB methods has not been previously observed in the context of smart city traffic modeling.

The literature includes studies that apply machine learning methods to traffic data. Table 2 compares the results of the current study with those from the literature.

As previously mentioned, Table 2 compares the results of the current study with those of the study conducted by Oyewola, Dada, and Jibrin [8], which used the same dataset. It is observed that the Adaboost and GB methods demonstrate a more successful prediction performance with fewer errors.

On the other hand, the best performance values from similar studies in the literature are highlighted in bold in Table 2. The lowest error amounts in terms of MSE and MAE were achieved by the study conducted by Savithramma, Sumathi & Sudhira [38] using the GB (Gradient Boosting) method. In terms of RMSE, the lowest error was also achieved with the GB method in the study by Alekseeva et al. [35] and in terms of R2, the best performance was achieved by the RF (Random Forest) method in the study by Mohammed and Kianfar [10].

In the current study, an attempt was made to determine the prediction performance of the Adaboost and GB method using a large traffic-related dataset. The method is found to make predictions with minimal error. Traffic congestion is considered one of the most significant problems in traffic management [34]. Therefore, due to the accuracy of the predictions made, measures implemented to reduce traffic density will prevent the formation of vehicle queues, facilitate quicker access to desired locations, and result in lower noise and environmental pollution. All these results will ultimately affect the quality of life for citizens.

In smart cities, the use of internet technology results in the collection of large amounts of data in traffic and other areas. Effectively analyzing and managing this vast data is crucial. Future studies could use the same dataset with different machine learning methods. The performance of these methods under various conditions can be compared.

REFERENCES

[1] Bawaneh, M., & Simon, V. (2023). Novel traffic congestion detection algorithms for smart city applications. Concurrency and Computation: Practice and Experience, 35(5), e7563.

[2] Balasubramanian, S. B., Balaji, P., Munshi, A., Almukadi, W., Prabhu, T. N., Venkatachalam, K., & Abouhawwash, M. (2023). Machine learning based IoT system for secure traffic management and accident detection in smart cities. PeerJ Computer Science, 9, e1259.

[3] Saleem, M., Abbas, S., Ghazal, T. M., Khan, M. A., Sahawneh, N., & Ahmad, M. (2022). Smart cities: Fusion-based intelligent traffic congestion control system for vehicular networks using machine learning techniques. Egyptian Informatics Journal, 23(3), 417-426.

[4] Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., ... & Chiroma, H. (2016). The role of big data in smart city. International Journal of information management, 36(5), 748-758.

[5] Hemanth, D. J. (Ed.). (2022). Machine Learning Techniques for Smart City Applications: Trends and Solutions. Springer Nature.

[6] Devi, T., Alice, K., & Deepa, N. (2022). Traffic management in smart cities using support vector machine for predicting the accuracy during peak traffic conditions. Materials Today: Proceedings, 62, 4980-4984.

[7] Djenouri, Y., Michalak, T. P., & Lin, J. C. W. (2023). Federated deep learning for smart city edge-based applications. Future Generation Computer Systems, 147, 350-359.

[8] Oyewola, D. O., Dada, E. G., & Jibrin, M. B. (2022). Smart City Traffic Patterns Prediction Using Machine Learning. In Machine Learning Techniques for Smart City Applications: Trends and Solutions (pp. 123-133). Cham: Springer International Publishing

[9] Ismaeel, A. G., Janardhanan, K., Sankar, M., Natarajan, Y., Mahmood, S. N., Alani, S., & Shather, A. H. (2023). Traffic pattern classification in smart cities using deep recurrent neural network. Sustainability, 15(19), 14522.

[10] Mohammed, O., & Kianfar, J. (2018, September). A machine learning approach to short-term traffic flow prediction: A case study of interstate 64 in Missouri. In 2018 IEEE International Smart Cities Conference (ISC2) (pp. 1-7). IEEE.

[11] Navarro-Espinoza, A., López-Bonilla, O. R., García-Guerrero, E. E., Tlelo-Cuautle, E., López-Mancilla, D., Hernández-Mejía, C., & Inzunza-González, E. (2022). Traffic flow prediction for smart traffic lights using machine learning algorithms. Technologies, 10(1), 5.

[12] Ramesh, K. (2021). Smart Traffic Prediction and Congestion Reduction in Smart Cities. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(12), 1027-1033.

[13] Boukerche, A., & Wang, J. (2020). Machine learning-based traffic prediction models for intelligent transportation systems. Computer Networks, 181, 107530.

[14] An, C., & Wu, C. (2020). Traffic big data assisted V2X communications toward smart transportation. Wireless Networks, 26, 1601-1610.

[15] Ibrahim, A., & Hafez, A. (2023). Adaptive IEEE 802.11 ah MAC protocol for Optimization Collision Probability in IoT smart city data traffic Based Machine Learning models.

[16] Lippmann, R., Fried, D., Piwowarski, K., & Streilein, W. (2003, November). Passive operating system identification from TCP/IP packet headers. In Workshop on Data Mining for Computer Security (Vol. 40).

[17] Yıldırım, P., Birant, U. K., & Birant, D. (2019). EBOC: Ensemble-based ordinal classification in transportation. Journal of Advanced Transportation, 2019, 1-17.

[18] Ozbayoglu, M., Kucukayan, G., & Dogdu, E. (2016). A real-time autonomous highway accident detection model based on big data processing and computational intelligence. In 2016 IEEE international conference on big data (Big Data) (pp. 1807–1813).

[19] Niu, X., Zhu, Y., Cao, Q., Zhang, X., Xie, W., & Zheng, K. (2015). An online-traffic-prediction based route finding mechanism for smart city. International Journal of Distributed Sensor Networks, 11(8), 970256.

[20] Smart City Traffic Patterns. Available online: https://www.kaggle.com/datasets/utathya/smart-city-traffic-patterns/code (accessed on 15 June 2023).

[21] Shrestha, D. L., & Solomatine, D. P. (2006). Experiments with AdaBoost. RT, an improved boosting scheme for regression. Neural computation, 18(7), 1678-1710.

[22] Shanmugasundar, G., Vanitha, M., Čep, R., Kumar, V., Kalita, K., & Ramachandran, M. (2021). A comparative study of linear, random forest and adaboost regressions for modeling non-traditional machining. Processes, 9(11), 2015.

[23] Gupta, K.K.; Kalita, K.; Ghadai, R.K.; Ramachandran, M.; Gao, X.-Z. Machine Learning-Based Predictive Modelling of Biodiesel Production—A Comparative Perspective. Energies 2021, 14, 1122. [CrossRef]

[24] Kummer, N., & Najjaran, H. (2014). Adaboost. MRT: Boosting regression for multivariate estimation. Artif. Intell. Res., 3(4), 64-76.

[25] Dahiya, N., Saini, B., & Chalak, H. D. (2021). Gradient boosting-based regression modelling for estimating the time period of the irregular precast concrete structural system with cross bracing. Journal of King Saud University-Engineering Sciences.

[26] Gumaei, A., Al-Rakhami, M., Al Rahhal, M. M., Albogamy, F. R., Al Maghayreh, E., & AlSalman, H. (2021). Prediction of COVID-19 confirmed cases using gradient boosting regression method. Comput Mater Continua, 66, 315-329.

[27] Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review, 54, 1937-1967.

[28] Touzani, S., Granderson, J., & Fernandes, S. (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. Energy and Buildings, 158, 1533-1543.

[29] Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. Transportation Research Part C: Emerging Technologies, 58, 308-324.

[30] Yangın, G. (2019). XGboost ve Karar Ağacı tabanlı algoritmaların diyabet veri setleri üzerine uygulaması (Master's thesis, Mimar Sinan Güzel Sanatlar Üniversitesi, Fen Bilimleri Enstitüsü).

[31] Islam, S., & Amin, S. H. (2020). Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and

Gradient Boosting Machine learning techniques. Journal of Big Data, 7(1), 65.

[32] Kaur, H., Malhi, A. K. & Pannu, H. S. (2020). Machine learning ensemble for neurological disorders. Neural Computing and Applications, 32, 12697-12714.

[33] Aksu, G. & Doğan, N. (2018). Learning Methods Used in Data Mining Comparison under Different Conditions. Ankara University Journal of Faculty of Educational Sciences (JFES). 51(3). 71-100.

[34] Zaman, M., Saha, S., & Abdelwahed, S. (2023, June). Assessing the Suitability of Different Machine Learning Approaches for Smart Traffic Mobility. In 2023 IEEE Transportation Electrification Conference & Expo (ITEC) (pp. 1-6). IEEE.

[35] Alekseeva, D., Stepanov, N., Veprev, A., Sharapova, A., Lohan, E. S., & Ometov, A. (2021). Comparison of machine learning techniques applied to traffic prediction of real wireless network. IEEE Access, 9, 159495-159514.

[36] Tiwari, P. (2024). The machine learning framework for traffic management in smart cities. Management of Environmental Quality: An International Journal, 35(2), 445-462.

[37] Zheng, G., Chai, W. K., & Katos, V. (2019, December). An ensemble model for short-term traffic prediction in smart city transportation system. In 2019 IEEE Global Communications Conference (GLOBECOM) (pp. 1-6). IEEE.

[38] Savithramma, R. M., Sumathi, R., & Sudhira, H. S. (2022). A comparative analysis of machine learning algorithms in design process of adaptive traffic signal control System. In Journal of Physics: Conference Series (Vol. 2161, No. 1, p. 012054). IOP Publishing.

BIOGRAPHIES

Özlem Bezek Güre graduated in Statistics from Anadolu University. She received her MSc in Biostatistics from Fırat University, and her PhD from Van Yuzuncu Yil University in Statistics. She is currently an instructor at Batman University, working on machine learning, data mining, and statistic. She has studies published in national and international scientific platforms.

European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.14, No.1, 2024

Research Article

Wind Energy Forecasting Based on Grammatical Evolution

Şehmus FİDAN[1*],

[1*]Batman University, Department of Electronics and Automation, 72060, Batman, Türkiye (e-mail: sehmus.fidan@batman.edu.tr).

ARTICLE INFO

Received: Nov., 22. 2023
Revised: Jan., 24. 2024
Accepted: Jan, 25. 2024

Keywords:
Wind Energy
Energy Forecasting
Symbolic Regression
Grammatical Evolution
Genetic Expression Programming

Corresponding author: *Şehmus FİDAN*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1394289

ABSTRACT

The energy generated by wind turbines exhibits a continually fluctuating structure due to the dynamic variations in wind speed. In addition, in the context of seasonal transitions, increasing energy demand, and national/international energy policies, the necessity arises for short and long-term forecasting of wind energy. The use of machine learning algorithms is prevalent in the prediction of energy generated from wind. However, in machine learning algorithms such as deep learning, complex and lengthy equations emerge. In this study, the grammatical evolution algorithm, a type of symbolic regression method, is proposed to obtain equations with fewer parameters instead of complex and lengthy equations. This algorithm has been developed to derive a suitable equation based on data. In the study, through the use of grammatical evolution, it has been possible to obtain a formula that is both simple and capable of easy computation, with a limited number of parameters. The equations obtained as a result of the conducted analyses have achieved a performance value of approximately 0.91 (R^2). The equations obtained have been compared with methods derived using the genetic expression programming approach. In conclusion, it has been ascertained that the grammatical evolution method can be effectively employed in the forecasting of wind energy.

1. INTRODUCTION

The continuously increasing population and expanding industry require the resolution of challenging problems to meet the growing energy demand. The traditional energy sources utilized to meet this energy demand have resulted in significant pollution and caused global warming, which is challenging to reverse, on our planet. Among the developed solution proposals, renewable energy sources are quite promising, leading many governments to make significant investments in them. Wind turbines, holding a significant place among renewable energy sources, are of critical importance for an ecological and sustainable world [1]. However, there are uncertainties in obtaining energy from wind farms, including fluctuations in wind speed, continuously increasing demand and population density, and price instability [2]. To overcome these uncertainties, it is necessary to make short-term and long-term wind speed predictions with high accuracy. Various machine learning algorithms exist for predicting wind speed, but these models possess complex mathematical structures. Moreover, the training of networks containing a large number of layers and neurons can sometimes take days, and the resulting models, having a closed structure, produce equations that are quite challenging to interpret. Conversely, mathematical equations that are simple to compute manually, easy to read, and highly accurate are essential to reduce this complexity. In this study, a grammatical evolution (GE) based algorithm, a significant type of symbolic regression (SR), is proposed for predicting the output energy of wind turbines.

SR is a method where various mathematical models are proposed to understand the relationships among data. SR methods, encompassing various algorithms, typically employ evolutionary algorithms like genetic programming (GP) to find mathematical expressions (formulas) that best fit the datasets. In SR, there is no predetermined model. To find the mathematical model that fits the data, a selection is made from a wide pool of expressions using various operations (such as addition, subtraction, multiplication, division, exponentiation, logarithms, trigonometric functions etc.) [3]. Genetic expression programming (GEP), a type of SR method, is an evolutionary algorithm derived from the foundations of GP. GEP automatically proposes mathematical models to solve complex problems, fundamentally transforming the principles of biological evolution and natural selection into an algorithmic process [4, 5].

GE is a specialized algorithm used for SR and is a type of GP algorithm. GE bases itself on linguistic rules to find mathematical models or expressions that best fit dataset [6, 7].

Dufek et al. [8] proposed a methodology using GE for ensemble wind speed forecasting in northeastern Brazil, developing models that reduced forecasting errors by 7%-56%

compared to existing techniques, including a real-world operational model. Valsaraj et al. developed a new method using SR to estimate high altitude wind speeds from lower altitude data, achieving up to 61.04% reduction in daily RMSE compared to traditional methods, enabling more efficient wind resource assessment [9]. In another study, bayesian SR are used to find new models that better link energy consumption and pollution to socioeconomic factors, challenging existing stochastic impacts by regression on population, affluence, and technology (STIRPAT) assumptions [10].

Rueda et al. proposed using ant colony optimization for straight line programs in SR, comparing it with traditional GP for modeling real energy consumption problems [11]. Hybrid SR with deep multi-layer perceptron model has been proposed for PV power forecasting, achieving higher accuracy and robustness with significant reductions in RMSE and MAE, and improved training efficiency using advanced feature selection techniques [12]. Porras et al. [13] proposed a study for predicting the energy generation of a small wind turbine in a bioclimatic house in northwest Spain, analyzing atmospheric data over a year and exploring regression techniques to accurately forecast short-term power generation levels.

Rueda et al. suggested using SR with single and multi-objective algorithms to develop a generalized model for forecasting energy consumption time series, learning shared properties across different series [14]. Ramon et al. explored using neural networks and symbolization techniques for electric demand prediction, testing various architectures and methods on Spanish electric data. Their symbolization approach yielded slightly lower accuracy but was trained much faster [15].

Kochueva et al. proposed novel predictive models using SR and genetic algorithms for CO and NOx emissions from gas turbines, enhancing interpretability and transparency, and introduced a new classification model based on fuzzy inference systems [16]. Li et al. used SR and Tapio decoupling analysis to identify key factors influencing rural energy consumption in Henan, China, revealing significant elements and various decoupling statuses affecting energy usage patterns from 2000 to 2015 [17]. Kefer et al. proposed an energy management system optimized through GEP based SR, outperforming existing systems in minimizing energy costs, supporting grid stability, and prolonging battery life in residential buildings [18].

There are studies where GE algorithms have been used in the context of wind energy. Rodriguez et al. developed a grammatical swarm algorithm to forecast a country's total energy demand using macroeconomic variables, successfully testing it in Spain and France for one-year predictions [19]. Colmenar et al. introduced a novel hybrid approach using GE and differential evolution (DE) to generate and optimize models for predicting a country's total energy demand from macro-economic variables, demonstrating high accuracy in Spain and France [20]. Aditya et al. used GE algorithm to predict the daily load of a power plant from Indonesia's 2019 Jamali region electricity system data, outperforming the autoregressive integrated moving average model with a lower mean average percentage error (MAPE) of 1.77% [21]. Jamil et al. proposed a novel approach combining GE and DE for one-year-ahead energy demand estimation in Turkey, achieving high accuracy and outperforming previous models with a low RMSE [22]. Lourenço et al. developed a structured GE algorithm, hybridized with DE, to create accurate models for predicting annual energy demand in Spain based on macro-economic indicators [23].

The GE algorithm has also been employed for various other purposes. Lujan et al. developed an AI-based method using GE and DE to predict poly-crystalline silicon PV module temperature, significantly reducing error rates compared to the traditional sandia model under varying weather conditions [24]. In another study, the authors have proposed a multi-layered LSTM model for wind speed prediction [33]. Jeschke et al. developed a parameterized model predictive control (PMPC) for urban traffic using GE, reducing the computational complexity of real-time implementations while maintaining high performance efficiency [25]. In another study, Christou et al. designed a GE based method for automatic feature selection in radial basis function networks, achieving the highest classification accuracy (90.07%) for distinguishing hemiplegia types in patients using accelerometer sensor data [26].

In summary, in this article:

1- A GE based algorithm is proposed, deriving an equation with a minimal number of parameters. This results in a straightforward, easily readable, and writable equation for predicting wind energy.

2- Statistical analyses have been conducted on a dataset from an actual wind turbine.

3- For comparative purposes, GEP, a significant method in SR, has been utilized.

4- Performance comparisons have been made for the considered SR methods.

5- Various mathematical functions were employed in the equation generation phase, and the one yielding the best performance results is presented.

The article is organized into four sections. Section 1 provides an introduction and a summary of the literature. In section 2, explanations of the GE algorithm and its pseudocode are presented. The fundamentals of GEP are discussed, and the statistical metrics used are explained. In section 3, the equations and performance metrics obtained using the proposed GE based algorithm are presented, along with comparisons to GEP methods. Finally, section 4 presents the conclusions of the study.

2-MATERIALS VE METHODS

2.1 Wind Energy

The energy that can be harnessed from wind is directly proportional to the kinetic energy of the wind, as shown in Eq. 1. According to Betz's law, the maximum power a wind turbine can extract from wind is defined as approximately 59.3% of the theoretical limit [27].

$$P_w = \frac{1}{2}.\rho.A.C_p(\lambda,\beta)V^3 \quad (1)$$

A represents the turbine blade area (m²), C_p the performance coefficient, P_w the turbine power, V the wind speed (m/s), ρ the air density (1.225 kg/m³), β the blade angle (°), and λ the blade speed ratio. The blade area can be represented as shown in Eq. 2.

$$A = \pi R^2/4 \quad (2)$$

The performance coefficient changes depending on the blade speed ratio λ, wind speed V, the blades' angular rotation speed ω, and the blade radius R. Eq. 3 provides the blade speed ratio.

$$\lambda = \frac{R\omega}{V} \tag{3}$$

2.2. Grammatical Evolution (GE)

GE [7], rooted in GP, leverages a grammar-oriented approach. This method integrates a context-free grammar (CFG) with a rule-selection system executed through a genetic algorithm, facilitating the mapping procedure. CFG defines the structural rules of the language and allows GE to create various expressions. These algorithms reach solutions by utilizing the fundamental principles of genetic algorithms (selection, crossover, and mutation). Chromosomes generated by the genetic algorithm are transformed into a program or expression using CFG rules, thereby automatically determining mathematical expressions suitable for the problem. CFG defines the solution space of GEP and the structures it can generate.

CFG is a concept used in linguistics and comprises rules and symbol sets that express the structure of a language. In CFG, *T* (terminal symbol), *N* (non-terminal symbols), *S* (start symbol), and *R* are defined as production rules. CFGs are often expressed in Backus-Naur form (BNF) [28]. An example template for BNF is given between Eq. 4-7. In these equations, it is possible to generate production rules using non-terminal symbols $N = \{expr, op, coef, var\}$ and terminal symbols $T = \{+, -, x, \div, v_1, v_2, c_1, c_2, (,)\}$.

$$< expr > \coloneqq (< expr >) < op > (< expr >) \tag{4}$$

$$< op > \coloneqq + | - | x | \div \tag{5}$$

$$< coef > \coloneqq c_1 | c_2 \tag{6}$$

$$< var > \coloneqq v_1 | v_2 \tag{7}$$

Figure 1 shows the flow of the algorithm related to GE. In this algorithm, an initial population of solutions is created, and a context-free grammar is defined. Each iteration involves generating a program for each individual using CFG and evaluating its effectiveness via a fitness function. A new population is then formed by selecting the best individuals and applying genetic operations like crossover and mutation. The current population is replaced with this new one. The process repeats until a termination criterion, such as a maximum number of iterations or a desired fitness level, is met. Upon meeting this criterion, the best solution found is returned.

Figure 2 illustrates an example tree structure generated by GE and GEP. Although these two algorithms have different working structures, the structures of the generated trees resemble each other. However, there are some differences in terms of certain fundamental mathematical functions in the constructed structural tree.

```
Initialize:
    Create initial population
    Define CFG
Each iteration:
  For each individual:
      Generate program using CFG
      Evaluate program (fitness function)
      Create new population:
      Select best individuals
      Apply crossover and mutation
  Replace old population
  If met, termination criteria:
      If met, return best solution
  Else, continue iterations
```

Figure 1. GE Pseudocode

2.3. Genetic Expression Programming (GEP)

GEP is used to obtain equations based on data, similar to what GE does [5]. In the GEP method, chromosomes are represented as mathematical operations. The GEP method consists of five main components: a function set, a terminal set, control parameters, fitness functions, and termination conditions [29]. The function set represents mathematical operations, while the terminal set includes the algorithm's input variables. Control parameters define the population size, and the fitness function serves as a measure of the solution's effectiveness. The termination criterion determines when the iterative process will end [30].

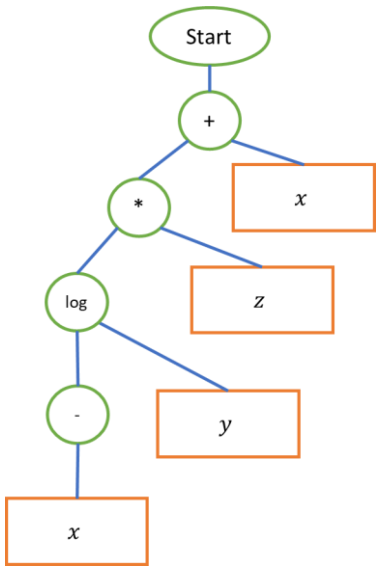


Figure 2. An example of a grammatical and genetic expression tree

In the genetic expression tree shown in Figure 2, the expression $(\log(y) - x) * z + x$ can be readily obtained. The orange-colored rectangular boxes represent terminals (x, y, z), while the circular green nodes represent mathematical operations $(+, -, *, log)$.

3. RESULTS AND DISCUSSION

3.1 Preliminary Analysis

Scada systems [34] in wind turbines measure and record data such as wind speed, direction, and generated power at 10-minute intervals, as exemplified by this file from a functioning wind turbine in Turkey. The dataset (50530) was recorded at 10-minute intervals between January 01, 2018, and December 31, 2018. The acquisition of the GEP tree and performance analyses were conducted using the HeuristicLab program developed by the Heuristic and Evolutionary Algorithms Laboratory [31]. The dataset was divided into three subsets: 15% for testing, 15% for validation, and 70% for training. This resulted in the use of 35371 data points for training, 7580 for validation, and 7580 for testing. In order to improve the performance during the testing phase, the data was randomly shuffled. The operation and function sets used in the construction of the GE algorithm are presented in Table 1. This table contains information about the function set, maximum generation count, population size, and depth used.

TABLE I
Parameter settings for the GE algorithm

Definitions	Parameters
Function Set	$+, -, *, /, \sqrt{}, \sqrt[3]{}, log, x^2,$
Number of Max. Generations	100
Population Size	1000
GE Depth	5-7

The operation and function sets used in the generation of genetic expression trees are presented in Table 2. Studies were conducted using the parameters in Table 2 to achieve the best performance of the genetic tree.

TABLE II
Parameter settings for the GEP

Definitions	Parameters
Function Set	$+, -, *, /, \sqrt{}, \sqrt[3]{}, log, x^2,$ $x^3, sin, cos, tan, tanh$
Number of Generations	200
Population Number	1000
GEP Depth	5
GEP Symbol Number	80
Mutation Probability	0.15

The sets of operations and functions used in the creation of the genetic expression tree are presented in Table 2. These parameters were determined as the most suitable through the research conducted. Table 3 displays the statistical values for the data collected from the wind turbine. Figure 3 displays the density distribution of four variables, including wind speed, direction, and generated power, measured during the operation of a wind turbine in Turkey.

TABLE III
Pre-descriptive statistics analysis

	Active Power (kW)	Wind Speed (m/s)	Theoretical Power (kWh)	Wind Direction (°)
count	50530	50530	50530	50530
mean	1307,684	7,557952	1492,175	123,6876
std	1312,459	4,227166	1368,018	93,44374
min	-2,47141	0	0	0
25%	50,67789	4,201395	161,3282	49,31544
50%	825,8381	7,104594	1063,776	73,71298
75%	2482,508	10,30002	2964,972	201,6967
max	3618,733	25,20601	3600	359,9976

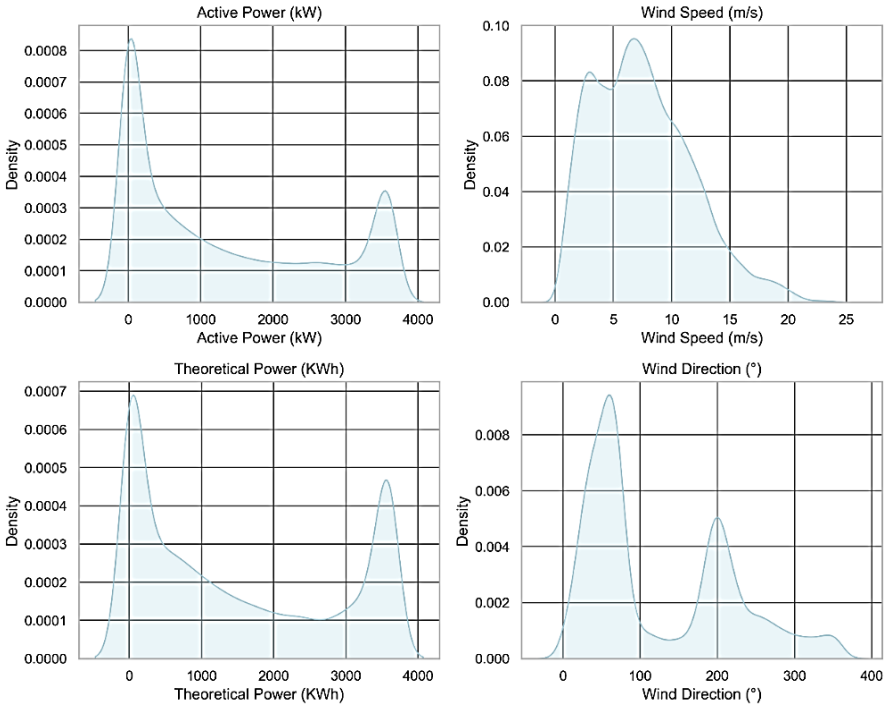


Figure 3. Density of active power, wind speed, theoretical power and wind direction

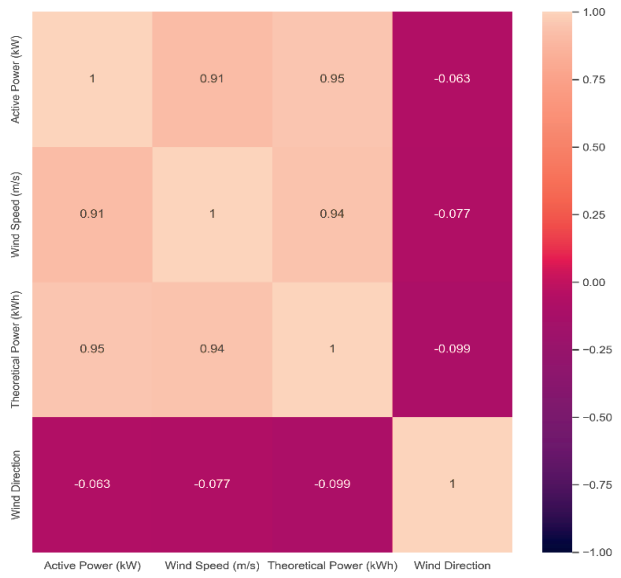


Figure 4. Correlation matrix

The correlation matrix is employed to examine the influence of variables on each other. If values are close to -1, there is a strong inverse relationship, if close to 1, there is a strong linear relationship between the data. If close to 0, there is no linear relationship between the data [32]. The correlation matrix presented in Figure 4 reveals a strong relationship (0.95) between active power and theoretical power. Additionally, as evident from the equations, there are robust relationships between wind and power (both theoretical and active).

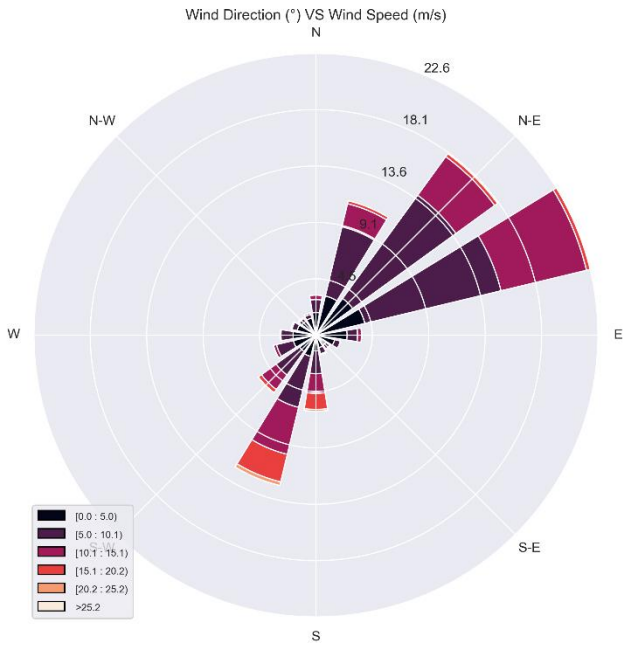


Figure 5. Relationship between wind speed and direction

Figure 5 illustrates the relationship between wind speed and wind direction. The figure demonstrates how wind direction varies with wind speed. The distribution of wind direction has been normalized based on wind speed. In this chart, the highest wind speeds (22.6 m/s) are observed in the easterly directions (NE and E). Particularly, in the east, wind speeds exceeding 20.2 m/s indicate that the wind is considerably strong. Conversely, in the northern and western directions, the wind appears to be relatively less frequent and predominantly registers speeds below 5.0 m/s.

3.2. Results of the Grammatical Evolution Analysis

The tree structure depicted in the graph integrates the values of theoretical power (T_p) and wind direction (W_d) within a complex equation to yield a result. Each node within this structure represents a mathematical operation, including addition, multiplication, and square rooting. The constants (-5.921, 1.823, and 3.69) are utilized to fine tune the dynamics of the equation, while average functions moderate the impact of input variables with a mean value. This model facilitates the analysis of the interactions between T_p and W_d, offering an output that is applicable in areas such as wind energy forecasting. For the proposed initial mathematical equation based on GE, the performance has achieved an R^2 value of 0.904 during the training phase and 0.902 (R^2) in the testing phase.

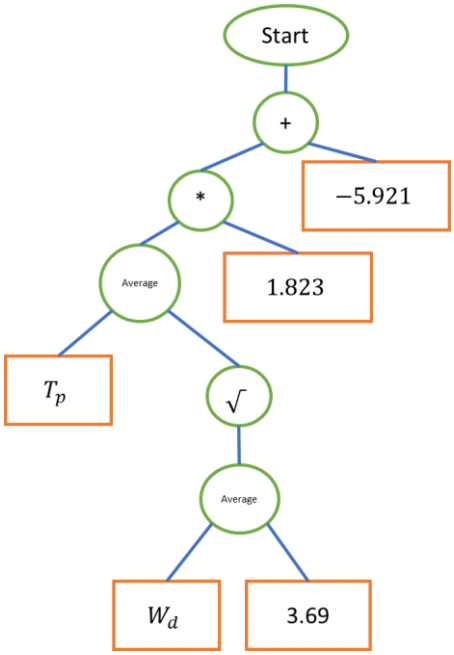


Figure 6. Proposed GE tree 1

The equation obtained using the tree depicted in Figure 6 is presented in Eq. 8. In this equation, P_A represents active power (kWh), P_T represents theoretical power (kWh), W_D denotes wind direction (°) and wind speed ($w_s - m/s$). Eq. 9 represents the second equation derived using GE. This equation has achieved a performance value of 0.9103 in the training phase and 0.909 in the testing phase.

$$P_A = 0.5 * \left(P_T + \sqrt{0.5 * W_D + 3.69}\right) * 1.823 - 59.21 \quad (8)$$

$$P_A = \left(P_T * log\left(3.82 + \frac{W_S}{\sqrt{3.69 * W_D}}\right)\right) * 0.321 + 2.817 \quad (9)$$

Figure 7 depicts the histogram of the number of elements in the tree when the tree depth is set to 5. The small number of symbols indicates that the structure of the predicted mathematical equation is simple. Figure 8 displays a tree structure created using the GE algorithm. Each node of the tree represents a mathematical operation or function. Basic arithmetic operations such as multiplication, division, and root

extraction, as well as the logarithm function, are included in this structure. Constants and variables (theoretical power, wind speed, wind direction) are combined with these operations to formulate an equation.

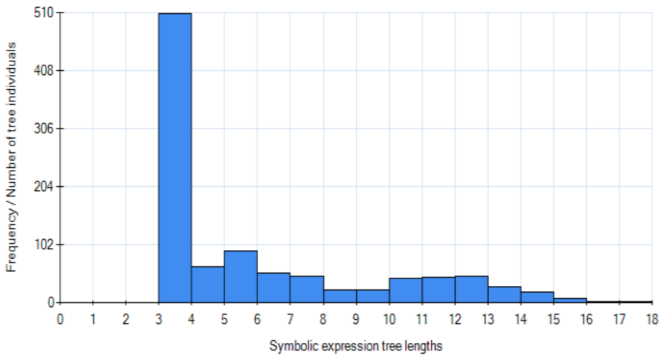


Figure 7. Proposed GE tree-1 histogram

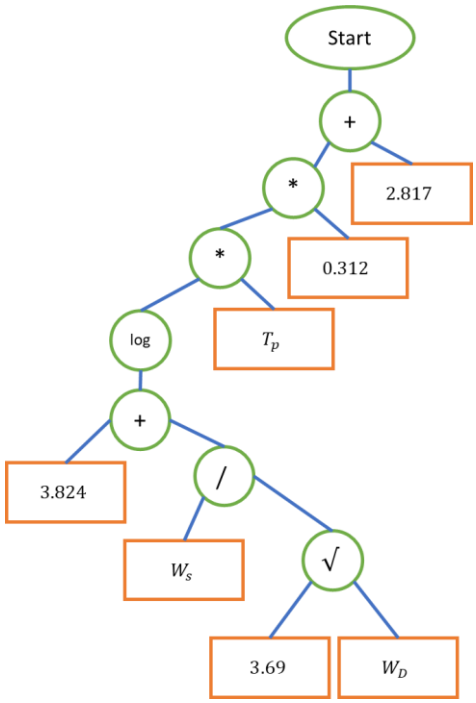


Figure 8. Proposed GE tree 2

Figure 8 presents the proposed GE tree, which consists of 14 symbols. Figure 9 shows the histogram of the number of tree elements when the tree depth is set to 7.

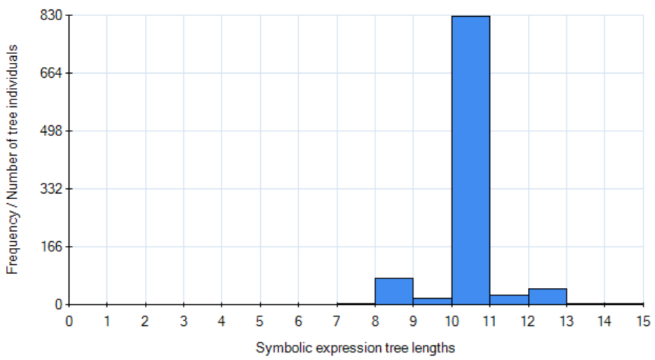


Figure 9. Proposed GE tree 2 histogram

3.3. Results of Genetic Expression Programming Analysis

Figure 10 displays the tree generated by the GEP algorithm. The tree is composed of operations such as logarithm, multiplication, and addition. Despite the simplicity of the resulting equation, it has attained a value of 0.906 during the training phase and 0.903 in the testing phase. Figure 11 illustrates how the tree depth varied in the attempts of the GEP algorithm to find Eq. 10.

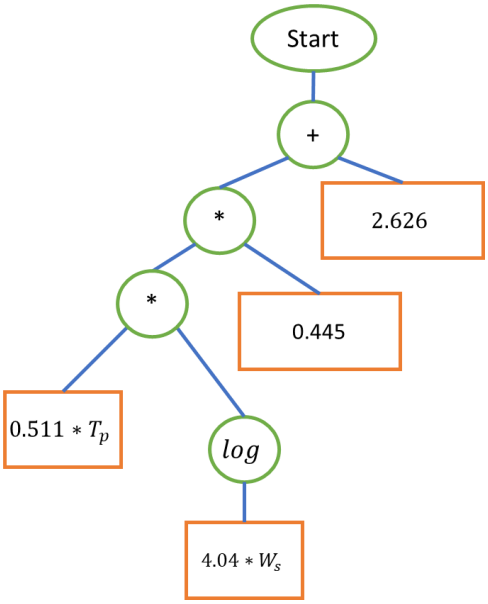


Figure 10. Proposed GEP tree

Eq. 10 represents the equation of the tree created in Figure 10. This equation expresses active power as dependent on theoretical power and wind direction. However, it should not be forgotten that the theoretical power is largely dependent on wind speed.

$$P_A = (0.445 * (0.511 * T_p * \log(4.04 * W_s))) + 2.626 \quad (10)$$

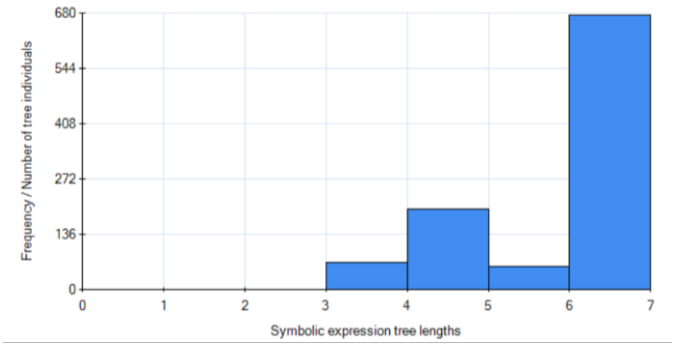


Figure 11. Proposed GEP tree-1 histogram

The three different equations proposed based on GE and GEP have been expressed with much simpler functions compared to the complex equations and time-consuming solutions of machine learning algorithms. Moreover, their overall performances possess values higher than 0.9, indicating that these R^2 performance values demonstrate the comprehensibility of the values in the dataset and the ease of generating equations. The Eq. 9, predicted using the GE

algorithm, not only exhibits superior performance (0.91) compared to other equations but can also be represented with just 14 symbols. Using 14 symbols demonstrates that the equations created are quite concise. Both GEP and GE have been able to generate highly successful results in producing equations.

4. CONCLUSION

Wind turbines hold a significant position among renewable energy sources and their usage is increasingly growing. However, the dynamic changes in wind speed, seasonal variations, increasing user demand, and the need to meet national policies make the prediction of wind energy essential. To make these predictions, numerous machine learning-based methods have been proposed in the literature. However, the training of these methods can be lengthy, and the resulting equations can be quite complex. This study utilizes wind speed, wind variation, and theoretical power information as inputs for predicting active power output from wind turbines and attempts to derive mathematical models. For this purpose, data collected via SCADA from a wind turbine in Turkey has been used. A Grammatical evolution-based algorithm has been proposed to derive equations dependent on this data. Through this algorithm, it is possible to generate equations containing a limited number of parameters and coefficients, thereby facilitating the creation of equations that are easily readable and writable. For comparison purposes, the grammatical expression programming method, one of the common symbolic regression techniques, has been proposed. The results indicate that the grammatical evolution algorithm has achieved a highly successful performance (0.91) in wind power prediction.

REFERENCES

[1] Ş. Fidan, M. Cebeci, ve A. Gündoğdu, "Extreme Learning Machine Based Control of Grid Side Inverter for Wind Turbines", *Teh. Vjesn.*, c. 26, sy 5, ss. 1492-1498, Eki. 2019, doi: 10.17559/TV-20180730143757.

[2] R. Saidur, N. A. Rahim, M. R. Islam, ve K. H. Solangi, "Environmental impact of wind energy", *Renew. Sustain. Energy Rev.*, c. 15, sy 5, ss. 2423-2430, Haz. 2011, doi: 10.1016/j.rser.2011.02.024.

[3] M. Schmidt ve H. Lipson, "Symbolic Regression of Implicit Equations", içinde *Genetic Programming Theory and Practice VII*, R. Riolo, U.-M. O'Reilly, ve T. McConaghy, Ed., içinde Genetic and Evolutionary Computation. , Boston, MA: Springer US, 2010, ss. 73-85. doi: 10.1007/978-1-4419-1626-6_5.

[4] Y. H. Çelik ve Ş. Fidan, "Analysis of cutting parameters on tool wear in turning of Ti-6Al-4V alloy by multiple linear regression and genetic expression programming methods", *Measurement*, c. 200, s. 111638, Ağu. 2022, doi: 10.1016/j.measurement.2022.111638.

[5] C. Ferreira, "Gene Expression Programming: a New Adaptive Algorithm for Solving Problems". arXiv, 30 Aralık 2001. doi: 10.48550/arXiv.cs/0102027.

[6] N. Lourenço, F. Assunção, F. B. Pereira, E. Costa, ve P. Machado, "Structured Grammatical Evolution: A Dynamic Approach", içinde *Handbook of Grammatical Evolution*, C. Ryan, M. O'Neill, ve J. Collins, Ed., Cham: Springer International Publishing, 2018, ss. 137-161. doi: 10.1007/978-3-319-78717-6_6.

[7] M. O'Neill ve C. Ryan, "Grammatical evolution", *IEEE Trans. Evol. Comput.*, c. 5, sy 4, ss. 349-358, Ağu. 2001, doi: 10.1109/4235.942529.

[8] A. S. Dufek, D. A. Augusto, P. L. S. Dias, ve H. J. C. Barbosa, "Data-driven symbolic ensemble models for wind speed forecasting through evolutionary algorithms", *Appl. Soft Comput.*, c. 87, s. 105976, Şub. 2020, doi: 10.1016/j.asoc.2019.105976.

[9] P. Valsaraj, D. A. Thumba, K. Asokan, ve K. S. Kumar, "Symbolic regression-based improved method for wind speed extrapolation from lower to higher altitudes for wind energy applications", *Appl. Energy*, c. 260, s. 114270, Şub. 2020, doi: 10.1016/j.apenergy.2019.114270.

[10] D. Vázquez, R. Guimerà, M. Sales-Pardo, ve G. Guillén-Gosálbez, "Automatic modeling of socioeconomic drivers of energy consumption and pollution using Bayesian symbolic regression", *Sustain. Prod. Consum.*, c. 30, ss. 596-607, Mar. 2022, doi: 10.1016/j.spc.2021.12.025.

[11] R. Rueda, L. G. B. Ruiz, M. P. Cuéllar, ve M. C. Pegalajar, "An Ant Colony Optimization approach for symbolic regression using Straight Line Programs. Application to energy consumption modelling", *Int. J. Approx. Reason.*, c. 121, ss. 23-38, Haz. 2020, doi: 10.1016/j.ijar.2020.03.005.

[12] M. Trabelsi *vd.*, "An Effective Hybrid Symbolic Regression–Deep Multilayer Perceptron Technique for PV Power Forecasting", *Energies*, c. 15, sy 23, Art. sy 23, Oca. 2022, doi: 10.3390/en15239008.

[13] S. Porras, E. Jove, B. Baruque, ve J. L. Calvo-Rolle, "A comparative analysis of intelligent techniques to predict energy generated by a small wind turbine from atmospheric variables", *Log. J. IGPL*, c. 31, sy 4, ss. 648-663, Tem. 2023, doi: 10.1093/jigpal/jzac031.

[14] R. Rueda, M. P. Cuéllar, M. Molina-Solana, Y. Guo, ve M. C. Pegalajar, "Generalised Regression Hypothesis Induction for Energy Consumption Forecasting", *Energies*, c. 12, sy 6, Art. sy 6, Oca. 2019, doi: 10.3390/en12061069.

[15] D. Criado-Ramón, L. G. B. Ruiz, ve M. C. Pegalajar, "Electric demand forecasting with neural networks and symbolic time series representations", *Appl. Soft Comput.*, c. 122, s. 108871, Haz. 2022, doi: 10.1016/j.asoc.2022.108871.

[16] O. Kochueva ve K. Nikolskii, "Data Analysis and Symbolic Regression Models for Predicting CO and NOx Emissions from Gas Turbines", *Computation*, c. 9, sy 12, Art. sy 12, Ara. 2021, doi: 10.3390/computation9120139.

[17] P. Li, C. Tian, Z. Zhang, M. Li, ve Y. Zheng, "Analysis of influencing factors of energy consumption in rural Henan based on symbolic regression method and Tapio model", *Energy Sources Part Recovery Util. Environ. Eff.*, c. 43, sy 2, ss. 160-171, Oca. 2021, doi: 10.1080/15567036.2019.1623951.

[18] K. Kefer *vd.*, "Simulation-Based Optimization of Residential Energy Flows Using White Box Modeling by Genetic Programming", *Energy Build.*, c. 258, s. 111829, Mar. 2022, doi: 10.1016/j.enbuild.2021.111829.

[19] D. Martínez-Rodríguez, J. M. Colmenar, J. I. Hidalgo, R.-J. Villanueva Micó, ve S. Salcedo-Sanz, "Particle swarm grammatical evolution for energy demand estimation", *Energy Sci. Eng.*, c. 8, sy 4, ss. 1068-1079, 2020, doi: 10.1002/ese3.568.

[20] J. M. Colmenar, J. I. Hidalgo, ve S. Salcedo-Sanz, "Automatic generation of models for energy demand estimation using Grammatical Evolution", *Energy*, c. 164, ss. 183-193, Ara. 2018, doi: 10.1016/j.energy.2018.08.199.

[21] I. A. Aditya, A. A. Simaremare, J. Raharjo, Suyanto, ve I. Wijayanto, "Daily Power Plant Load Prediction using Grammatical Evolution", içinde *2022 International Conference on Electrical Engineering, Computer and Information Technology (ICEECIT)*, Kas. 2022, ss. 122-126. doi: 10.1109/ICEECIT55908.2022.10030558.

[22] B. Jamil, L. Serrano-Luján, ve J. Colmenar, "On the Prediction of One-Year Ahead Energy Demand in Turkey using Metaheuristic Algorithms", c. 7, ss. 79-91, Ağu. 2022, doi: 10.25046/aj070411.

[23] N. Lourenço, J. M. Colmenar, J. I. Hidalgo, ve S. Salcedo-Sanz, "Evolving energy demand estimation models over macroeconomic indicators", içinde *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, içinde GECCO '20. New York, NY, USA: Association for Computing Machinery, Haz. 2020, ss. 1143-1149. doi: 10.1145/3377930.3390153.

[24] L. Serrano-Luján, C. Toledo, J. M. Colmenar, J. Abad, ve A. Urbina, "Accurate thermal prediction model for building-integrated photovoltaics systems using guided artificial intelligence algorithms", *Appl. Energy*, c. 315, s. 119015, Haz. 2022, doi: 10.1016/j.apenergy.2022.119015.

[25] J. Jeschke, D. Sun, A. Jamshidnejad, ve B. De Schutter, "Grammatical-Evolution-based parameterized Model Predictive Control for urban traffic networks", *Control Eng. Pract.*, c. 132, s. 105431, Mar. 2023, doi: 10.1016/j.conengprac.2022.105431.

[26] V. Christou *vd.*, "Grammatical Evolution-Based Feature Extraction for Hemiplegia Type Detection", *Signals*, c. 3, sy 4, Art. sy 4, Ara. 2022, doi: 10.3390/signals3040044.

[27] Ş. Fidan ve H. Çimen, "Rüzgâr Türbinlerinde Tork ve Kanat Eğim Açısı Kontrolü", *Batman Üniversitesi Yaşam Bilim. Derg.*, c. 11, sy 1, Art. sy 1, Haz. 2021.

[28] F. Noorian, A. M. de Silva, ve P. H. W. Leong, "gramEvol: Grammatical Evolution in R", *J. Stat. Softw.*, c. 71, ss. 1-26, Tem. 2016, doi: 10.18637/jss.v071.i01.

[29] B. Peng, S. Wan, Y. Bi, B. Xue, ve M. Zhang, "Automatic Feature Extraction and Construction Using Genetic Programming for Rotating

Machinery Fault Diagnosis", *IEEE Trans. Cybern.*, c. 51, sy 10, ss. 4909-4923, Eki. 2021, doi: 10.1109/TCYB.2020.3032945.

[30] I. Arnaldo, K. Krawiec, ve U.-M. O'Reilly, "Multiple regression genetic programming", içinde *Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation*, içinde GECCO '14. New York, NY, USA: Association for Computing Machinery, Tem. 2014, ss. 879-886. doi: 10.1145/2576768.2598291.

[31] S. Wagner *vd.*, "Architecture and Design of the HeuristicLab Optimization Environment", içinde *Advanced Methods and Applications in Computational Intelligence*, R. Klempous, J. Nikodem, W. Jacak, ve Z. Chaczko, Ed., içinde Topics in Intelligent Engineering and Informatics. , Heidelberg: Springer International Publishing, 2014, ss. 197-261. doi: 10.1007/978-3-319-01436-4_10.

[32] J. H. Steiger, "Tests for comparing elements of a correlation matrix", *Psychol. Bull.*, c. 87, sy 2, ss. 245-251, 1980, doi: 10.1037/0033-2909.87.2.245.

[33] Çelebi, S. B., & Fidan, Ş. RNN-Based Time Series Analysis for Wind Turbine Energy Forecasting. *International Journal of Engineering and Innovative Research*, *6*(1), 15-28.

[34] Çelebi, S. B., & Karaman, Ö. A. Multilayer LSTM Model for Wind Power Estimation in the Scada System. *European Journal of Technique (EJT)*, *13*(2), 116-122.

BIOGRAPHIES

Sehmus Fidan, completed his Bachelor of Science in Electrical Teaching at Gazi University in 2006. He then obtained his Master of Science degree from Afyon Kocatepe University in 2010. Continuing his academic journey, he earned his PhD from Fırat University in 2018. Since 2019, he has been serving as a Doctoral Lecturer at Batman University. Throughout his professional career, Dr. Fidan has been involved in numerous projects, enriching his experience in the field. His research interests include the Internet of Things, machine learning, and meta-heuristic algorithms, in which he has conducted extensive studies. His work in these areas has contributed significantly to the academic community and the broader field of electrical engineering.

European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.14, No.1, 2024

Research Article

Design and Performance Analysis of Single-Phase Squirrel-Cage Induction Machine with High Efficiency

Serhat Akşun[1*], Sezai Taşkın[2], Ali Bakbak[3]

[1] Volt Elektrik Motor Company, Kazım Karabekir Cad. No:84 Kemalpaşa-Izmir, Türkiye (e-mail: serhat.aksun@voltmotor.com.tr).
[2*3] Manisa Celal Bayar University, Electrical and Electronics Engineering Department, 45140, Yunusemre, Manisa, Türkiye.
(e-mail: sezai.taskin@cbu.edu.tr), (ali.bakbak@cbu.edu.tr)

ARTICLE INFO

Received: March, 09, 2024
Revised: May., 15. 2024
Accepted: May, 15. 2024

Keywords:
Electrical machine design
Single phase induction motor
IE2 Efficiency
IEC 60034-2-1:2014

Corresponding author: *Sezai Taşkın*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1449620

ABSTRACT

This study presents an approach to designing single-phase squirrel-cage induction motors with a focus on achieving high efficiency. The objective is to improve the performance of the motors, particularly in meeting IE2 efficiency standards, without compromising on power or reliability. Specifically addressed is a single-phase squirrel-cage induction motor with a rated power of 0.55 kW. Through comprehensive design iterations involving dimensioning, slot-pole selection, and parametric optimization of stator and rotor geometry, the efficiency of the single-phase squirrel-cage induction motor has been improved from the current market design of 64.9% to a minimum of 77.1%, as defined in the IEC 60034-30 standard. Analysis includes key parameters such as slot opening, end-ring length, and rotor bar width, highlighting their significant influence on performance metrics. Experimental validation, conducted at an ISO/EN 17025 accredited R&D Laboratory, confirms the design improvements, achieving an efficiency of 80.1%. This research highlights the potential for significant energy savings across various applications by improving the efficiency of single-phase squirrel-cage induction motors to meet IE2 standards.

1. INTRODUCTION

The development of energy-efficient electric motors has great significance to address sustainability and energy conservation challenges. Most small power, generally below 2 kW, induction machines are operated with single-phase power supplies [1]. SPIMs are extensively used in appliances, small machinery, and domestic applications, such as compressors, air conditioners, heating-circulating pumps, fans, centrifugal pumps, vacuum cleaners, sewing, washing and milking machines, and so on [1]. Due to their wide range of applications, achieving high efficiency in SPIMs is high importance [2,3]. According to new energy efficiency regulations (EU) 2019/1781 and amending (EU)2021/341 starting 1st July 2023, the SPIMs must provide IE2 efficiency level in Europe [4]. Efficiency standards and regulations play a crucial role in driving the development of energy-efficient induction motors. The International Electrotechnical Commission (IEC) has established standards, including IE2 efficiency requirements, to encourage the adoption of motors with higher efficiency ratings. Compliance with these standards ensures that motors meet specific efficiency requirements, contributing to energy savings and reduced environmental impact.

In recent years, various research endeavors have focused on improving the performance of SPIM motors, with extensive analysis undertaken on aspects encompassing motor modeling, design, manufacturing, and control. Researchers have initially prioritized the development of more accurate models to assess SPIM losses, factoring in phenomena such as eddy currents and nonlinear effects such as magnetic saturation affecting motor components [5].

The focus of many research efforts has been on optimizing the configuration of the squirrel-cage rotor in electric motors. The geometry and layout of the rotor bars play a critical role in determining the motor's efficiency, power factor, and starting torque [6]. Various designs for rotor slots, including trapezoidal, oval, circular, rectangular, and pentagonal shapes, as well as different configurations of rotor bars such as open, semi-closed, and closed, with varying depths, have been thoroughly investigated [7]. Additionally, studies have examined the impact of factors such as the size of the rotor bar slot opening and the aspect ratio of the bar's width to its height. Despite advancements in rotor design methodologies, the process still heavily relies on the experience and intuition of the designer. Incorporating optimization algorithms and other artificial techniques into the design process has become increasingly necessary. The selection of the number of rotor bars, in conjunction with the combination of pole pairs and

stator slots, significantly influences the motor's efficiency. Furthermore, experimental materials have been utilized in the construction of parts for SPIMs, with aluminum rotors being replaced by die-cast copper ones. Among various manufacturing processes, die-casting has emerged as the most efficient method for mass-producing squirrel-cage components in SPIMs [8].

The characteristics of capacitors represent another significant factor. Researches have centered on several key areas such as determining the most suitable placement of the capacitor, selecting the optimal value of the capacitor, examining its influence on the phase shift between primary and auxiliary winding currents, and evaluating how this phase shift affects the quality of the motor's torque. Increasing capacitance boosts starting torque, although at the expense of motor efficiency beyond a certain point. Consequently, determining the ideal value for the run capacitor is important.

Factors such as poor power factor, low starting torque, and unbalanced magnetic fields pose challenges in achieving high efficiency. Enhanced winding designs, such as the split-phase, capacitor start, and capacitor run configurations, have been investigated to optimize performance. These designs help mitigate power factor issues and increase motor efficiency by improving the distribution of magnetic fields [9]. Power factor correction methods have also been explored to enhance the efficiency of SPIMs. Techniques such as capacitor banks, reactive power compensation, and active power factor correction circuits have been employed to improve power factor and reduce energy losses [10].

In the manufacturing process, the optimal design for SPIM is presented in [11], aiming to obtain the dimensions and electrical parameters of a SPIM. In addition, [12] delves into a study focusing on a single-phase permanent slip capacitor induction motor for a washing machine with the parameters of 373 W, 4 pole, 220 V, 50 Hz.

This study represents a significant contribution to the field of single-phase squirrel-cage induction motors. By focusing on the design, analysis, and performance testing of a 0.55 kW 4P SPIM with high efficiency (IE2), our research fills a crucial gap in current knowledge and industry practice. Utilizing Siemens Simcenter SPEED software for motor design, notable improvements in efficiency have been achieved, enhancing the SPIM's performance from the current market standard of 64.9% to a minimum of 77.1%, as defined in the IEC 60034-30 standard. This advancement not only aligns with global energy efficiency regulations but also offers tangible benefits to industries and households worldwide. A comprehensive layout of the design process, test methods, and systems is provided in this paper, offering valuable insights for researchers, engineers, and practitioners in the field. Meticulous attention to detail in each section, from design steps to prototyping and performance testing, aims to set a new standard for SPIM efficiency and contribute to the ongoing evolution of sustainable motor technologies.

2. ELECTROMAGNETIC DESIGN

A SPIM can be classified into different types based on its starting mechanism, namely (i) split phase, (ii) capacitor start, (iii) shaded pole, and (iv) repulsion type [13]. This study focuses on examining the characteristics of capacitor start motors. The single-phase capacitor start induction motor is classified into three specific types as follows.

(i) Two-value capacitor motor
(ii) Permanent-split capacitor motor
(iii) Capacitor-start induction run motor

In this study, a permanent-split capacitor motor is considered. The design algorithm of the motor is given in Figure 1. The operations in each step of the algorithm have been explained in detail in the following sections. Performance calculations, including electromagnetic analysis, have been conducted using FEA. The prototype of the designed motor was produced and tested.

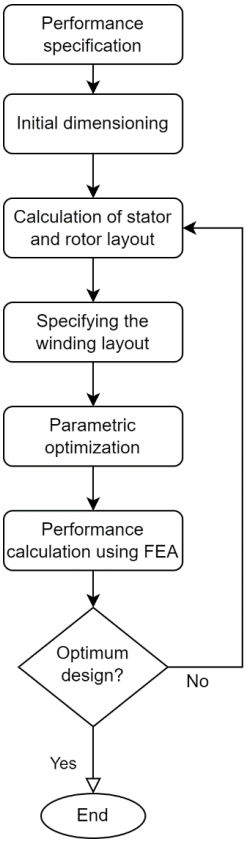


Figure 1. Design algorithm

2.1. Performance specifications

The specified motor is intended for use in milking machines. In the field, most milking machines use motors with the following specifications 550W, 4 poles (4P), 230V, 50Hz. Other requirements for milking machine motors include:

(i) Moderate starting torque: 0.3-0.6 times the nominal torque.
(ii) Low starting current: 4-5 times the nominal current.
(iii) High breakdown torque: 2-3 times the nominal torque.
(iv) Max. Capacitor Voltage: 400V.
(v) Nominal efficiency is 77.1% and minimum efficiency is 73.7% according to IEC 60034-30.

To prevent excessive temperature rise, current densities should be kept below 7 A/mm² for aux winding and 4 A/mm² for main winding. Temperature rise up is maximum 80°C for aux winding and main winding.

The design targets specified in Table I. In the table; *ILr* is the starting current, *In* is the nominal current, *TLr* is the starting torque, *TBr* is the breakdown torque and *Tn* is the nominal torque. The two target values specified for some parameters, such as speed and efficiency, represent a range within which

the parameter value is acceptable. This range accounts for variations that may occur due to different operating conditions or manufacturing tolerances. Therefore, if the parameter value falls within this range, it is considered satisfactory for our design targets.

TABLE I
DESIGN TARGETS

Parameter	Unit	Target value
Power	*W*	550
Speed	*rpm*	1400-1450
Torque	*Nm*	3.66
Voltage	*V*	220
Frequency	*Hz*	50
Efficiency	-	73.7-77.1
ILr/In	-	4-5
TLr/Tn	-	0.3-0.6
TBr/Tn	-	2-3
Capacitor voltage	*V*	<400
Current density of main winding	-	<4
Current density of aux winding	-	<7

2.2. Dimensioning

2.2.1 Stator Outer and Inner diameter

The machine utilization factor (*Cu*) is used to determine the stator's outer diameter and stack length, calculated according to Equation 1.

$$C_u = D_o^2.L \quad (1)$$

where *Do* is the outer diameter, and *L* is the calculated stack length.

Figure 2 shows the utilization factor (*Cu*) specifically for 4-pole (4P) induction machines.

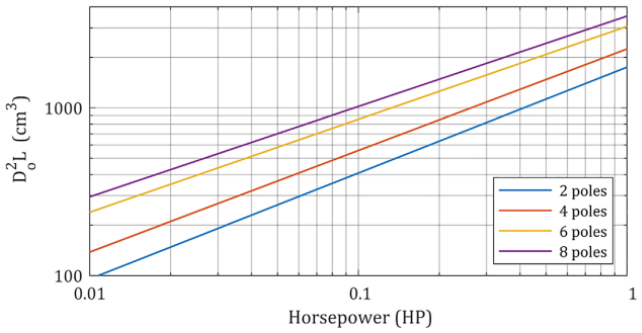


Figure 2. Machine utilization factor [7]

The optimal ratio for the inner diameter to outer diameter (*Di/Do*) is typically specified as given in Table II [1].

TABLE II
STATOR INNER/OUTER DIAMETER RATIO

2P	2	4	6	8
$\frac{D_i}{D_o}$	0.54-0.58	0.61-0.63	0.68-0.71	0.72-0.74

2.2.2 Stator and rotor slot combination

The combination of stator and rotor slot numbers plays a crucial role in motor operation. Determining the right combination is essential to avoid various issues during the motor operation such as electromagnetic noise, parasitic torque in the torque-speed curve and low efficiency.

In the case of 4-pole (4P) motor, a stator slot number of 36 is selected. For this stator slot number, different rotor slot numbers are investigated according to reference [1]. To mitigate magnetic noise and parasitic torque, a rotor slot number of 28 is determined as the optimal configuration. Failing to select the appropriate slot combination may lead to the following issues: electromagnetic noise, synchronous torque problems in the torque-speed curve, and reduced efficiency. To address these concerns, it is recommended to refer to Table III for the stator-rotor slot combination.

TABLE III
STATOR ROTOR SLOT NUMBER COMBINATION

2P	Stator slot number (Ns)	Rotor slot number (Nr)
2	24	18, 20, 22, 28, 30, 33, 34
	36	25, 27, 28, 29, 30, 43
	48	30, 37, 39, 40, 41
4	24	16, 18, 20, 30, 33, 34, 35, 36
	36	**28**, 30, 32, 34, 45, 48
	48	36, 40, 44, 57, 59
	72	42, 48, 54, 56, 60, 61, 62, 68, 76
6	36	20, 22, 28, 44, 47, 49
	54	34, 36, 38, 40, 44, 46
	72	44, 46, 50, 60, 61, 62, 82, 83

2.2.3 Air gap and dimensioning stator and rotor slots

The air gap (*g*) dimension for 4-pole (4P) motor is calculated using Equation 2 [1].

$$g = \left(0{,}1 + 0{,}012.\sqrt[3]{Pn}\right) \quad (2)$$

For the specified motor, the air gap (*g*) is found as 0.2 mm.

The dimensioning of stator slots in the motor design is taken into account with the following important parameters:

(i) Production Capability: The design should consider the practical constraints and capabilities of the production process.

(ii) Magnetic Flux Density of Stator Yoke (at *Hsty*): The magnetic flux density in the stator yoke should be controlled to limit iron losses. A maximum preferred value of 1.6 Tesla (T) is considered.

(iii) Magnetic Flux Density of Stator Tooth Width: The magnetic flux density in the stator tooth width is another crucial factor to optimize for efficient motor performance.

To determine optimum values for stator rotor slots, parametric analysis is used. Optimum values are given in Table IV. The dimensions represented by the parameters in the Table IV are depicted in Figure 3.

TABLE IV
STATOR PARAMETRIC DIMENSION

Parameters	Dimensioning Boundaries (mm)	Optimum value (mm)
Hsty	9-10	9.5
Wstt	3-4	3.38
Hso	0.5-1	0.75
Sso	2-3	2.23
Hrts	14.5-15.5	14.79
Wrtt	4-5	4.79
Rs	0-0.5	0.2
Sro	0.5-1.5	1

The dimensions are carefully determined to attain the best possible motor performance and efficiency. Considering the selected stator and rotor slot combinations, as well as the dimensioning requirements, the lamination design is successfully finalized. The specific design of the lamination is given in Figure 3.

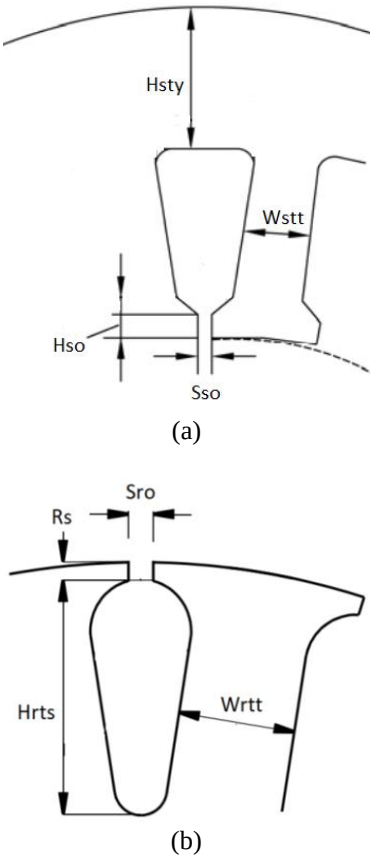


Figure 3. Slot parameters of (a) stator and (b) rotor.

In parametric optimization, one of the key parameters is the height of the rotor bars. Although increasing this height also increases the starting torque, it reduces efficiency. Therefore, the *Hrts* value that provides the required starting torque with the highest efficiency has been selected. The impact of the *Hrts* value on efficiency and starting torque is shown in Figure 4.

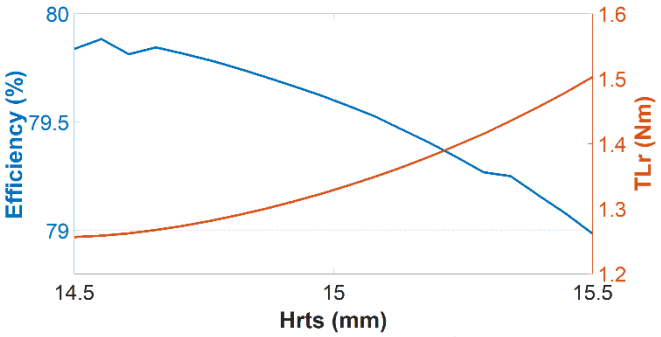


Figure 4. Parametrization of Hrts

These dimensions are established with the aim of ensuring proper casting in the rotor slots, while simultaneously achieving desirable performance outcomes in terms of speed, efficiency, and the torque-speed curve.

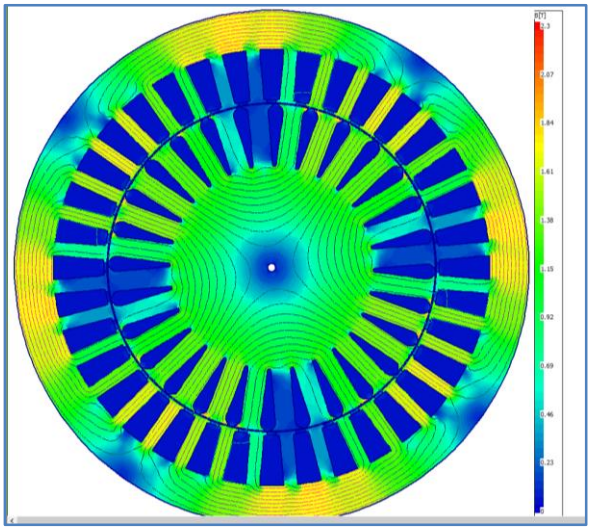


Figure 5. Magnetic flux density distribution of the machine

The Finite Element Method (FEM) analysis yielded magnetic flux density results, which are presented in Figure 5. The analysis indicates that there is no saturation observed in the steel material. Table V presents the magnetic flux density values for specific points."

2.2.4 Winding Design

In the winding design phase, the number of turns per pole, wire diameters, and the winding scheme are determined based on the following criteria:

(i) Number of Turns (*Nm*) Calculation: The number of turns in the main winding is determined based on the electromagnetic force (EMF) using the Equation 3.

$$E_m = \pi.\sqrt{2}.\Phi_m.Nm.K_{wm}.f \quad (3)$$

where K_{wm} represents the winding factor and f is the frequency.

(ii) Wire Diameter Determination: Wire diameters are chosen considering factors such as current densities, copper losses, and slot fill factor. The slot fill factor area should not

exceed 0.45 to ensure manufacturability. The wire diameter is calculated using Equation 4.

$$J = \frac{I}{\frac{\Pi . d^2}{4}} \quad (4)$$

where J represents the current density, I is the current, and d is the wire diameter. Based on the criteria, the following specifications are determined:

- Number of turns in the main winding (*N_m*): 15-18-50-50.
- Number of turns in the auxiliary winding (*N_aux*): 45-49-68 / 45-49.
- Wire diameter of the main winding: 0.8mm
- Wire diameter of the auxiliary winding: 0.7mm

The distributed winding scheme, as depicted in Figure 6, is utilized in the motor design.

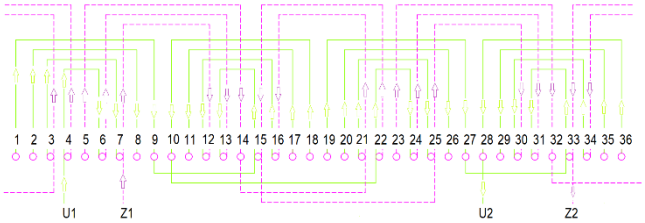


Figure 6. Winding scheme

3. ELECTROMAGNETIC ANALYSIS

The design parameters are given in Table V. The torque-speed curve for the motor design is shown in Figure 7. This curve shows the relationship between the motor's torque output and its rotational speed. By examining the torque-speed curve, important information for the motor performance characteristics can be obtained. As seen in Figure 7, the starting torque is 1.3 Nm, and maximum torque is 6.7 Nm at the 1285 rpm.

The torque-speed curve demonstrates that there is no presence of synchronous torque. This indicates that the determined combination of stator and rotor slot numbers is appropriate.

TABLE V
ANALYSIS RESULTS

Parameter	Unit	Result
Current	*A*	3.2
Speed	*rpm*	1433
Rated torque	*Nm*	3.66
Rated power	*W*	550
Efficiency		78.9
Locked rotor torque p.u		0.4
Locked rotor current p.u		3.8
Breakdown torque p.u		2.14
Capacitor voltage	*V*	330
Stator yoke magnetic flux	*T*	1.59
Stator tooth magnetic flux	*T*	1.53
Rotor tooth magnetic flux	*T*	1.4

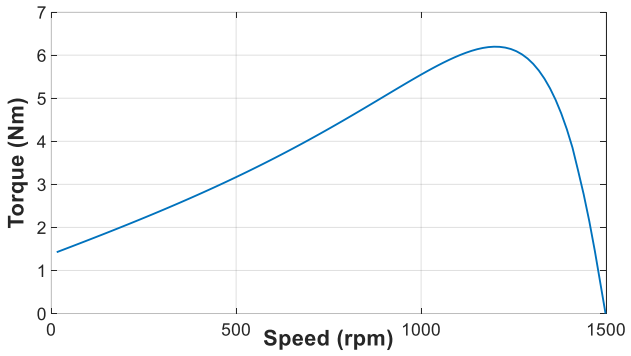


Figure 7. Torque-speed curve

The temporal torque characteristic is given in Figure 8, showing a well-balanced graph with no detectable harmonic effects. The FEM results and analytical analysis both confirm that the design targets have been successfully achieved.

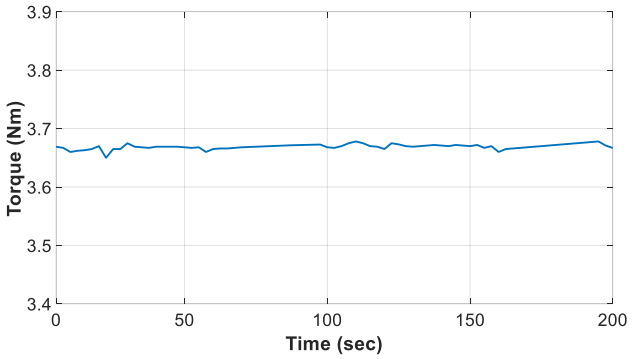


Figure 8. Torque profile over time

4. EXPERIMENTAL RESULTS

During the prototyping phase, it is crucial to ensure that semi-finished products, such as shafts and housings, meet the specified tolerances. Deviations from the required tolerances can result in an unstable air gap. An unstable air gap can lead to increased losses and harmonics, the occurrence of synchronous torque on the torque-speed curve, and the potential for magnetic sound generation. Therefore, meticulous attention should be given to maintaining the desired tolerances to ensure a rigid air gap and mitigate these adverse effects. The prototyped motor is shown in Figure 9.

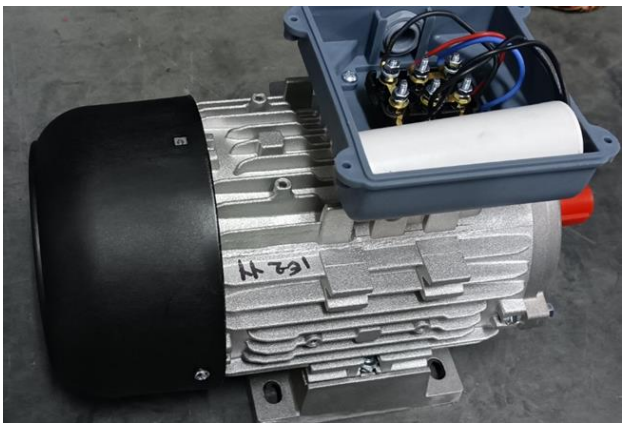


Figure 9. The prototyped motor

For the motor testing process, a torque-controlled dynamometer is used. This dynamometer incorporates a power analyzer, torque sensor, and speed sensor. The power analyzer provides the measurement of current, frequency, voltage, and power factor. The torque and speed sensor allow for the measurement of torque and speed, respectively. Experimental set-up of the system is shown in Figure 10 and 11.

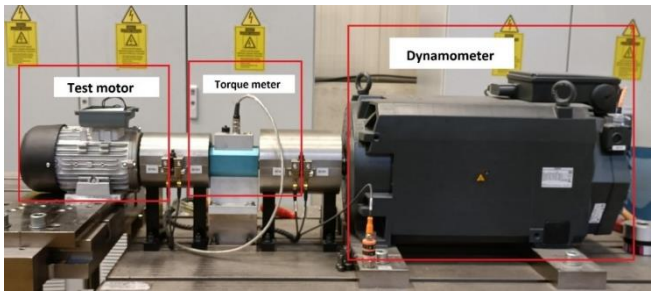


Figure 10. Test system

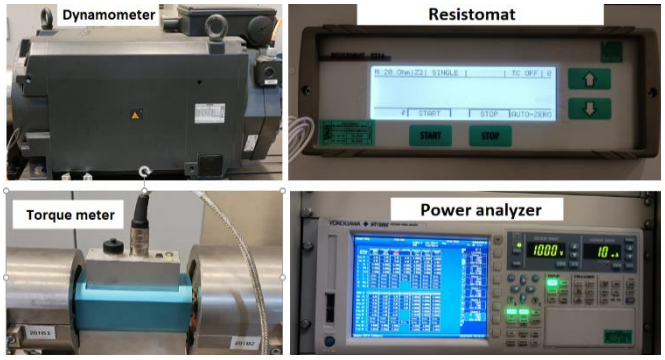


Figure 11. Other components of the test system

By utilizing the values obtained from the power analyzer, electrical power can be calculated, while the values acquired from the torque and speed sensor enable the determination of mechanical power. The combination of electrical and mechanical power facilitates the calculation of motor efficiency. It is essential to ensure proper alignment and coupling between the motor and dynamometer during testing. Aligning them in the same line is crucial to obtain reliable test results and ensure that the measured efficiency corresponds to the nominal value. The test procedure for efficiency measurement is outlined as follows [14]:

(i) Couple the motor to the dynamometer.

(ii) Apply a load to the motor up to its rated power.

(iii) Calculate the efficiency.

To measure locked rotor current, locked rotor torque, and breakdown torque, the following test procedure is specified [10]:

(i) Couple the motor to the dynamometer.

(ii) Initiate the locked rotor condition.

(iii) Measure the torque and current.

Following these testing procedures ensures accurate assessment of motor performance, including efficiency, locked rotor characteristics, and breakdown torque. All the tests are conducted in accordance with the IEC 60034-2-1 standard, which outlines three efficiency determination methods:

Method 2-1-1A: This method is applicable to all single-phase machines.

Method 2-1-1B: This method is used for three-phase machines with a rated output power of up to 2 MW.

Method 2-1-1C: This method is employed for three-phase machines with a rated output power exceeding 2 MW.

In this study, Method 2-1-1A is utilized to determine the efficiency of the motor. According to this method, the motor is loaded to its rated power until thermal equilibrium is achieved, with a rate of change of 1 K or less per half hour. Measurements of voltage, current, electrical power, speed (n), torque (T), and temperatures are saved during the test. The temperature measurement graph for full-load operation starting at 30 degrees Celsius is given in Figure 12. The temperatures reached by the motor in a steady state are within the permitted limits.

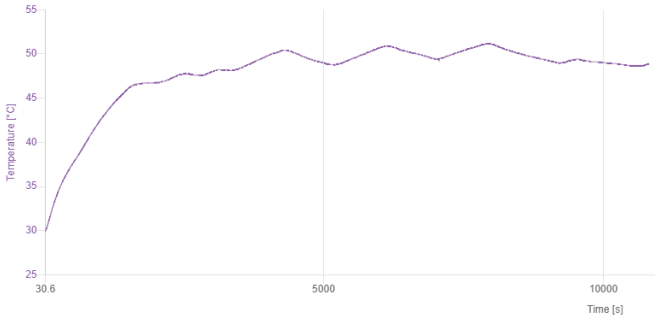


Figure 12. Temperature rise up of the motor

The test results are compared with the analysis results in Table VI. The results show that the design targets are successfully achieved based on the comparison of these results.

TABLE VI
COMPARING OF THE ANALYSIS AND TEST RESULTS

Parameter	Unit	Analysis Result	Test Result
Current	*A*	3.2	3.3
Speed	*rpm*	1433	1440
Rated torque	*Nm*	3.66	3.66
Rated power	*W*	550	550
Efficiency	-	77.8	80.1
Locked rotor torque p.u	-	0.4	0.5
Locked rotor current p.u	-	3.8	3.9
Breakdown torque p.u	-	2.14	2.1
Capacitor voltage	*V*	330	360
Stator yoke magnetic flux	*T*	1.59	Design value
Stator tooth magnetic flux	*T*	1.53	
Stator tooth magnetic flux	*T*	1.4	

5. CONCLUSION

In this study, the development and performance testing of a 0.55 kW single-phase induction motor with high efficiency in accordance with the IE2 class requirements are successfully accomplished. The design process involved proper dimensioning, selection of the slot-pole combination, parametric optimization of the stator and rotor geometry, and

rigorous finite element analyses to validate the design. By analyzing the impact of parameters such as slot opening, end-ring length, and rotor bar width, we identify the specific design enhancements responsible for optimizing the performance of the SPIM. Additionally, the significance of these modifications in aligning the SPIM with IE2 efficiency standards and maximizing energy savings across various industrial and residential applications is discussed.

The analysis results demonstrated that the designed SPIM achieved notable performance specifications, including a power factor of 0.98 and an efficiency of 77.8% at full load. Subsequently, the prototyped motor underwent comprehensive testing according to the IEC 60034-2-1:2014 standard methods for determining losses and efficiency. The efficiency of the SPIM, based on the measurement results obtained from testing in an accredited R&D laboratory compliant with ISO/EN 17025 standards, is calculated as 80.1%.

The successful realization of the SPIM design with enhanced efficiency not only aligns with the new regulations (EU)2019/1781 and (EU)2021/341 but also contributes to overall energy savings. By attaining high efficiency, SPIMs in the specified power range (0.12 kW ≤ PN ≤ 0wi.75 kW) can now meet the IE2 efficiency class requirements, enabling improved energy utilization in various industrial and household applications.

The findings contribute to the advancement of energy-efficient motor technology and have practical implications for reducing energy consumption and promoting sustainability in various industrial and household applications.

REFERENCES

[1] I. D. Chasiotis and Y. L. Karnavas, “On the design and manufacturing of small single phase induction motors toward super premium efficiency standards,” presented at the 2020 Int. Conf. on Electrical Machines (ICEM), Gothenburg, Sweden, pp. 2321-2327, 2020.

[2] Y. L. Karnavas, I. D. Chasiotis, “Design and manufacturing of a single-phase induction motor:A decision aid tool approach”, Int. Tran Electr Energ Syst. 27:e2357, 2017.

[3] D. Liang, V. Zhou, "Recent market and technical trends in copper rotor for high-efficiency induction motors", in Proc. of 2018 Int. Power Electronic Conf. (IPEC-Niigata 2018- ECCE Asia), Niigata, Japan, May 20-24, 2018.

[4] Udroiu, Nicoleta-Alina, and Carmen Georgeta Nicolae. "POSSIBILITIES TO REDUCE CO 2 EMISSIONS BY USING ELECTRIC MOTORS WITH HIGH ENERGY EFFICIENCY." Scientific Papers. Series D. Animal Science 65.1 (2022).

[5] T. Lubin, S. Mezani and A. Rezzoug, "Analytic Calculation of Eddy Currents in the Slots of Electrical Machines: Application to Cage Rotor Induction Motors," IEEE Transactions on Magnetics, vol. 47, no. 11, pp. 4650-4659, Nov. 2011.

[6] I. Chasiotis, Y. Karnavas, F. Scuiller, “Effect of rotor bars shape on the single-phase induction motors performance: An analysis toward their efficiency improvement”, Energies, 10.3390/en15030717, 15, 3, (717), 2022.

[7] Gui-Yu Zhou and Jian-Xin Shen, "Current harmonics in induction machine with closed-slot rotor," 2015 IEEE 2nd Int. Future Energy Electronics Conf. (IFEEC), Taipei, Taiwan, 2015.

[8] Lee KS, Ho Lee S, Park JH, Kim JM, Choi JY. “Experimental and analytical study of single-phase squirrel-cage induction motor consideringend-ring porosity rate,” IEEE Trans Magnet. 53(11) pp. 1-4. 2017.

[9] D. Sharma, R. K. Saini, and S. Pathania, “Design, analysis and selection of electric and magnetic loading for different rating of squirrel cage induction motors by using matlab gui software,” Int. J. Innovative Res. Elec. Electron. Instrum. Control Eng., vol. 3, no. 4, pp. 20-22, Apr. 2015.

[10] S.H. Choudhury, M.A. Uddin, N. Hasan, M. S. Alam, M. F. I. Bashar, “Impact of skin effect for the design of a squirrel cage induction motor on its starting performances”, International Journal of Engineering Science and Technology (IJEST), vol. 4, no.01 Jan. 2012. pp:362-366.

[11] C. Mademlis, I. Kioskeridis, and T.Theodoulidis, “Optimization of single-phase induction motors-Part I: Maximum energy efficiency control,” IEEE Trans. Energy Convers., vol. 20, no. 1, pp. 187–195, Mar. 2005.

[12] S. Shekhar, S. Kumar, S.K. Anand, S.K. Duve, P.D. Bharadwaj, “Design and programming for single phase induction motor”, International Journal of Engineering Science and Computing, April 2016.

[13] T.Z. Aung , E.E Cho, “Design calculation of single-phase permanent slip capacitor induction motor used in washing machine”, IRE Journals, vol. 3, no. 2, Aug 2019, pp:206-213.

[14] M. Farasat, A. M. Trzynadlowski, M. S. Fadali, “Efficiency improved sensorless control scheme for electric vehicle induction motors”, IET Electr. Syst. Transp., vol. 4, no. 4, pp. 122–131, 2014.

ACKNOWLEDGEMENT

We would like to thank Volt Elektrik Motor Company for providing laboratory facilities for the design, fabrication, and testing of the electric motor presented in this study.

BIOGRAPHIES

Serhat Aksun received the B.Sc. degree in electrical engineering from Kocaeli University in 2013. He received MSc. degree in Electrical&Electronics Engineering from Manisa Celal Bayar University in 2023. He works for the Volt Elektrik Motor Company as electrical design engineer.

Sezai Taskin received the B.Sc., M.Sc., and Ph.D. degrees from Marmara University, Istanbul, Türkiye, in 1999, 2001 and 2007, respectively. He worked as a Research Assistant at the Marmara University during his graduate studies. He joined the Electrical&Electronics Engineering Department at Manisa Celal Bayar University in 2012. His research interests include control system applications, solar PV systems, battery energy storage systems, and the smart operation of future power systems.

Ali Bakbak received the B.Sc. degree in electrical and electronics engineering from Niğde University, Niğde, Türkiye, in 2012, the M.Sc. degree in electrical and electronics engineering from Celal Bayar University, Manisa, Türkiye, in 2016, and the Ph.D. degree in electrical and electronics engineering from Ege University, Izmir, Türkiye, in 2021. He is currently a Faculty Member with Manisa Celal Bayar University, Manisa. His research interests include modeling, design, and control of electrical machines.

European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.14, No.1, 2024

Research Article

Design of Sesame Peeling Machine and Performance Analysis with Threshold-Based Image Segmentation Technique

Ahmet Gündoğdu[1*], Emrullah Acar[2] and Reşat Çelikel[3]

[1*]University of Batman, Electrical and Electronic Engineering Department, Batman, Türkiye. (ahmet.gundogdu@batman.edu.tr)
[2] University of Batman, Electrical and Electronic Engineering Department, Batman, Türkiye. (emrullah.acar@batman.edu.tr)
[3] University of Firat, Mechatronic Engineering Department, Elazığ, Türkiye. (rcelikel@firat.edu.tr)

ARTICLE INFO

Received: Mar., 074. 2024
Revised: June., 01. 2024
Accepted: June, 07. 2024

Keywords:
Sesame
Sesame Peeling Machine
Machine Design
Threshold-based Image Segmentation

Corresponding author: *Ahmet Gündoğdu*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1448580

ABSTRACT

In this study, peeled sesame seeds were analyzed using a threshold-based image segmentation technique. Peeled sesame seeds were obtained from a sesame membrane peeling machine whose design and production was carried out within the project's scope. By applying the image segmentation technique to the samples taken from the machine, the machine's sesame peeling performance and sesame peeling percentage were determined. Automatic calculation of the peeled sesame weights in the image is presented by the threshold-based image segmentation technique, which is one of the basic operations of digital image processing and is extensively employed in many application areas, including extraction, detection, and identification of features. This technique aims to divide an image into meaningful regions according to a particular operation. Based on the results obtained using the threshold-based image segmentation technique, the success achieved in the peeling process is 48%, 56%, 60% for periods of 60 min, 120 min, 180 min respectively. By using the designed machine, the processing time of peeling sesame has been reduced by 75%, water consumption has been reduced by 90%, and salt use has been eliminated. Moreover, operating, personnel, and energy costs have been minimized.

1. INTRODUCTION

Sesame seeds, whose motherland is India, Indonesia, Nigeria, China and Burma in the world, are found in yellow, white, red and black colours, also have geometrical dimensions such as 2,26-3,01 mm in height, 1,55-1,86 mm in width and 0,75-0,97 mm in thickness [1,2]. Sesame, which is used widely in dough products such as fat, tahini, halvah, bread, pastries, pies, is an oily seed that contains minerals such as Ca, Fe, Mn, Zn, Mg, Cu, and Se, as well as fat, carbohydrate and protein with high nutritive properties [3]. However, it also contains substances that are harmful to human health, especially in the thin membrane surrounding the grains, as well as oxalic acid, phytic acid and cellulose. Sesame membrane contains a large amount of oxalic acid, phytic acid and cellulose, which form a structure that disrupts human health by reacting chemically with Ca, Mg, Zn and Fe in sesame grains [2]. Moreover, the membrane on sesame seed has undesirable outcomes in the food industry like the dark colour and sour flavour. Therefore, before the sesame is consumed, it is a necessity to peeling of the membrane on the seed [1].

Sesame seeds contain a thin membrane around them. Approximately 15-20% of sesame seeds consist of this thin membrane. The ratio of oxalic acid, which is approximately 2.53% in a sesame seed, in the membrane can reach 15%. Oxalic acid combines with calcium in the environment to form oxalate, which can cause health problems such as kidney stones. Phytic acid and Cellulose ratio in the membrane is around 5% separately. These substances can cause digestive problems by combining magnesium, zinc, calcium, and iron found in sesame [2, 4]. In order to purify the sesame of these harmful active substances and make them suitable for consumption in the food industry, the membrane must be peeled. The amount of oxalic acid in a peeled sesame seed can decrease up to 0.25%.

In the food industry, the membranes of sesame seeds are peeled by mechanical methods, using different chemical disinfectants and solvent salts. Disinfectants and salts, used in the peeling processes of the sesame membrane, cause environmental pollution and require high water consumption. Therefore, environmental pollution and a large amount of water consumption establish a crucial problem. In order to reduce water consumption and shortened the process duration, sesame grains are kept in salt and chemical disinfectants such as NaClO and H_2O_2 before the mechanical peeling process [1]. In another decorticating process, the sesame grains were

kept at 60°C for 5 to 10 hours in NaOH and Na_2CO_3 saline solutions and then subjected to the mechanical stirring. After the stirring process, the sesame seed from the saline solution was dried at 120°C for 2 hours. Thus the peeling processing time can be reduced to 40 minute [5]. In another method, while sesame seeds are in the water, they simultaneously stir with rotating the tank, driving by the engine. After the stirring process, the sesame grains were taken into the boiler and transferred to the saline solution with higher density. The peeled grains from the saline solution were washed in another medium and then dried [6].

In order to peel the sesame membrane, the mixer rods in A, I and L profiles were fixed on the rotating shaft. Moreover, during the mixing process, the sharp edges of the bars are rounded to prevent the breakage and deformation of the sesame seeds. Better results were obtained in the peeling process with L-shaped stirrer bars at 500 rpm rotating speed [7]. In another method, sesame grains, which are water in the sesame grains and put into the water heated simultaneously with the electrical resistance, were kept at a temperature of 60ºC-70ºC for 1 to 1.5 hours. During the holding process, the moistened sesame grains were stirred with the rotational speed of 1800 rpm in the acidic bath. Then the sesame grains were separated from the membrane with the effect of the centrifugal movement inside the sieve [7]. In the existing sesame membrane peeling methods, chemical disinfectants and salts, causing environmental pollution, are used. Furthermore, large amounts of water are consumed during the solution waiting and subsequent washing.

Programmable Logic Controller-PLC is widely used in many industrial applications such as driving AC servo motors used in robotic manipulators [8], automatic liquid filling systems [9], solar power generation systems [10], hydraulic positioning systems [11], and machine automation. In [12], the liquid level control process was carried out automatically without human observation and using a buoy. A neural network-based PID controller has been designed in [13]. Speed control of an asynchronous motor is realized by using this controller and PLC together. An electrical energy generation system has been established in [14] by using a thermoelectric generator. The current and voltage values of this system are monitored instantaneously by using a PLC-SCADA system. In [15], the control of a water pump has been experimentally provided using a PLC. The most important advantage of this study is that the wireless communication application has been implemented in the system controlled by PLC. Thus, observation and control operations could be made farther from where the water pump is located. In [16], a system with a Stirling engine (providing thermal energy), electric generator, and secondary heat exchanger were controlled by an automation system using PLC in a solar energy conversion system. This automation system aims to ensure the operating safety of the system and monitor its parameters during operation. At the same time, the solar concentrator is positioned towards the sun to obtain maximum power. In [17], the engine oil production process was carried out using PLC automation. Instead of star-delta starting, a motor driver is used. Worker safety has been increased by taking high-security measures. Apart from that, many studies have been done on image processing in industrial applications. In a vegetable chopping automation, by using image processing, human labor was reduced, and various parameters related to vegetables such as texture, hardness, color, size, and shape were analyzed. Appropriate knife selection was made [18]. In another study, sesame oil quality was determined using Hyper Spectral Image-HSI analysis [19]. Industrial orange grading is done manually and using expensive technologies. A more efficient orange selection system has been developed in a very recent study using artificial intelligence and image processing [20].

Segmentation (border detection) is one of the most significant steps in image-processing stage because of that (a) it ensures major information for example: border shape, border disorder, and asymmetry (b) it helps extraction of features depending on the accuracy of the border detection [21]. Considering the literature studies, a large number of segmentation methods have been suggested to segment related object from the background. [22] and [23] employed the common and simple image segmentation method, which is the Otsu thresholding (adaptive thresholding). In segmentation, obtaining the outline of the object from the noisy image is one of the most important difficulties. To manage this problem, a technique called active contour is often used to outline an object from an image [24, 25]. The precision of the segmentation can be advanced by combining multiple segmentation techniques, which is also known as fusion based segmentation [26]. In addition, alternative segmentation methods are used for object segmentation, for example, neural network-based [27], edge-based [28] and cluster-based [29]. In this study, the threshold-based segmentation process was preferred, which gives extremely easy and fast results. After the segmented image of the object is obtained, the next step is to calculate the pixel-based area of the relevant object. The reason for calculating the segmented object area is that finding percentage of it in whole image. In literature, a few studies have been done on the calculation of the relevant object area by image segmentation methods. These studies are summarized respectively. [30] investigated the segmentation of the brain tumor using brain MR images and computed the area of the segmented tumor regions employing two algorithms: K-mean clustering and Fuzzy C-Mean algorithms. [21] suggested a new approach for skin lesion detection. Moreover, they computed the diameters of the segmented lesions. In this study, the properties (coordinates, area etc.) of the segmented image regions were calculated to determine the area of the peeled sesame in each image.

This study, it is aimed to test the performance of a sesame skin peeling machine designed and manufactured within the scope of the project numbered TUBITAK / SAN-TEZ. Thus, the peeled sesame samples obtained from the machine were analyzed using the threshold-based image segmentation technique. This study contributed to the literature on both machine design and automation, and image processing. Since machine design is not the subject of this article, detailed mechanical calculations are not included. The paper is divided into four sections. In section II machine design, machine automation and threshold-based image segmentation technique is presented. In section III, the simulation results obtained from the threshold-based image segmentation technique are given. Section IV describes conclusions.

2. MATERIALS AND METHOD

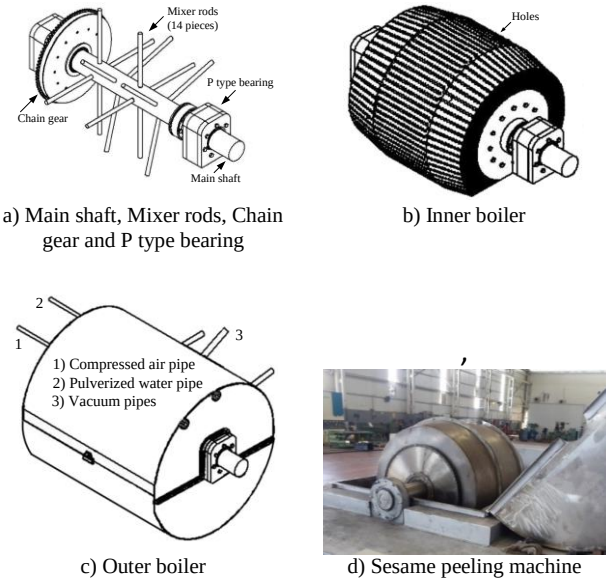
When looking through the literature, it is clear that the salty water method is commonly used to peel sesame skins. This procedure is carried out in two stages: For the first stage, 1 ton

of sesame is kept in 5 tons of water for 6-12 hours at a maximum temperature of 600°C. The sesame seeds softened in the first stage are preserved in a solution of 5 tons of water and 1 ton of salt for 6-12 hours in the second stage, after which they are washed with approximately 12 tons of water to desalinate, and finally dried and peeled off. Similarly, peeling sesame skins with different mechanisms using water and salt have been developed in the sources numbered [31, 32, 33]. The intensive use of inputs such as water and salt, high labor, time, energy, and operating costs, large area coverage, and wastes they leave in the environment are all drawbacks of these techniques. Furthermore, since sesame is a delicate and oily seed, it can be deformed by the skin peeling process's impact and friction. Fat, carbohydrate, protein, and other useful minerals and the peeled membranes are discarded as trash. A sesame skin peeling machine was developed and a prototype produced to remove these drawbacks and to be distinct from the brine method. Sections 2.1 and 2.2 describe the mechanical and electrical characteristics of the designed machine.

2.1. Machine Design

Fig.1 shows the main components of the sesame skin peeling machine, which includes 1) the main shaft, stirring rods, chain gear, and P-type bearings, 2) the sieve perforated inner vessel, and 3) the external boiler [34]. One of the most important machine components is the main shaft, as shown in Fig.1a. A 15 kW motor can drive it in both directions. It has a 100 mm diameter and is held in place by two P-type bearings. It comes with 14 detachable mixing rods. During rotation, these stirring rods are intended to avoid damaging sesame seeds. Also, the main shaft has a chain gear mechanism that rotates the sieve hole inner tank. P-type bearing bearings can support approximately 600 kg of weight, including the weight of the inner tank. The inner tank with the sieve hole is shown in Fig.1b as another key component of the machine. It can hold 500 kilograms of sesame seeds. It has a 560 mm radius, 1120 mm length, and a 1 mm wall thickness. It has 1 mm diameter holes with a 2 mm spacing between centers and a filling and discharges opening. With the assistance of hinges, this mouth can be opened and closed. A 15 kW motor powers the screen hole inner boiler. The inner bowl can rotate independently from the main shaft thanks to its chain gear assembly. Depending on demand, the main shaft and inner bowl may rotate in the same direction or the opposite direction.

Peeling sesame skins is easier when you rotate in the opposite direction. The outer boiler, shown in Fig.1c, has a pulverized water pipe, compressed air pipe, and vacuum pipe. External boiler; the perforated inner vessel also acts as a casing, keeping the main shaft and stirring rods safe from the elements. Evaporated water is sprayed from the pulverized water pipe to moisten the sesame seeds in the inner boiler at regular intervals. The peeled sesame membranes are vented using compressed air pipes, and the ventilated sesame membranes are transferred to the outside using vacuum pipes. An external compressor provides compressed air. Sesame membranes that have been vacuumed into the outside environment are preserved in a separate environment for various purposes. Fig.1d [34] shows a real image of a sesame skin peeling machine made of A304 stainless steel.



a) Main shaft, Mixer rods, Chain gear and P type bearing

b) Inner boiler

c) Outer boiler

d) Sesame peeling machine

Fig 1. Main parts of sesame peeling machine [34].

2.2. Machine Automation

The machine is operated by a PLC-based automation system, which was designed and manufactured. The main shaft carrying the mixing rods was driven by a 15 kW drive motor, while the inner bowl was driven by a 15 kW drive motor. Each motor is an asynchronous type motor with a Delta-VFD150E43A model industrial type motor driver controlling its speed. Direct torque control-DTC and field-oriented control-FOC are two advanced control techniques used by this driver. In practice, an open-loop control was conducted. The central processing unit is a DVP-14SS211R CPU, and the touch panel is a Delta-DOP-B07S410 model HMI panel. All control operations on the HMI panel are carried out by touch, thanks to software designed for machine automation. The HMI panel was used to adjust the motor rotation direction, and the speed settings were made on the drivers. Excessive currents and the dangers that can occur when the direction of rotation changes during idle and load are protected by software. Fig.2a depicts the machine's control and control panel, while Fig.2b depicts the established HMI panel interface.

a)Automation panel b)HMI user interface

Fig 2. Machine Automation.

2.3. Practise

The sesame skin peeling process, made with the machine manufactured, was carried out as in practice.

Practise : 400 kg of untreated dried sesame seeds were submerged in 280°C water for 3 hours. After the sesame seeds

had softened in the skin, they were loaded into the machine and peeled. The inner bowl with sesame seeds was rotated at 45 rpm during the peeling process, while the main shaft with stirring rods was rotated at 25 rpm. The cauldron's rotation period has been determined to be 180 minutes. Fig.3 shows the photographs taken every 60 minutes from peeled sesame seeds. An operator using a high-resolution camera captured these images. The image acquisition unit has not yet been integrated into machine automation because the machine is a prototype. However, during the mass production phase, the machine will be equipped with image acquisition and a processing unit. As a result, it is expected that the sesame skin peeling process will be improved in terms of energy and time.

During the application, the effect of the rotational movement separated the sesame seeds from their membranes after they had been moistened by soaking in water. Vacuuming was used to transfer the split sesame membranes to the outside world. The process was completed by removing the peeled sesame seeds from the discharge cover at the bottom of the inner boiler.

For peeling, 500 kg of sesame, 20 tons of water, and 1 ton of salt are used in the conventional saltwater peeling process. It will take about 24 hours to complete the task. The use of salt was eliminated with this machine's development, and only 1 ton of water was used to peel 500 kg of sesame in each application. The peeled sesame images in Fig.3 were subjected to the threshold-based image segmentation technique to test the machine's performance. In section *D*, the results of the study are reviewed.

Unprocessed Sesame | Image1-Processed Sesame for 60 min.

Image2-Processed Sesame for 120 min. | Image3-Processed Sesame for 180 min.

Fig 3. Machine-processed sesame images

2.4. Image Processing

2.4.1 Sesame Image Dataset

The dataset used in our study during the image-processing phase consists of 16 sesame images, which belong to three groups of images (Group-1: 60 minutes; Group-2:120 minutes and Group-3: 180 minutes). These images were obtained from the mobile phone cameras as shown in Fig.3.

2.4.2 Image Pre-Processing and Segmentation

In this stage, each sesame image was resized to 1/4 original image size. The resized images were then converted to 8-bit gray scale images via MATLAB-2016. After that, these images were turned into binary images based on thresholding. Region filling morphological algorithm were then applied to the thresholded binary images in order to fill image holes and regions.Later, image segmentation approach was implemented to the preprocessed sesame images to determine connected label components in the binary images. Lastly, the properties (coordinates, area etc.) of image regions were computed to find the collected area of the peeled sesame in the each image. These stages were considered in the following sections in details.

A) Image Thresholding

The threshold approach is the simplest method of image segmentation, which can be used to create a binary image from a gray scale image by selecting a single threshold (T) value. It is also used to divide the image into smaller sections using at least one color or gray scale value to determine its boundaries. The advantage of attaining a binary image is the simplification of the recognition and classification process, together with reducing the data complexity. In the threshold operation, input is generally a gray or color image *f(x, y)* and output is a segmented binary image *g(x, y)*, where black pixels match with the background and white pixels comply with the foreground (or vice versa). Moreover, a single criterion is implemented to all pixels in the image at the same time. The steps of threshold-based image segmentation are simply given in below:

- *The image is partitioned into pixel sets or continuous regions for segmentation.*
- *The partition is made according to intensity value of pixel sets.*
- *Then, the image is segmented as the background, g(x, y)=1, while the pixels of input image f(x, y) smaller than the threshold (T) value;*
- *Image is segmented as the foreground, g(x, y)=0, while the pixels of input image f(x, y) bigger than the threshold (T) value.*

$$g(x,y) = \begin{cases} 0 \; if \; f(x,y) < T \\ 1 \; if \; f(x,y) \geq T \end{cases} \quad (1)$$

B) Region Filling Morphological Algorithm

Region filling is one of the most frequently used morphological algorithms in image processing and it endeavor filling the image regions with specific colors. The algorithm is built on the complementation, dilations and intersection operators. In this algorithm, all non-boundary pixels are first categorized as ‘white’ and a value of black is then assigned to p. Region filling with ‘black’ is carried out by the following equation.

$$x_m = (x_{m-1} \oplus S) \cap I^C \qquad m = 1,2,3 \ldots \quad (2)$$

Here $x_0 = p$, S is the symmetric structuring matrix, $\oplus$ is dilation operator, $\cap$ is intersection operator, I is image matrix and I^C is the complement of I matrix [35].

2.4.3 Calculation of the Area of the Peeled Sesame

After pre-processing and segmentation stage, the pixels that form the outer boundaries of the peeled sesame were determined automatically and the area of these boundaries was calculated using Matlab R2016a commands such as 'regionprops'. This command calculates the object measurements like area, bounding box, centroid …etc. [36].

2.4.4 The Proposed Threshold-Based Image Segmentation Approach

The proposed approach consists of five stages as indicated in Fig.4.

1) A dataset consisting of 16 different sesame images was obtained from the mobile phone cameras over the machine we designed.
2) These images were resized to 1/4 original image size and turned into 8-bit gray scale images.
3) Resized gray scale images were converted to binary images based on thresholding.
4) Morphological algorithm was implemented to the binary images.
5) The weights of the peeled sesames were computed thanks to segmented areas.

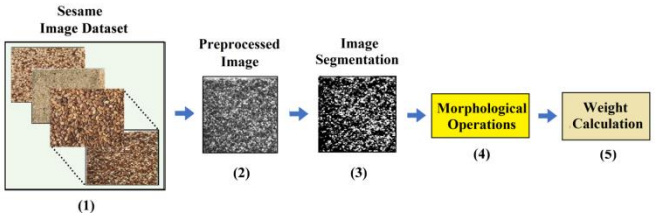


Fig 4. Proposed Threshold-Based Image Segmentation system

3. EXPERIMENTS RESULTS AND DISCUSSION

In this section, the attained 16 Sesame images (3 different Groups) were resized in 25% of original image size and they were converted to 8-bit resolution gray scale images. The original and resized 8-bit gray scale image samples for Group-1 are shown in Fig.5 (a) and (b). After that, segmentation of the peeled sesames was made based on thresholding, which was chosen according to the pixel value of peeled sesame in the image. Moreover, region filling morphological algorithm was applied to the segmented binary images as shown in Fig.5 (c). Finally, the approximate weights of the peeled sesames were computed by using the Equation-3 and tabulated in Table-1.

$$Weights(\%) = \left(\frac{Pixel\ areas\ of\ the\ segmented\ sesames}{Total\ pixel\ area}\right) \times 100 \qquad (3)$$

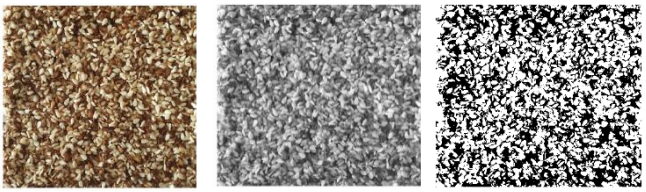
The same methodology used for Group-1 images above was implemented to the other groups (Group-2 and Group-3) and the output images for these groups were shown in Fig.6 and Fig.7, respectively. Moreover, the ratio of peeled sesame for each image groups was given in Table-2 and Table-3, respectively.

a) b) c)

Fig 5. The output images for a sample Group-1 image after 60 minutes: (a) Original Image (b) Resized Gray Scale Image (c) Segmented Binary Image based on Thresholding and morphological algorithm

TABLE I
THE APPROXIMATE WEIGHTS OF THE PEELED SESAMES FOR A SAMPLE IMAGE OF GROUP-1

Pixel areas of the Segmented Sesames	328862
Total Pixel areas	689500
Ratio of Peeled Sesame in Whole Image	**48 %**

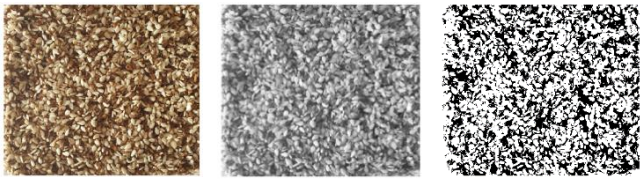


a) b) c)

Fig 6. The output images for a sample Group-2 image after 120 minutes:(a) Original Image (b) Resized Gray Scale Image (c) Segmented Binary Image based on Thresholding and morphological algorithm

TABLE II
THE APPROXIMATE WEIGHTS OF THE PEELED SESAMES FOR A SAMPLE IMAGE OF GROUP-2

Pixel areas of the Segmented Sesames	361239
Total Pixel areas	650655
Ratio of Peeled Sesame in Whole Image	**56 %**



a) b) c)

Fig 7. The output images for a sample Group-3 image after 180 minutes: (a) Original Image (b) Resized Gray Scale Image (c) Segmented Binary Image based on thresholding and morphological algorithm

TABLE III
THE APPROXIMATE WEIGHTS OF THE PEELED SESAMES FOR A SAMPLE IMAGE OF GROUP-3

Pixel areas of the Segmented Sesames	373978
Total Pixel areas	630591
Ratio of Peeled Sesame in Whole Image	**60 %**

Taking everything into account, it can be said that. The threshold-based image processing technique provides an advantage because it has a much easier structure compared to architectures such as U-Net, in terms of both time and processing complexity. However, in architectures such as U-Net, threshold-based image processing technique will be disadvantageous in this respect, as much better training can be achieved by marking the area to be segmented beforehand.

4. CONCLUSIONS

In this study, in order to test the performance of the sesame-peeling machine, which designed within the scope of the project before, the images of the peeled sesame samples obtained from the machine were analysed by threshold-based image segmentation technique. The proposed system consists of several stages. In the first stage, sesame images obtained from three different groups were pre-processed as well as these images were reduced in size and transformed into grayscale images. In the second stage, the pre-processed images were converted to the binary images thanks to threshold-based image segmentation approach. In the last stage, morphological operations were implemented to the binary images and the weights of the peeled sesames were calculated via segmented pixel areas. Considering the results obtained with the proposed threshold-based image segmentation technique, it was observed that a great success was achieved in peeling sesame grains.

When the results of the threshold-based image segmentation technique were investigated, the success achieved in the peeling process is 48%, 56%, 60% for periods of 60 min, 120 min, 180 min respectively. In terms of fictionalized processing time, these results are satisfactory. With this machine, the time it takes to peel sesame seeds is reduced by 75%, and the amount of water used is reduced by 90%, eliminating the use of salt. Furthermore, operating, staffing and energy costs have all been reduced. Image acquisition and processing unit will be mounted on the machine in future experiments. With various applications to be made, the machine's rotation time and rotation speed are optimized, aiming to improve the machine's performance. Reducing the amount of water used in the sesame peeling process may be a interesting research topic for other researchers.

ACKNOWLEDGEMENT

This study is supported by Ministry of Science, Industry and Technology/Turkey with 0905.STZ.2015 numbered project, titled "*Design and Manufacture of Technological Sesame Peeling Machine*". Thanks for support.

REFERENCES

[1] Carbonell-Barrachina AA, Lluch MA, Perez-Munera I, Hernando I, Castillo S (2009) Effect of chemical dehulling of sesame on color and microstructure. Food Sciences and Technology International 15(3): 229-234.

[2] Tunde-Akintunde TY, Oke MO, Akintunde BO (2012) Sesame Seed, Oilseeds, Dr. Uduak G. Akpan (Ed.), ISBN: 978-953-51-0665-4, InTech, Available from: http://www.intechopen.com/books/oilseeds/sesameseed

[3] Akbulut M (2008) Comparative studies of mineral contents of hulled sesame paste (tahin), unhulled sesame paste (bozkır tahin) and their blends. Asian Journal of Chemistry 20(3):1801-1805.

[4] Röbbelen G, Downey RK. Ashri A (1989) Oil Crops of the World-their Breeding & Utilization. Mc Graw-Hill Publishing Company.

[5] Yehia G, Moharram-Hussein OA, Osman IA, Yaldes IA, Abou-EI-Khier, Wet decortications of sesame seeds by new methods. United Nations University Website, http://unu.eduarchive.unu.edu/unupress/food/8F121e/8F121E09.htm

[6] Suzanne L (1988) Produce Peeling Machine Particularly for Shelling Nuts, (November 22 1988 4785729 USA patent).

[7] Lee KS, Noh HK (2015) Development of threshing machine for shatter-resistant sesame. Journal of Biosystems Engineering 40(2):110-114.

[8] Srinivas GL, Singh SP, Javed A (2020) Experimental evaluation of topologically optimized manipulator-link using PLC and HMI based control system. Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2020.08.023

[9] Baladhandabany D, Gowtham S, Kowsikkumar T, Gomathi P, Vijayasalini P (2015) PLC based automatic liquid filling system. International Journal of Computer Science and Mobile Computing 4(3):684-692.

[10] Prinsloo G, Dobson R, Schreve K (2014) Carbon footprint optimization as PLC control strategy in solar power system automation. Energy Procedia 49:2180-2190.

[11] Papoutsidakis M, Xatzopoulos A, Smyraiou GP, Tseles D (2017) PLC Programming Case Study for Hydraulic Positioning Systems Implementations. International Journal of Computer Applications, 167(12):49-53.

[12] Das R, Dutta S, Sarkar A, Samanta K (2013) Automation of tank level using Plc and establishment of Hmi by Scada. IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE) 7(2):61-67.

[13] Maraba V, Kuzucuoğlu A (2011) PID neural network based speed control of asynchronous motor using programmable logic controller. Advances in Electrical and Computer Engineering 11(4):23-28.

[14] Ahiska R, Mamur H (2012) A test system and supervisory control and data acquisition application with programmable logic controller for thermoelectric generators. Energy conversion and management 64:15-22.

[15] Bayindir R, Cetinceviz Y (2011) A water pumping control system with a programmable logic controller (PLC) and industrial wireless modules for industrial plants-An experimental setup. ISA transactions. 50(2):321-328.

[16] Mocanu DA, Bădescu V, Bucur C, Ștefan I, Carcadea E, Răboacă MS, Manta I (2020) PLC Automation and Control Strategy in a Stirling Solar Power System. Energies 13(8):1-19.

[17] Singh R, Verma HK (2018) Development of PLC-based controller for pneumatic pressing machine in engine-bearing manufacturing plant. Procedia Computer Science 125:449-458.

[18] Pratap KA, Marjorie SR, Saikumar M (2020) Automatic vegetable chopper using image processing. Materials Today: Proceedings 33(7):4787-4789.

[19] Choi JY, Moon KD (2020) Non-destructive discrimination of sesame oils via hyperspectral image analysis. Journal of Food Composition and Analysis 90:103505.

[20] Fermo IR, Cavali TS, Bonfim-Rocha L, Srutkoske CL, Flores FC, Andrade CM (2020) Development of a low-cost digital image processing system for oranges selection using hopfield networks. Food and Bioproducts Processing 125:181-192.

[21] Nasir M, Attique Khan M, Sharif M, Lali IU, Saba T, Iqbal T (2018) An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach. Microscopy research and technique 81(6):528-543.

[22] Norton KA, Iyatomi H, Celebi ME, Schaefer G, Tanaka M, Ogawa K (2010) Development of a novel border detection method for melanocytic and non-melanocytic dermoscopy images. In 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology.

[23] Garnavi R, Aldeen M, Celebi ME, Varigos G, Finch S (2011) Border detection in dermoscopy images using hybrid thresholding on optimized color channels. Computerized Medical Imaging and Graphics 35(2):105–115.

[24] Satheesha T, Satyanarayana D, Giriprasad M, Nagesh K (2016) Detection of melanoma using distinct features. In 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC), IEEE.

[25] Abbas Q, Fondón I, Rashid M (2011) Unsupervised skin lesions border detection via two-dimensional image analysis. Computer Methods and Programs in Biomedicine 104(3):e1-e15.

[26] Do TT, Zhou Y, Zheng H, Cheung NM, Koh D (2014) Early melanoma diagnosis with mobile imaging. In 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE.

[27] Sadri AR, Zekri M, Sadri S, Gheissari N, Mokhtari M, Kolahdouzan F (2013) Segmentation of dermoscopy images using wavelet networks. IEEE Transactions on Biomedical Engineering 60(4):1134-1141.

[28] Saez A, Serrano C, Acha B (2014) Model-based classification methods of global patterns in dermoscopic images. IEEE Transactions on Medical Imaging 33(5):1137-1147.

[29] Zhou H, Schaefer G, Sadka AH, Celebi ME (2009) Anisotropic mean shift based fuzzy c-means segmentation of dermoscopy images. IEEE Journal of Selected Topics in Signal Processing 3(1):26-34.

[30] Kabade Rohit S, Gaikwad M (2013) Segmentation of brain tumour and its area calculation in brain MR images using K-mean clustering and fuzzy C-mean algorithm. International Journal of Computer Science & Engineering Technology 4(05):524-531.

[31] Girgis MF, Method for Peeling and Processing Grain, (December 31 1968- 3419056 patent).

[32] Dopp S, Device for the Wet Treating of Seed Material, (December 16 1986-4628807 patent).

[33] Beeler ML, Peeler with Inclined Grooves in Side Wall, (April 8 1997-5617783 patent).

[34] Demir Z (2019) Susam Zarı Soyma Makinasının Tasarımı, Firat University Journal of Engineering Science 31(1):241-248.

[35] Raid AM, Khedr WM, El-Dosuky MA, Aoud M (2014) Image restoration based on morphological operations. International Journal of Computer Science, Engineering and Information Technology (IJCSEIT) 4(3):9-21.

[36] Patel S, Trivedi P, Gandhi V, Prajapati GI (2013) 2D basic shape detection using region properties. International Journal of Engineering Research & Technology 2(5):1147-1153.

BIOGRAPHIES

Ahmet Gündoğdu Elazig in 1974. He received the B.S. and M.S. degrees in electrical teaching from the University of Firat. Elazig, in 2004 and the Ph.D. degree in electrical engineering from Firat University, Elazig, in 2012. He has been working as an Assoc. Prof. Dr. at Electrical Engineering Department of Batman University. Research interests include electrical machines, motor control, renewable energy, MPPT.

Emrullah Acar received the MSc degree in electrical engineering (2012 year) and the PhD degree in electrical engineering (2017 year) from the University of Dicle, Diyarbakır/Turkey. From 2010 to 2018, he was a Research Assistant with the electrical engineering department. Since 2022, he has been an Assoc. Prof. Dr. with the Electrical Engineering Department, Batman University.

Reşat Çelikel was born in 1980. He received B.S, M.S and PhD degrees from Firat University, Elazig, Turkey. He has been working as an Assoc. Prof. Dr. at Firat University. His current interests are BLDC motors, flywheel energy storage, FPGA, renewable energy, MPPT.

European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.14, No.1, 2024

Research Article

Frame Error Rate Approximations in Coded LoRa Systems

Rıfat Volkan Şenyuva[1*]

[1*]Maltepe University, Electrical-Electronics Engineering Department, 34857, Maltepe, Istanbul, Turkey. (e-mail: rifatvolkansenyuva@maltepe.edu.tr).

ARTICLE INFO

Received: Dec., 05. 2023
Revised: Mar., 26. 2024
Accepted: Mar., 27. 2024

Keywords:
Frame Error Rate
Hamming Coding
LoRa
Internet of Things (IoT)

Corresponding author: *Rıfat Volkan Şenyuva*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1400982

ABSTRACT

In this study, we consider the approximations for calculating the frame error rate (FER) of a communication system using the coded low power long range (LoRa) modulation for the additive white Gaussian noise (AWGN) channel. The discrete-time baseband model of the LoRa communication system encompasses the (7,4) Hamming encoder, a diagonal interleaver, a Gray encoder, and the LoRa baseband modulator. Two FER approximations using a tighter uncoded symbol error rate (SER) bound are derived. The derived approximations are compared against the numerical evaluation of the FER in Monte Carlo trials and the approximations proposed by Afisiadis *et al*. Numerical results show that the proposed FER approximations are closer to the Monte Carlo trials.

1. INTRODUCTION

In this paper we address the problem of calculating the frame error rate (FER) of the low power long range (LoRa) systems for the additive white Gaussian noise (AWGN) channel with Hamming codes which are specific linear block codes. Due to emerging Internet of Things (IoT) applications, the transmission of shorter data units, that is shorter frames, using short to medium-length linear block codes have gained renewed interest within the information theory community [1-2]. Thus, developing tighter bounds on the achievable performance via the best error correcting codes is an important research topic.

This paper focuses on the FER of a single user in a coded LoRa system using (7,4) Hamming codes. The computational complexity of calculating the exact FER of the coded LoRa systems for practical signal dimensions and constellation sizes is very high due to the involvement of combinatorics. Several approaches proposed in [3-5] tackle this problem. While [3] and [4] analyze the FER and bit error rate (BER) of a single user respectively, [5] investigates both BER and FER of the successive interference cancellation method proposed for two interfering users. Instead of the symbol error rate (SER) bound of [6] used in [3], we adapt the SER bound proposed for Rician fading channels in [7] to the AWGN channel and then employ it in the FER approximations. The proposed BER expression of [4] involves combinatorics without any advantage computational complexity-wise and does not investigate the FER. The setup of [5] differs from our single-user setup due to the collision of two users, and the performance results of [5] are empirical. We present our numerical results using the normalized signal-to-noise ratio (SNR) defined the SNR per information bit for the power-limited LoRa modulation.

The summary of this paper's contributions is as follows:

- We developed two FER approximations. These are closer to the Monte Carlo results than those in [3].
- We use a tighter SER bound without the high SNR assumption.
- The numerical results are given against the normalized SNR, which is valid for the power-limited LoRa modulation.

The remaining parts of this paper are organized as: the literature review is in Section 2. The transceiver chain and the frame structure along with demodulation and Hamming decoding of LoRa is introduced in Section 3. Then the frame error rate approximations are derived using tighter bit error rate bound in Section 4. The numerical evaluations of the frame error rate approximations are compared in Section 5. Finally, Section 6 reports the conclusions.

2. LITERATURE REVIEW

IoT is a new communication perspective where billions of interconnected devices such as sensors, actuators are envisioned to be integrated into the same network and the Internet through wireless links. IoT has many applications including smart home systems [8-9], improved security [10], smart metering, logistics, localization, tracking, health monitoring, smart farming, and environment monitoring [11-12]. Since most of the devices within this massive network are resource constrained and have low power, the wireless communication between these devices must require new specialized protocols which are investigated under the machine-type communications (MTC). All the seven layers including the physical layer up to the application layer of the open systems interconnection (OSI) reference communication model must be covered by these new specialized protocols. Amongst the viable solutions for the implementation of the IoT networks, the low-power wide-area networks (LPWANs) is becoming a rising substitute for the multihop short-range transmission technologies such as ZigBee and Bluetooth or the wireless cellular standards like the Third Generation Partnership Project (3GPP) and Long Term Evolution (LTE). LPWANs use the unlicensed industrial, scientific, and medical (ISM) frequency bands of 2.4 GHz, 868/915 MHz, 433 MHz, and 169 MHz [11-12]. The physical layer design of the LPWAN devices allow them to have rural area ranges between 10-15 km and urban area ranges between 2-5 km. The tradeoff for having these cellular-like coverage ranges is the significantly reduced data rate of LPWAN devices compared to Zigbee and Bluetooth [11-12].

One of the solutions used by the LPWANs for designing the physical layer is LoRa. LoRa is an M-ary digital modulation technique using chirp spread spectrum (CSS). LoRa employs the chirp waveforms for baseband modulation. The instantaneous frequency of the chirp waveforms is linearly increased within the symbol interval. The number of samples obtained in the output of the LoRa baseband demodulator at the end of each symbol interval depends on the spreading factor (SF). The coverage range of LoRa can be extended by increasing the SF, but the data rate is reduced [13-14].

Although the data rate can be increased by decreasing the SF and so the coverage range, it is still limited for many IoT applications such as smart homes/buildings, image transmission, and indoor IoT [15]. Thus, there is significant ongoing research on increasing the data rate and analyzing the achievable performance of the LoRa systems. It is shown that the data rate can be increased by modifying the nominal LoRa modulation proposed in the literature such as the interleaved chirp spreading (ICS) LoRa [16], the slope-shift keying (SSK) LoRa [17], and the frequency-bin-index (FBI) LoRa [15]. [18] compare the modified LoRa modulations with no coding applied and focuses on the differences in the spectral efficiency and the demodulator complexity amongst the modulations. The bounds proposed in [4,6,19-20] for the SER of the uncoded LoRa modulation use the AWGN channel model. The performance of the uncoded LoRa under Rayleigh and Rician fading channels has been investigated by [6-7,21]. The error rate for the uncoded LoRa with another interfering terminal using the same SF is analyzed in [19-20]. The expressions for calculating the BER and the FER of the coded LoRa modulation are given for the AWGN channel in [4] and [3] respectively.

3. SYSTEM MODEL

The physical layer of LoRa systems is explained in this section. The transceiver chain along with the structure of the frames used in LoRa systems is given in Figure 1. LoRa systems transmit data using the frame structure [3,5,14] shown in Figure 1. The preamble part of the frame consists of several upchirp symbols, and 4.25 symbols used as frame delimiters for synchronization. The header has the frame information such as its length, the coding rate, the cyclic redundancy check (CRC) and the checksum. The data is contained in the payload and the frame rate approximations are given for the payload part of the frame [3,5,14].

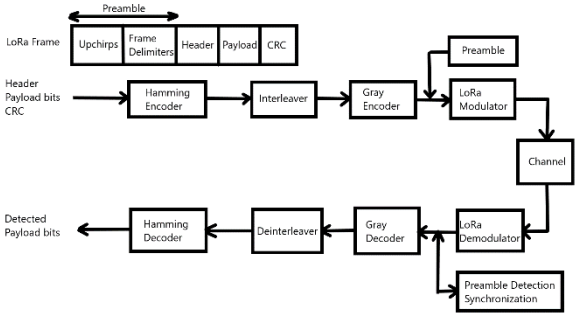


Figure 1. LoRa frame structure and transceiver chain

As it can be observed from Figure 1, the transmitter of the LoRa transceiver chain start with bit level operations including Hamming encoding, interleaving, and Gray encoding. Then LoRa modulation is used to convert Gray encoded bits into symbols and then into passband signals suitable for transmission. At the receiver chain the preamble part of the incoming signal is used for synchronization first before demodulation. Then the demodulated LoRa symbols are applied to Gray decoder, deinterleaver, and Hamming decoder in order to detect the payload bits.

2.1. LoRa Modulation and Demodulation

LoRa is an M-ary digital modulation scheme. The SF-length tuples, $(b_0, \dots, b_{\mathrm{SF}-1})$, of information bits are converted into M-dimensional signal samples, $\mathbf{a}_m = \left[a_{m,0}, \dots, a_{m,M-1}\right]^{\mathrm{T}} \in \mathcal{A}$, from the constellation $\mathcal{A} = \{\mathbf{a}_0, \dots, \mathbf{a}_{M-1}\}$ at every symbol interval, T_{symbol}, by the encoder [7]. The signal dimension, M, is equal to $M = 2^{\mathrm{SF}}$ where SF can take integer values from SF$\in\{7, \dots, 12\}$. Each symbol in the LoRa constellation, $\mathbf{a}_m$, has M samples of which only one is nonzero that is

$$a_{m,l} = \begin{cases} \sqrt{E}, l = m \\ 0, \text{else.} \end{cases} \quad (1)$$

E shows the energy of the LoRa symbol in Equation (1) and the nonzero sample index is equal to $m = \sum_{j=0}^{\mathrm{SF}-1} b_j 2^j$ [7]. The baseband signal of the LoRa modulation is written in terms of the symbols as follows

$$x_m(t) = \sum_{l=0}^{M-1} a_{m,l}\phi_l(t), \quad 0 \le t \le T_{\mathrm{symbol}} \quad (2)$$

The frequency shifted chirp waveform, $\phi_l(t)$, in Equation (2) is

$$\phi_l(t) = \exp\left\{i2\pi Wt\left[\frac{l}{M} - \frac{1}{2} + \frac{Wt}{2M} - u\left(t - \frac{M-l}{W}\right)\right]\right\} \quad (3)$$

where the chirp waveform has bandwidth W and the unit step function are shown as $u(t)$ [7,13]. Then the baseband waveform in Equation (3) is modulated to passband and the signal received at the demodulator becomes [7]

$$Y(t) = 2\Re\{[x_m(t) + N(t)]e^{i2\pi f_c t}\}, \quad 0 \le t \le T_{\text{symbol}} \quad (4)$$

In Equation (4), $\Re(\cdot)$ takes the real part of its complex argument and the additive baseband noise process, $N(t)$, is white Gaussian with its single-sided spectral density given as N_0. The sub-GHz ISM band, 863-870 MHz band, is allocated for the LoRa carrier frequencies, f_c, in Europe. The received signal in Equation (4) is first passed through the Hilbert filter and the output of the filter multiplied with the complex carrier yields the baseband signal. For the sampling interval of $T_{\text{sampling}} = 1/W$ the discrete-time baseband signal samples are

$$Y[n] = x_m[n] + N[n], \quad n = 0, \dots, M-1 \quad (5)$$

where $M = WT_{\text{symbol}}$ [7]. Both the baseband noise samples and the received signal samples in Equation (5) are circularly symmetric complex Gaussian random variables described as $N[n] \sim \mathcal{CN}(0, N_0)$ and $Y[n] \sim \mathcal{CN}(x_m[n], N_0)$ respectively. The baseband LoRa samples are given as

$$x_m[n] = \sqrt{\frac{E}{M}}(-1)^n e^{i\pi n^2/M} e^{i2\pi mn/M} = x_0[n]e^{i2\pi mn/M} \quad (6)$$

where $x_0[n] = \sqrt{E/M}(-1)^n e^{i\pi n^2/M}$ is defined as the upchirp signal [7,17]. The maximum a-posteriori probability (MAP) rule gives the optimum detector for the baseband samples in Equation (5) as

$$\hat{m} = \underset{0\le\tilde{m}\le M-1}{\text{argmax}}\, f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}_m|\mathbf{y}) = \underset{0\le\tilde{m}\le M-1}{\text{argmax}}\, \ln\left[f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}_m|\mathbf{y})\right] \quad (7)$$

where $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}_m|\mathbf{y})$ is the conditional probability density function of $\mathbf{X} = \mathbf{x}_m = (x_m[0], \dots, x_m[M-1])^{\mathrm{T}}$ given $\mathbf{Y} = \mathbf{y} = (y[0], \dots, y[M-1])^{\mathrm{T}}$ and $\ln(\cdot)$ is the natural logarithm of its argument. According to the Bayes theorem, $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}_m|\mathbf{y})$ is equal to

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}_m|\mathbf{y}) = \frac{p_{\mathbf{X}}(\mathbf{x}_m) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}_m)}{f_{\mathbf{Y}}(\mathbf{y})} \quad (8)$$

$$= C\exp\left[2\Re\left\{\sum_{n=0}^{M-1} y[n]x_m^*[n]\right\}\right] \quad (9)$$

where the probability mass function for m-th message signal is shown as $p_{\mathbf{X}}(\mathbf{x}_m)$ and it is equal to $p_{\mathbf{X}}(\mathbf{x}_m) = 1/M$ due to equal probability assumption and $f_{\mathbf{Y}}(\mathbf{y})$ is the marginal probability density function (PDF) of $\mathbf{Y}$ in Equation (8) [17]. The C term in Equation (9) is not explicitly given since it is independent of the message, m, and does not affect the detection rule. Plugging Equation (9) into Equation (7) yields the optimum detector

$$\hat{m} = \underset{0\le\tilde{m}\le M-1}{\text{argmax}}\, \Re\left\{\sum_{n=0}^{M-1} y[n]x_0^*[n]e^{-i2\pi\tilde{m}n/N}\right\} \quad (10)$$

$$= \underset{0\le\tilde{m}\le M-1}{\text{argmax}}\, \Re\left\{\sum_{n=0}^{M-1} \check{R}[\tilde{m}]\right\} \quad (11)$$

where $\check{R}[\tilde{m}] = \sum_{n=0}^{M-1} r[n]e^{-i2\pi\tilde{m}n/M}$ is the discrete Fourier transform (DFT) of the dechirped signal, $r[n] = y[n]x_0^*[n]$, evaluated at $\tilde{m}$-th frequency [7]. The fast Fourier transform (FFT) algorithm may calculate the DFT of the dechirped signal in $\mathcal{O}(M \log M)$ operations.

2.2. Hamming Encoding and Decoding

(n_c, k_c) Hamming codes are implemented in LoRa systems. There are 2^{k_c} codewords of length n_c in a (n_c, k_c) Hamming code. The length of the dataword and the codewords chosen in practice are $k_c = 4$ and $n_c \in \{5,6,7,8\}$ respectively [3,5,14]. This paper investigates the frame error rate approximations of the LoRa systems for the (7,4) Hamming code which can correct single-bit errors. The LoRa payload bits, $\mathbf{u} \in \{0,1\}^{k_c}$, are Hamming encoded using

$$\mathbf{c} = (\mathbf{uG})_2 \quad (12)$$

where $\mathbf{G} = [\mathbf{I}_4\ \mathbf{P}]$ is the generator matrix for the systematic (7,4) Hamming code and $(\cdot)_2$ denotes the elementwise modulo-2. $\mathbf{I}_4$ represents the identity matrix of size 4×4 and the parity matrix is given as

$$\mathbf{P} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (13)$$

Error bursts can be corrected if the codewords are interleaved such that the location of errors is distributed over many codewords. To accomplish this, LoRa systems employ a diagonal interleaver which reads SF number of codewords each with length n_c and reorder them in a block of SF rows and n_c columns shown in Figure 2.

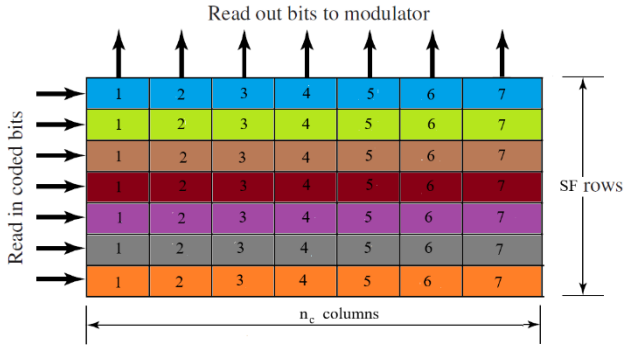


Figure 2. LoRa Interleaver for SF=7 and $n_c = 7$

Then the columns of the interleaver is Gray mapped and modulated using the LoRa signaling scheme explained in Section 2.1. Once the demodulated signal is Gray demapped in the receiver chain (Figure 1), the deinterleaver constructs an $n_c \times$ SF block structure by reordering its rows and outputs its columns. The decoding of the deinterleaved codewords is

realized by hard decision. The first step is to calculate the syndrome of the Hamming decoder input $\mathbf{v}$

$$\mathbf{s} = (\mathbf{v}\mathbf{H}^{\mathrm{T}})_2 \tag{14}$$

Then, we find the row called a coset corresponding to $\mathbf{s}$ in the standard array for the Hamming code. The first element of the corresponding coset gives the error pattern. The rows of $\mathbf{E}$ matrix in Equation (15) include all of the single-bit error patterns

$$\mathbf{E} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{15}$$

If the corresponding coset is the l-th row of $\mathbf{E}$ shown as $\mathbf{e}_l$, then the corrected codeword can be obtained by adding it to the input vector as in

$$\hat{\mathbf{c}} = \mathbf{v} \oplus \mathbf{e}_l \tag{16}$$

where $\oplus$ is the modulo-2 elementwise vector addition.

4. FRAME ERROR RATE

The FER approximations are derived in this section. First the uncoded bit error rate is bounded and then the codeword error rate is calculated using the given bound for the uncoded bit error rate. Finally, two approximations for the FER are derived in terms of the codeword error rate.

3.1. Uncoded Bit Error Rate

The probability of a symbol error, $P_s = P(m \neq \hat{m})$, for the uncoded LoRa modulation [6] is given as

$$P_s = \sum_{k=1}^{M-1} \frac{(-1)^{k+1}}{k+1} \binom{M-1}{k} \exp\left(-\frac{k}{k+1}\frac{E}{N_0}\right) \tag{17}$$

The calculation of the binomial coefficient in Equation (17) is subject to precision errors for large M. Since the smallest M for practical LoRa systems is equal to $M = 2^7 = 128$, the calculation of Equation (17) becomes challenging. There exist several approximations in the literature for the evaluation of Equation (17) [4,6,19-21]. The upper bound proposed for Rician channels in [7] can be modified for AWGN channels by setting the mean of the channel variable to one and the variance of the channel variable to zero as in

$$P_s \approx 1 - Q_1(\alpha, \beta) + \frac{M-1}{2} e^{-E/2N_0} Q_1\left(\alpha\sqrt{2}, \beta\sqrt{2}\right) \tag{18}$$

where $Q_1(\cdot)$ is the Marcum Q function and its arguments are $\alpha = \sqrt{2E/N_0}$ and $\beta = \sqrt{2\ln(M-1)}$. The probability of a bit error can be approximated to the half of the probability of a symbol error

$$P_b = \frac{2^{\mathrm{SF}-1}}{2^{\mathrm{SF}}-1} P_s \approx \frac{P_s}{2} \tag{19}$$

3.2. Codeword Error Rate

If the output of the Hamming decoder is represented as $\hat{\mathbf{c}}$, then the probability that $\hat{\mathbf{c}}$ is not equal to the Hamming encoder output, $\mathbf{c}$, is the codeword error probability and it is given as $P_c = P(\mathbf{c} \neq \hat{\mathbf{c}})$. Since the (7,4) Hamming decoder is capable of correcting only the single-bit errors, the event of a codeword error occurs when there are at least two bits in error that is

$$P_c = P(\{w_H(\mathbf{v} - \mathbf{c}) \geq 2\}) \tag{20}$$

where $w_H(\mathbf{v} - \mathbf{c})$ shows the Hamming distance, the number of ones in the difference of $\mathbf{v}$ and $\mathbf{c}$. The bit errors at the Hamming decoder input, $\mathbf{v}_j$, are independent and identically distributed with Equation (19). The codeword error probability from Equation (20) can be calculated as

$$P_c = 1 - \left[(1-P_b)^{n_c} + \binom{n_c}{1} P_b (1-P_b)^{n_c-1}\right] \tag{21}$$

3.3. Frame Error Rate

We assume that a LoRa frame has N_p payload symbols. Choosing N_p as an integer multiple of the codeword length, n_c, results in $N_c = N_p \mathrm{SF}/n_c$ number of codewords in a LoRa frame. For one block of the interleaver, the total number of transmitted codewords is equal to SF that is $\mathbf{c}_j$ for $j \in \{1, \ldots, \mathrm{SF}\}$, and they are shown as $\mathbf{C}_{\mathrm{bl}} \in \{0,1\}^{\mathrm{SF}\times n_c}$. The block of the corrected codewords at the output of the Hamming decoder $\hat{\mathbf{c}}_j, j \in \{1, \ldots, \mathrm{SF}\}$ are given as $\hat{\mathbf{C}}_{\mathrm{bl}} \in \{0,1\}^{\mathrm{SF}\times n_c}$. For one deinterleaver block, the probability that none of the codewords are erroneously decoded is

$$P(\hat{\mathbf{C}}_{\mathrm{bl}} = \mathbf{C}_{\mathrm{bl}}) = \prod_{j=1}^{\mathrm{SF}} P(\hat{\mathbf{c}}_j = \mathbf{c}_j | \hat{\mathbf{c}}_1 = \mathbf{c}_1, \ldots, \hat{\mathbf{c}}_{j-1} = \mathbf{c}_{j-1}) \tag{22}$$

The probability that the last j-th codeword is erroneously decoded given that all the decoding's for the previous $j-1$ codewords are error free is given as

$$P_c^{(j)} = P\left(\hat{\mathbf{c}}_j \neq \mathbf{c}_j \middle| \hat{\mathbf{c}}_1 = \mathbf{c}_1, \ldots, \hat{\mathbf{c}}_{j-1} = \mathbf{c}_{j-1}\right) \tag{23}$$

The correct decoding probability in Equation (22) may be expressed in terms of Equation (23) as in

$$P(\hat{\mathbf{C}}_{\mathrm{bl}} = \mathbf{C}_{\mathrm{bl}}) = \prod_{j=1}^{\mathrm{SF}} (1 - P_c^{(j)}) \tag{24}$$

The FER for N_p symbols can be calculated as in

$$P\left(\hat{\mathbf{C}}_{\mathrm{pl}} \neq \mathbf{C}_{\mathrm{pl}}\right) = 1 - \left(\prod_{j=1}^{\mathrm{SF}} \left(1 - P_c^{(j)}\right)\right)^{N_p/n_c} \tag{25}$$

where $\hat{\mathbf{C}}_{\mathrm{pl}}, \mathbf{C}_{\mathrm{pl}} \in \{0,1\}^{N_c \times n_c}$ show the matrices containing the decoded and the transmitted codewords respectively and N_p/n_c is the number of interleaver blocks within the frame.

The calculation of Equation (25) is quite involved due to large number of possible error patterns inside a deinterleaver block. Straightforward removal of the conditioning in Equation (23) yields a first FER approximation as given in

$$P(\hat{\mathbf{C}}_{\text{pl}} \neq \mathbf{C}_{\text{pl}}) \approx 1-(1-P_c)^{\frac{N_p \text{SF}}{n_c}}, \quad (26)$$

which has a much lesser computational complexity compared Equation (25). Since the codewords errors within the same interleaver block are dependent, the first FER approximation in Equation (26) is a looser upper bound for the FER [3].
If the single-bit errors in Equation (23) is to be ignored, then a much tighter bound for the FER can be found. Equation 23 can be approximated to

$$P_c^{(j)} \approx P(\hat{\mathbf{c}}_j \neq \mathbf{c}_j | \mathbf{v}_1 = \mathbf{c}_1, \dots, \mathbf{v}_{j-1} = \mathbf{c}_{j-1}) \quad (27)$$

The condition given in the codeword error probability in Equation (27) means that the error is due to the last j-th position in $\mathbf{v}_j$. To calculate Equation (27), the conditional probability of a symbol error must be given as

$$P_s^{(j)} = P(\hat{m} \neq m | \hat{b}_1 = b_1, \dots, \hat{b}_{j-1} = b_{j-1}) \quad (28)$$

where $\hat{b}_j = b_j$ means that j-th bit of the symbol m is decoded correctly. Given that the error is in $\text{SF}-j+1$ bit locations due to Equation (27), the LoRa signal dimension parameter, M, in Equation (18) can be modified to express the conditional symbol error probability as follows

$$P_s^{(j)} \approx 1-Q_1(\alpha, \beta^{(j)}) + \frac{M^{(j)}}{2} e^{-E/2N_0} Q_1(\alpha\sqrt{2}, \beta^{(j)}\sqrt{2}) \quad (29)$$

where $M^{(j)} = 2^{\text{SF}-j-1}-1$ and $\beta^{(j)} = \sqrt{2\ln(M^{(j)})}$. The conditional bit error probability is going to be the half of Equation (29) that is $P_b^{(j)} = P_s^{(j)}/2$. The codeword error probability in Equation (27) can be written as

$$P_c^{(j)} \approx 1-\left[\left(1-P_b^{(j)}\right)^{n_c} + \binom{n_c}{1} P_b^{(j)} \left(1-P_b^{(j)}\right)^{n_c-1}\right] \quad (30)$$

Rewriting the exact FER expression from Equation (25) using the new codeword error probability shown in Equation (30) yields the second FER approximation given as

$$P(\hat{\mathbf{C}}_{\text{pl}} \neq \mathbf{C}_{\text{pl}}) \approx 1-\left(\prod_{j=1}^{\text{SF}} \left(1-P_b^{(j)}\right)^{n_c} + \binom{n_c}{1} P_b^{(j)} \left(1-P_b^{(j)}\right)^{n_c-1}\right)^{N_p/n_c} \quad (31)$$

5. NUMERICAL RESULTS

The empirical and analytical evaluations of the FER are presented against the signal energy per bit over the noise energy that is $E_b/N_0 = E/(\text{SF}N_0)$, since the AWGN channel using LoRa modulation is considered power-limited regime due to its spectral efficiency and the normalized SNR is E_b/N_0 for the power-limited regime. The numerical results are given for the (7,4) Hamming code. The LoRa frame has $N_p = 35$ payload symbols which is chosen as an integer multiple of $n_c = 7$. The numerical FER results are averaged over 10^6 LoRa frames. Figure 3 shows the numerical and the analytical evaluations using the proposed approximations for the FER of the (7,4) Hamming coded LoRa modulation using SF=7 under the AWGN channel. The (7,4) Hamming code has provided about 2 dB coding gain over the uncoded LoRa modulation. The first FER approximation in Equation (26) is looser to the numerical FER by more than 1 dB since it is derived by ignoring the conditioning in Equation (23). The second FER approximation in Equation (31) is within 1 dB away from the numerical result and so much tighter compared to the first approximation. Due to employing the tighter SER bound of [7], both proposed FER approximations are tighter than the approximations proposed in [3].

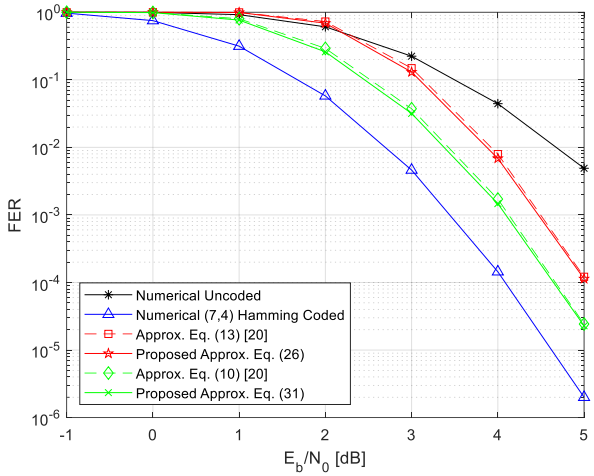


Figure 3. FER results for the (7,4) Hamming coded LoRa modulation with SF=7 under AWGN

The numerical evaluation of the first FER approximation in Equation (26) is presented for the whole range of SF values used in practice, $\text{SF}\in\{7,\dots,12\}$, in Figure 4. The results in Figure 4 show that there is about 1.5 dB gap between SF=7 and SF=12 for FER greater than 10^{-4}. Thus, increasing the SF results in decreased FER for the coded LoRa modulation. We can see from Figure 4 that the proposed FER approximation in Equation (26) is better than the first approximation given in [3] for E_b/N_0 less than 4 dB.

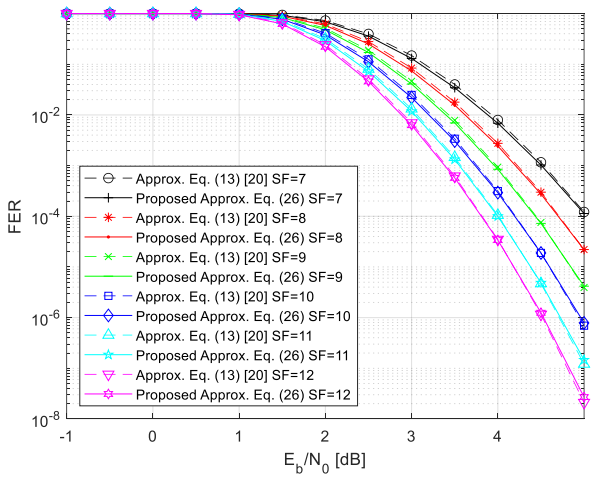


Figure 4. The FER approximation in Eq. (26) for the (7,4) Hamming coded LoRa modulation with $\text{SF}\in\{7,\dots,12\}$ under AWGN

Figure 5 shows the results for the second FER approximation in Equation (31). Our second FER approximation slightly improves upon the corresponding approximation proposed in [20] across all SF values in the low E_b/N_0 region where E_b/N_0 less than 4 dB. For E_b/N_0 greater than 4 dB, both approximations yield the same results. The reason for this difference is while the SER bound used in the FER approximations of [3] is valid under a high SNR assumption, which does not hold in the low E_b/N_0 region, our SER bound does not rely on any SNR approximation.

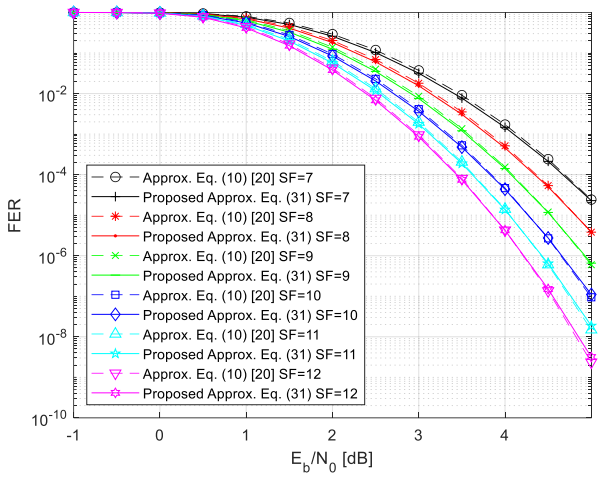


Figure 5. The FER approximation in Eq. (31) for the (7,4) Hamming coded LoRa modulation with SF∈{7, … ,12} under AWGN

6. CONCLUSION

This paper studies the approximations for calculating the FER of coded LoRa systems under the AWGN channel model. The LoRa frames of our system model pass through a (7,4) Hamming encoder, a diagonal interleaver, and a gray encoder before entering the LoRa baseband modulator. First, we derive the uncoded bit error probability of the LoRa modulation for the AWGN channel. Then, we represent the codeword error probability for the (7,4) Hamming decoder in terms of the uncoded bit error probability and give the exact FER expression. We develop two FER approximations with reduced complexity to replace the exact FER expression. We relax the conditioning between the codeword errors within the same interleaver block for the first FER approximation. As for the second FER approximation, we ignore the single-bit errors and derive a new conditional codeword error probability. We compare the proposed approximations against the numerical evaluations for varying SF. The performances of the proposed FER approximations are better than the ones proposed by Afisiadis *et al*.

REFERENCES

[1] G. Durisi, T. Koch and P. Popovski, "Toward Massive, Ultrareliable, and Low-Latency Wireless Communication With Short Packets," in *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1711-1726, Sept. 2016.

[2] M. C. Coşkun et al., "Efficient error-correcting codes in the short blocklength regime," Phys. Comm., vol. 34, pp. 66-79, Jun. 2019.

[3] O. Afisiadis, A. Burg and A. Balatsoukas-Stimming, "Coded LoRa Frame Error Rate Analysis", *IEEE ICC*, Dublin, Ireland, 2020, pp. 1-6.

[4] G. Baruffa, L. Rugini, L. Germani, F. Frescura, "Error probability performance of chirp modulation in uncoded and coded LoRa systems," Digit. Signal Process., vol. 106, Nov. 2020.

[5] J. Tapparel, M. Xhonneux, D. Bol, J. Louveaux and A. Burg, "Enhancing the reliability of dense LoRaWAN networks with multiuser receivers", IEEE Open J. Commun. Soc., vol. 2, pp. 2725-2738, 2021.

[6] T. Elshabrawy and J. Robert, "Closed-Form Approximation of LoRa Modulation BER Performance," in *IEEE Comm. Letters*, vol. 22, no. 9, pp. 1778-1781, Sept. 2018.

[7] R. V. Şenyuva, "Union bounds on the symbol error probability of LoRa modulation for flat Rician block fading channels", *Phys. Comm.*, vol. 58, 2023.

[8] Z. L. OO, T. W. Lai, and A. Moe, "IoT Based Home Automation System using a REST API Architecture", *EJT*, vol. 12, no. 2, pp. 123-128, 2022.

[9] R. Daş, and T. Ababaker, "Design and Application of a Smart Home System based on Internet of Things", *EJT*, vol. 11, no. 1, pp. 34-42, 2021.

[10] F. Abdulkafi, S. Kurnaz, and F. A. Abdulkafi, "Security Improvements of Internet of Things Systems." *EJT*, vol. 10, no. 2, pp. 476 - 488, 2020.

[11] U. Raza, P. Kulkarni and M. Sooriyabandara, "Low Power Wide Area Networks: An Overview," in *IEEE Comm. Surveys & Tutorials*, vol. 19, no. 2, pp. 855-873, Secondquarter 2017.

[12] M. Centenaro, L. Vangelista, A. Zanella and M. Zorzi, "Long-range communications in unlicensed bands: the rising stars in the IoT and smart city scenarios," in *IEEE Wireless Comm.*, vol. 23, no. 5, pp. 60-67, Oct. 2016.

[13] M. Chiani and A. Elzanaty, "On the LoRa Modulation for IoT: Waveform Properties and Spectral Analysis," in *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8463-8470, Oct. 2019.

[14] R. Ghanaatian, O. Afisiadis, M. Cotting and A. Burg, "Lora Digital Receiver Analysis and Implementation," *IEEE ICASSP*, Brighton, UK, 2019, pp. 1498-1502.

[15] H. Ma, Y. Fang, G. Cai, G. Han and Y. Li, "A New Frequency-Bin-Index LoRa System for High-Data-Rate Transmission: Design and Performance Analysis," in *IEEE Internet Things J.*, vol. 9, no. 14, pp. 12515-12528, July 2022.

[16] T. Elshabrawy and J. Robert, "Interleaved Chirp Spreading LoRa-Based Modulation," in *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3855-3863, April 2019.

[17] M. Hanif and H. H. Nguyen, "Slope-Shift Keying LoRa-Based Modulation," in *IEEE Internet Things J.*, vol. 8, no. 1, pp. 211-221, 1 Jan.1, 2021.

[18] R. V. Şenyuva, "Comparison of LoRa-Based Modulations," IEEE SIU, Safranbolu, Turkey, 2022, pp. 1-4.

[19] O. Afisiadis, M. Cotting, A. Burg and A. Balatsoukas-Stimming, "On the Error Rate of the LoRa Modulation With Interference," in *IEEE Trans. Wireless Comm.*, vol. 19, no. 2, pp. 1292-1304, Feb. 2020.

[20] O. Afisiadis, S. Li, J. Tapparel, A. Burg and A. Balatsoukas-Stimming, "On the Advantage of Coherent LoRa Detection in the Presence of Interference," in *IEEE Internet Things J.*, vol. 8, no. 14, pp. 11581-11593, July 2021.

[21] J. Courjault, B. Vrigneau, O. Berder and M. R. Bhatnagar, "A Computable Form for LoRa Performance Estimation: Application to Ricean and Nakagami Fading," in *IEEE Access*, vol. 9, pp. 81601-81611, 2021.

BIOGRAPHIES

Rıfat Volkan Şenyuva (8p) graduated from Istanbul Technical University, Istanbul, Turkey, with dual B.Sc. degrees in control engineering (2007) and computer engineering (2008). He then continued his studies at Boğaziçi University, Istanbul, Turkey, where he received M.Sc. (2009) and Ph.D. (2016) degrees in electrical–electronics engineering. His Ph.D. dissertation was focused on sparse signal recovery algorithms. From 2016 to 2018, he worked as a researcher in NETAS Telecommunication Inc. Since 2018, he has been with the Faculty of Engineering and Natural Sciences of Maltepe University, Istanbul, Turkey, where he is an Assistant Professor at the Department of Electrical–Electronics Engineering. His current research interests include multidimensional harmonic retrieval, machine-type communications, channel estimation in reconfigurable intelligent surface based systems.

European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.14, No.1, 2024

Research Article

Optimization of Optimal Power Flow considering Location of FACTS Devices using Partial Reinforcement Optimizer

Burçin Özkaya[1*]

[1*] Bandirma Onyedi Eylul University, Electrical Engineering Department, 10200, Bandırma, Balıkesir, Turkey. (e-mail: bozkaya@bandirma.edu.tr).

ARTICLE INFO

Received: May., 06. 2024
Revised: June., 11. 2024
Accepted: June., 13. 2024

Keywords:
Optimization
Partial reinforcement optimizer
Optimal power flow
FACTS devices

Corresponding author: *Burçin Özkaya*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI:
https://doi.org/10.36222/ejt.1479409

ABSTRACT

Optimal power flow (OPF) is the most commonly addressed modern power system planning and operating optimization problem. It becomes a very difficult and high complexity optimization problem with the inclusion of the optimal location and sizing of flexible AC transmission system (FACTS) devices. Therefore, to obtain the optimal solution for the OPF problem, it is necessary to use the most suitable meta-heuristic search (MHS) algorithm. In this study, an up-to-date and strong MHS algorithm known as the partial reinforcement optimizer (PRO) was used to solve the OPF problem considering optimal location and sizing of the multi-type FACTS devices. The objectives considered in the study were minimization of total cost, minimization of total cost with valve-point loading effect, and minimization of the real power loss. In the simulation studies, four case studies were considered and solved by PRO algorithm and its three rivals including dingo optimization algorithm, evolutionary mating algorithm, and snow geese algorithm. According to the results of the case studies, compared to its closest competitor, the evolutionary mating algorithm algorithm, its objective function values for Case-1, Case-2, Case-3, and Case-4 were lower by 0.0059%, 0.2814%, 0.0062%, and 0.0003%, respectively. The performance of PRO algorithm was evaluated using Friedman and Wilcoxon tests against its rivals. The Friedman test results show that PRO algorithm achieved the best rank first with 1.2333 score value among them. In summary, PRO algorithm achieved superior performance in solving these case studies.

1. INTRODUCTION

One of the most significant operating and planning issues in modern power systems is optimal power flow (OPF). Because of the nonlinear power flow equations, generator cost curves, and operational limitations, it is usually defined as a non-convex and nonlinear optimization problem. Its aim is to minimize the objective function while meeting a variety of operational limitations to specify the optimal operating conditions for a power system, usually an electrical grid, by adjusting control variables [1-3].

In power system operations, the absence of reactive power causes problems such as voltage fluctuations, reduction, collapse, and instability. Moreover, the generators and capacitor banks generate reactive power at a slow pace, making it challenging to adapt to sudden load changes. Therefore, flexible alternating current transmission system (FACTS) devices have an important role, such as controlling the power flow, increasing the power loss, and solving operational problems including congestion, voltage fluctuations, and line losses, and as to improve power system stability, transmission efficiency, and grid reliability [4]. In fact, FACTS devices can control many parameters of transmission lines; the most important of these are the shunt impedance, series impedance, and phase angles. FACTS devices are very helpful for a power system; however, the OPF problem with FACTS devices is challenging because of the complexity, nonlinearities, non-convexity, and modeling accuracy required for an accurate solution [3-5]. Therefore, meta-heuristic search (MHS) algorithms have been used by researchers in the literature to solve the OPF problem including FACTS devices. Some of MHS algorithms used for the solution of the OPF including FACTS devices were the lightning search algorithm [1], the krill herd algorithm [3], the enhanced bacterial foraging algorithm [5], the symbiotic organisms search algorithm [6], the new partitioned ant lion optimizer [7], the modified Runge Kutta optimizer [8], the lightning attachment procedure optimization [9], etc.

In the above mentioned research, FACTS devices were placed on fixed branches in the power system. However, the placement and sizing of the FACTS devices directly affect the power flow, voltage profiles, system stability, and overall grid performance [2, 10]. The optimal allocation of FACTS

devices can optimize the power flow, minimize the power losses, and improve the voltage stability. To sum up, the placement of the FACTS devices for the OPF problem is crucial for the efficient, reliable and cost-effective operation of power systems [10, 11]. The optimal placement and ratings of the FACTS devices for the OPF problem can be considered highly constrained, multimodal, and complex optimization problems.

In the literature, the optimal placement of FACTS devices in optimal power flow has been solved using MHS algorithms. Sulaiman and Mustaffa presented a comprehensive study using seven MHS algorithms in solving the OPF problem considering the optimal placement and rating of the TCSC and SVC devices on IEEE 14-bus test system [2]. Biswas et al. performed optimal power flow incorporating wind energy sources and multi-type FACTS devices using the success history-based adaptive differential evolution [10]. Inkollu and Kota proposed a hybrid algorithm to improve the voltage stability of the system and optimize the placement of the FACTS devices [12]. Sakr et al. proposed an adaptive differential evolution algorithm to optimize the location of the TCSC in the network [13]. Ziaee and Choobineh modeled the location of the TCSC problem as a mixed integer nonlinear program and Benders' decomposition method was used to solve the problem [14]. Raj and Bhattacharyya presented a comparative study using five different MHS algorithms for solving the reactive power planning with FACTS devices. Moreover, the optimal location of the TCSC and static VAR compensator (SVC) was determined [15]. In another study, Agrawal et al. presented a comparative study for the optimal location and sizing of the TCSC using evolutionary optimization techniques. The performance of the algorithms was tested on IEEE 14-bus, 30-bus, and 75-bus systems [16]. Khan et al. proposed a modified lightning attachment procedure optimizer for solving the optimal reactive power dispatch including TCSC and SVC devices. In this study, the optimal placement of them were also considered [17]. Nusair et al. performed a comprehensive study about the solution of the OPF problems including multi-type FACTS devices and wind renewable energy sources using four MHS algorithms [18]. Mohamed et al. proposed a hybrid algorithm including a gradient-based optimizer and a moth flame optimization algorithm to solve the OPF considering optimal allocation and rating of multi-type FACTS devices and wind power [19]. Hassan et al. introduced an enhanced hunter-prey optimization algorithm to solve the OPF problem including wind energy sources and FACTS devices such as SVC, TCSC, and thyristor-controlled phase shifter (TCPS). Here, the optimal location and sizing of the FACTS devices were also considered in the OPF problem [20]. Awad et al. proposed the der-based walrus optimization algorithm for solving the OPF problems on the IEEE 30-bus test system. The optimal location of multi-type of FACTS devices were also considered [21]. Mohamed et al. introduced chaotic african vultures optimization algorithm for solving the OPF problem incorporating wind power and multi-type FACTS devices. The objectives were the minimization of generation cost and minimization of active power loss. Moreover, the optimal location and rating of the FACTS devices were optimized [22]. Mahapatra et al. presented a hybrid cuckoo search algorithm and ant-lion optimizer for determining the optimal placement and sizing of the TCSC. In the study, the placement and sizing of the TCSC were determined by using CS and ALO algorithms, respectively [23].

In this study, the OPF problem including multi-type FACTS devices was solved by using partial reinforcement optimizer (PRO). The PRO algorithm was introduced to the literature by Taheri et al. in 2024 [24]. In the OPF problem, three objective functions were considered: (i) minimization of total cost, (ii) minimization of total cost with the valve-point loading effect, (iii) minimization of the real power loss. On the other hand, the allocation and rating of the TCSC and TCPS devices were determined to minimize the relevant objective function. Accordingly, four different case studies were carried out on the IEEE 30-bus test system using the objective functions and FACTS devices. The performance of the PRO algorithm was compared with the dingo optimization algorithm (DOA) [25], the evolutionary mating algorithm (EMA) [26], and the snow geese algorithm (SGA) [27]. Accordingly, the contributions of this study can be summarized as follows:

- The PRO algorithm was applied to solve the OPF problem considering location and sizing of the FACTS devices.
- Four different case studies were considered using the three objective functions and two FACTS devices.
- The performance of the PRO algorithm was compared with the up-to-date three MHS algorithms.
- A comprehensive analysis was performed using the convergence analysis and statistical analysis methods to prove the performance of the PRO algorithm in solving the OPF problem against its rivals.

The remaining sections of the paper are organized as follows: Section 2 presents the formulation of the problem, where the modeling of the TCSC and TCPS devices, the objective functions and the constraints considered in the OPF problem are defined. In section 3, the partial reinforcement optimizer is explained. Section 4 presents the simulation studies and results. Moreover, the results of the case studies, the convergence and the statistical analysis results were given. Section 5 presents the conclusion of the study.

2. FORMULATION OF THE PROBLEM

The primary goal of the OPF problem is to minimize the objective function to determine the optimal control variables, taking into account the equality and inequality constraints. It is typically expressed as in Eq. (1).

$$\begin{aligned} &Minimize \quad OF(\boldsymbol{v},\boldsymbol{u}) \\ &subject\ to \begin{cases} H(\boldsymbol{v},\boldsymbol{u}) = 0 \\ G(\boldsymbol{v},\boldsymbol{u}) \le 0 \end{cases} \end{aligned} \tag{1}$$

Here, OF, G, and H represent the objective function, the inequality and equality constraints, respectively. Also, $\boldsymbol{u}$ and $\boldsymbol{v}$ denote the control and state variables vector, respectively [1, 6].

In this study, the placement of TCSC and TCPS devices presented in section 2 is considered. The placement and sizing of these devices are optimized to minimize the related objective function. Accordingly, the state and control variables of this problem are expressed as follows:

$$\boldsymbol{u} = \begin{bmatrix} P_{THG_2} \dots P_{THG_{N_{THG}}}, V_{G_1} \dots V_{G_{N_{THG}}}, T_1 \dots T_{N_T}, \\ TCSC_1 \dots TCSC_m, \tau_{TCSC_1} \dots \tau_{TCSC_m}, \\ TCPS_1 \dots TCPS_p, \phi_{TCPS_1} \dots \phi_{TCPS_p} \end{bmatrix} \tag{2}$$

$$\boldsymbol{v} = \left[P_{THG_1}, V_{L_1} \ldots V_{L_{N_{PQ}}}, Q_{THG_1} \ldots Q_{THG_{N_{THG}}}, S_{L_1} \ldots S_{L_{N_{TL}}} \right] \quad (3)$$

where P_{THG}, Q_{THG}, and V_G active power, reactive power, and the voltages of the thermal generating units. T is the tap changing transformer settings, V_L is the voltage of the load buses, S_L is the transmission line loading. N_{THG}, N_T, N_{TL}, m, and p are the number of the transformers, the number of thermal generating units, the number of transmission lines, the number of TCSC devices, and the number of TCPS devices, respectively.

2.1. Modeling of FACTS devices

In this study, the thyristor-controlled phase shifter (TCPS) and thyristor-controlled series capacitor (TCSC) devices were considered, which are used to increase the power flow and loading ability of the line. This section includes two sub-sections, where the modeling of the TCSC and TCPS devices is presented.

2.1.1. Modeling of TCSC device

TCSC is used in power systems to enhance the system performance and control the power flow. The circuit model of the TCSC between the i^{th} and k^{th} buses in a power grid is given in Fig. 1. The equivalent reactance (X_{eq}) of the transmission line is given as follows [6]:

$$X_{eq} = X_{ik} - X_{TCSC} = X_{ik}(1-\tau) \quad (4)$$

where X_{ik} and X_{TCSC} are the inductive reactance of the line and TCSC devices, respectively. τ is called the series compensation degree with X_{ik}.

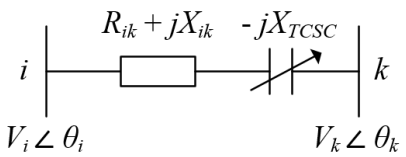


Figure 1. Circuit model of TCSC

The power flow equations between i^{th} and k^{th} buses including TCSC can be written as follows [6]:

$$P_{ik} = V_i^2 G_{ik} - V_i V_k G_{ik} \cos(\theta_{ik}) - V_i V_k B_{ik} \sin(\theta_{ik}) \quad (5)$$

$$Q_{ik} = -V_i^2 B_{ik} - V_i V_k G_{ik} \sin(\theta_{ik}) + V_i V_k B_{ik} \cos(\theta_{ik}) \quad (6)$$

$$P_{ki} = V_k^2 G_{ik} - V_i V_k G_{ik} \cos(\theta_{ik}) + V_i V_k B_{ik} \sin(\theta_{ik}) \quad (7)$$

$$Q_{ki} = -V_k^2 B_{ik} + V_i V_k G_{ik} \sin(\theta_{ik}) + V_i V_k B_{ik} \cos(\theta_{ik}) \quad (8)$$

where P_{ik} and Q_{ik} are the active and reactive power flows from i^{th} to k^{th} bus, θ_k and θ_i are the angles of the k^{th} and i^{th} buses, V_k and V_i are the voltages of the k^{th} and i^{th} buses, respectively.

The conductance (G_{ik}) and the susceptance (B_{ik}) of the transmission line are computed as follows [6]:

$$G_{ik} = \frac{X_{ik}}{X_{ik}^2 + X_{new}^2}, \; B_{ik} = -\frac{X_{new}}{X_{ik}^2 + X_{new}^2} \quad (9)$$

2.1.2. Modeling of TCPS device

TCPS can be modelled using a phase-shifting transformer with control parameter (Φ) in the power system [6]. The circuit model of the TCPS between i^{th} and k^{th} bus in a power grid is shown in Fig. 2.

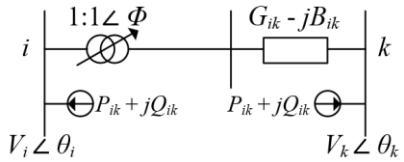


Figure 2. Circuit model of TCPS

The power flow equations between i^{th} and k^{th} bus including TCPS can be expressed as follows [6]:

$$P_{ik} = \frac{V_i^2 G_{ik}}{\cos^2\phi} - \frac{V_i V_k}{\cos\phi}\left[B_{ik}\sin(\theta_{ik}+\phi) + G_{ik}\cos(\theta_{ik}+\phi)\right] \quad (10)$$

$$Q_{ik} = -\frac{V_i^2 B_{ik}}{\cos^2\phi} - \frac{V_i V_k}{\cos\phi}\left[-B_{ik}\cos(\theta_{ik}+\phi) + G_{ik}\sin(\theta_{ik}+\phi)\right] \quad (11)$$

$$P_{ki} = V_k^2 G_{ik} - \frac{V_i V_k}{\cos\phi}\left[-B_{ik}\sin(\theta_{ik}+\phi) + G_{ik}\cos(\theta_{ik}+\phi)\right] \quad (12)$$

$$Q_{ki} = -V_k^2 B_{ik} + \frac{V_i V_k}{\cos\phi}\left[B_{ik}\cos(\theta_{ik}+\phi) + G_{ik}\sin(\theta_{ik}+\phi)\right] \quad (13)$$

The injected reactive and active power of TCPS at i^{th} and k^{th} bus is written as [6]:

$$P_{is} = -V_i^2 G_{ik}\tan^2\phi - V_i V_k \tan\phi\left[-B_{ik}\cos\theta_{ik}) + G_{ik}\sin(\theta_{ik})\right] \quad (14)$$

$$Q_{is} = V_i^2 B_{ik}\tan^2\phi + V_i V_k \tan\phi\left[B_{ik}\sin(\theta_{ik}) + G_{ik}\cos(\theta_{ik})\right] \quad (15)$$

$$P_{ks} = -V_i V_k \tan\phi\left[B_{ik}\cos(\theta_{ik}) + G_{ik}\sin(\theta_{ik})\right] \quad (16)$$

$$Q_{ks} = -V_i V_k \tan\phi\left[-B_{ik}\sin(\theta_{ik}) + G_{ik}\cos(\theta_{ik})\right] \quad (17)$$

2.2. Objective functions

In this study, three objective functions are considered.

(i) *Minimization of the total cost*: The cost of the thermal generating units is expressed as in Eq. (18), where $\alpha_j, \beta_j, \delta_j$ are the cost coefficients of the j^{th} thermal generating unit. P_{THG} is the active power of the j^{th} thermal generating unit and N_{THG} is the number of thermal generating units [1].

$$OF_1 = \sum_{j=1}^{N_{THG}} \alpha_j + \beta_j P_{THG_j} + \delta_j P_{THG_j}^2 \quad (18)$$

(ii) *Minimization of the total cost including valve-point loading effect*: The total cost including VPLE of the thermal generating units can be expressed as in Eq. (19). Here, x_j and

y_j are the coefficients of valve-point loading of the j^{th} thermal generating unit [1].

$$OF_2 = \sum_{j=1}^{N_{THG}} \alpha_j + \beta_j P_{THG_j} + \delta_j P_{THG_j}^2 + \left| x_j \times \sin\left(y_j \times \left(P_{THG_j}^{\min} - P_{THG_j} \right) \right) \right| \quad (19)$$

(iii) *Minimization of active power loss*: It can be mathematically expressed as in Eq. (20), where *i* and *k* are the bus in a power grid [1].

$$OF_3 = \sum_{l=1}^{N_{TL}} G_{m(ik)} \left(V_i^2 + V_k^2 - 2V_i V_k \cos(\theta_{ik}) \right) \quad (20)$$

2.3. Constraints

Equality constraints: The equality constraints defined as follows [6]:

$$P_{G_i} - P_{D_i} + P_{is} = V_i \sum_{k=1}^{N_B} V_k Y_{ik} \cos(\psi_{ik} + \theta_{ik}), \ \forall i \in N_B \quad (21)$$

$$Q_{G_i} - Q_{D_i} + Q_{is} = V_i \sum_{k=1}^{N_B} V_k Y_{ik} \sin(\psi_{ik} + \theta_{ik}), \ \forall i \in N_B \quad (22)$$

where P_{D_i} and Q_{D_i} represent the active and reactive power load demand, respectively. Q_{G_i} and P_{G_i} are the reactive and active power of the i^{th} generating unit, respectively. N_B denotes the number of buses. P_{is} and Q_{is} are the injected active and reactive power by the TCPS of the m^{th} bus, respectively. ψ_{ik} is the angle value of the ik^{th} element of the bus admittance matrix and Y_{ik} is the magnitude of it.

Inequality constraints [6]:

- Generator constraints:

$$\begin{aligned} &P_{THG_i,min} \le P_{THG_i} \le P_{THG_i,max}, \ \forall i \in N_{THG} \\ &Q_{THG_i,min} \le Q_{THG_i} \le Q_{THG_i,max}, \ \forall i \in N_{THG} \\ &V_{G_i,min} \le V_{G_i} \le V_{G_i,max}, \ \forall i \in N_G \end{aligned} \quad (23)$$

- Security constraints:

$$S_{L_i} \le S_{L_i,max}, \ \forall i \in N_{TL} \quad (24)$$

$$V_{L_i,min} \le V_{L_i} \le V_{L_i,max}, \ \forall i \in N_{PQ} \quad (25)$$

- Transformer constraints:

$$T_{j,min} \le T_j \le T_{j,max}, \ \forall j \in N_T \quad (26)$$

- TCSC and TCPS constraints:

$$\tau_{TCSC_r}^{\min} \le \tau_{TCSC_r} \le \tau_{TCSC_r}^{\max}, \ \forall r \in m \quad (27)$$

$$\phi_{TCPS_z}^{\min} \le \phi_{TCPS_z} \le \phi_{TCPS_z}^{maz}, \ \forall z \in p \quad (28)$$

2.4. Constraint handling method

The widely used constraint avoidance technique for solving OPF problems with MHS algorithms is the penalty function method. In this method, the constraint violation values are added to the objective function value in an attempt to modify each infeasible individual's fitness status [28]. Accordingly, the fitness function obtained by adding the constraint violation values to the objective functions discussed in this study using the penalty function method can be expressed with Equation (29). Here, $Fitness_i$ and $F_{penalty,i}$ represent the fitness function and the penalty function of the i^{th} objective function. γ_1, γ_2, γ_3, γ_4, γ_5, and γ_6 are the penalty coefficients.

$$Fitness_i = OF_i + F_{penalty,i} = OF_i + \begin{bmatrix} \gamma_1 \sum_{j=1}^{N_B} \left(P_{G_i} - P_{D_i} + P_{is} \right) + \gamma_2 \sum_{j=1}^{N_B} \left(Q_{G_i} - Q_{D_i} + Q_{is} \right) + \\ \gamma_3 \left(P_{THG_1} - P_{THG_1}^{lim} \right) + \gamma_4 \left(Q_{THG_1} - Q_{THG_1}^{lim} \right) + \\ \gamma_5 \sum_{i=1}^{N_{PQ}} \left(V_{L_i} - V_{L_i}^{lim} \right) + \gamma_6 \sum_{i=1}^{N_{TL}} \left(S_{L_i} - S_{L_i}^{lim} \right) \end{bmatrix} \quad (29)$$

3. PARTIAL REINFORCEMENT OPTIMIZER

Partial Reinforcement Optimizer (PRO) was presented to the literature in 2024 by Taheri et al. [24]. Its inspiration is the Partial Reinforcement Extinction (PRE) theory presented by Ferster and Skinner in 1957 [29]. In the PRO algorithm, there are some basic concepts that need to be defined before examining the structure of it: learner, behavior, population, and response. A learner refers to an individual or animal whose behaviors require training through the PRE theory, and a solution represents the learner. The behavior of the learners is viewed as a solution involving decision variables. The group of learners forms the population. A response refers to the prospective improvement in the objective function.

The PRO algorithm includes three important stages: schedule, stimulation, and reinforcement, which are explained as follows [24]:

Schedule: The idea behind a schedule is to model behaviors for a data structure at different intervals and determine when and how they should be reinforced. Each learner has their own schedule. The scale used in this schedule indicates the priority or score of the behavior of the learner; higher scores mean a better chance of selection in the next round. Furthermore, a stochastic analysis is represented dynamically using the variable-interval scheduling method using Eqns. (30)-(32).

During the search process, the schedule's priorities are sorted from highest to lowest, and a subset of behaviors with the highest priorities for each learner is chosen in each iteration. This subset, which serves as the candidate behaviors utilizing Eq. (31) is made up of the first λ items depending on priority [24].

$$\tau = \frac{FEs}{maxFEs} \quad (30)$$

$$SR = e^{-(1-\tau)} \quad (31)$$

$$\mu \subseteq \{1,2,...,N\} \ | \ \forall j \in \mu, schedule^j \ge schedule^{*,\lambda}, \quad \lambda = \{ \|\mu\| \ | \ \|\mu\| = [u(1, N \times SR)] \} \quad (32)$$

Here, *FEs* and *maxFEs* are the number of fitness function evaluations and the maximum number of FEs, respectively. τ, μ, λ, *SR*, and *N* are the time factor, a subset of behaviors chosen in accordance with scheduling, the selected subset's size, the selection rate, and the number of behaviors, respectively. $schedule^{*}$ and $schedule^{*,\lambda}$ represent the prioritized schedule and the λ^{th} item in the $schedule^{*}$.

Stimulation: Applying operations that change the decision variables of a proposed solution represents an attempt to elicit a response from the behaviors of the learners. In the PRO algorithm, these operations are used to produce the new solutions performed by Eqns. (33)-(35). Here, the decision factors chosen for the student based on the scheduler are represented by the symbols SF_i and $\overline{\beta}$, which stand for the stimulation factor and the mean value of the normalized score, respectively [24].

$$SF_i = \tau + u\left(0,\overline{\beta}\right), \text{ where } \overline{\beta} = \sum_{j\in\mu}\left(\frac{schedule_{i,j}}{\max\left(schedule_i\right)}\right) \tag{33}$$

$$S_i^{\mu} = \begin{cases}\left(x_{best}^{\mu} - x_i^{\mu}\right), & if\ \ rand < 0.5\\ \left(x_i^{\mu} - x_j^{\mu}\right), & otherwise\end{cases} \tag{34}$$

$$x_{i,new}^{\mu} = x_i^{\mu} + SF_i \times S_i^{\mu} \tag{35}$$

Reinforcement: The PRO algorithm uses both positive and negative reinforcement. A certain behavior's score is raised by the application of positive reinforcement. Following improvement during the stimulation phase, the learner's objective function, as stated in Eq. (36) is employed as a response. Here, $schedule_j^{\mu}$ denotes the priority of the chosen behaviors for the j^{th} learner and *RR* is the reinforcement rate [24].

$$schedule_j^{\mu} = schedule_j^{\mu} + \left(schedule_j^{\mu} \times RR\right) \tag{36}$$

Conversely, negative reinforcement is applied in the absence of a response, which lowers the behavior's score and objective function. In the following iteration, only the behaviors with higher scores are chosen to receive reinforcement and stimulation.

$$schedule_j^{\mu} = schedule_j^{\mu} - \left(schedule_j^{\mu} \times RR\right) \tag{37}$$

Rescheduling: This process entails introducing a new timetable for a learner, with the learner consistently receiving negative reinforcement for all behaviors during training. The PRO uses the schedule's standard deviation to assess when rescheduling the learner is necessary. This procedure is implemented using Eqns. (38) and (39) [24].

$$schedule_j = \begin{cases}u(0,1), & if\ \ std\left(schedule_j\right) = 0\\ \text{Do nothing}, & else\end{cases} \tag{38}$$

$$x_j = \begin{cases}u\left(l_b,u_b\right), & if\ \ std\left(schedule_j\right) = 0\\ \text{Do nothing}, & else\end{cases} \tag{39}$$

Here, std is the abbreviation of the standard deviation, l_b and u_b represent the lower and upper bound, respectively. $u(l_b, u_b)$ and $u(0,1)$ denote the randomly uniform distributed values between the (l_b, u_b) and (0,1), respectively.

The flowchart of the PRO algorithm is given in Algorithm-1.

Algorithm-1: The pseudocode of the PRO algorithm [24]

Inputs: Population size, *maxFEs*, *RR*

```
Initialize the population and schedules
Compute the responses of the learners in the population
while FEs<maxFEs
    for j = 1 : N      // for all learners
        Calculate the time parameter using Eq. (30)
        Calculate the SR using Eq. (31)
        Choose λ number of behaviors that have the highest priority
        in schedule j using Eq. (32)
        Update the β and SF based on Eqns. (33) and (34)
        Compute the x_{j,new}^{μ} using Eq. (35)
        Compute the response of j^th learner
              //Apply negative or positive reinforcement
                    based on the response//
        if F(x_{j,new}^{μ}) < F(x_j)
            x_j = x_{j,new}^{μ}
            Employ positive reinforcement using Eq. (36)
        else
            Employ negative reinforcement using Eq. (37)
        end if
        Update the best solution
        Perform the rescheduling process based on Eqns. (38) and
    (39)
        FEs = FEs+1
    end for
end while
```

4. SIMULATION STUDIES AND RESULTS

In this section, the performance of the PRO algorithm was investigated on the IEEE 30-bus test system with TCSC and TCPS devices. The IEEE 30-bus test system consists of six generators at buses 1, 2, 5, 8, 11, and 13 connected to each other with 41 transmission lines and four tap changing transformers. The detailed data of the IEEE 30-bus power network are obtained from [30]. The cost coefficients of the IEEE 30-bus test system are taken from [31].

The main purpose is to optimize the relevant objective function by placing FACTS devices in the optimal locations on the IEEE 30-bus test system. According to this, four different case studies were considered in this study and the summary of them is presented in Table 1. Table 1 provides information on which objective functions and which FACTS devices were used in the case studies.

TABLE I
CASE STUDIES USED IN THE STUDY

Case(s)	Objective Function			FACTS Devices	
	OF_1	OF_2	OF_3	TCSC	TCPS
Case-1	●			●	
Case-2			●	●	
Case-3	●			●	●
Case-4		●		●	●

When Table 1 was examined, two TCSC devices were added to the test system for Case-1 and Case-2, while two TCSC and two TCPS devices were added for Case-3 and Case-4. In these case studies, for each FACTS device, two control variables are assigned that represent the device's location and rating. The variables representing the location are discrete integers indicating the numbers of the branches and

buses, and are obtained by rounding to the nearest integer during the optimization process. Two criteria were taken as the basis for including the FACTS devices to the test system; (*i*) it is not possible to add two FACTS devices in the same location, (*ii*) TCSC and TCPS cannot be added to branches with tap changing transformers. In the case studies, the TCSC agreed to accept a 50% maximum compensation of the installed line reactance, and the lower and upper limits for the phase shifting angle for TCPS were -5° to +5°.

To compare the performance of the PRO algorithm for solving the OPF problem with TCSC and TCPS devices, three up-to-date MHS algorithms, which were the dingo optimization algorithm (DOA), the evolutionary mating algorithm (EMA), and the snow geese algorithm (SGA), were applied to solve all case studies. The parameters of the algorithms are presented in Table 2. The maximum number of fitness function evaluations (*maxFEs*) was set as 30000 for all case studies. Besides, all algorithms performed 30 runs for each case study.

TABLE II
PARAMETERS OF THE ALGORITHMS

Algorithm	Year	Parameter(s)
PRO [24]	2024	Number of population = 30, RR = 0.7
SGA [27]	2024	Number of population = 30,
EMA [26]	2022	Number of population = 30, Cr = 0.8
DOA [25]	2021	Number of population = 100, P=0.5, Q=0.7

4.1. Results of case studies

The PRO, DOA, EMA, and SGA algorithms were applied to solve the four case studies. All algorithms were performed 30 runs for all cases. Accordingly, the results obtained from them were assessed using the minimum (min), mean, maximum (max), and standard deviation (std). In Table 3, the min, max, mean, and std values of all algorithms for all case studies were tabulated.

TABLE III
THE RESULTS OF ALL ALGORITHMS FOR CASE STUDIES

Case(s)	Parameter	PRO	DOA	EMA	SGA
Case-1	Min	**800.7291**	801.1595	800.7766	801.0689
	Mean	**800.8452**	802.7512	802.8931	801.3670
	Max	800.9696	811.8790	818.2737	801.6416
	Std	0.0630	2.6984	5.1248	0.1595
Case-2	Min	**3.2112**	3.2629	3.2202	3.2486
	Mean	**3.2371**	3.5567	3.3629	3.3146
	Max	3.2886	4.6973	4.3397	3.5159
	Std	0.0226	0.3365	0.2809	0.0562
Case-3	Min	**800.6926**	801.3040	800.7419	801.0178
	Mean	**800.8189**	804.4709	801.7940	801.5760
	Max	800.9153	832.9893	803.9818	802.9945
	Std	0.0550	6.2821	1.1489	0.4639
Case-4	Min	**831.6959**	832.9082	831.6988	831.9242
	Mean	**831.9700**	840.2567	833.7043	833.7467
	Max	832.5879	873.2547	845.6788	837.5147
	Std	0.1972	10.5736	3.1634	1.4470

Case-1: *Minimization of the total cost with TCSC*

In this case, the objective was to minimize the total cost given in Eq. (18). The optimal variables obtained from the PRO algorithm for Case-1 are presented in Table 4. It can be seen from Table 4 that all control variables were within the specified lower and upper limit bounds. From Table 3, the total cost values of the PRO, DOA, EMA, and SGA were **800.7291 $/h**, **801.1595 $/h**, **800.7766 $/h**, and **801.0689 $/h**, respectively. That is, the total cost of the PRO was **0.0537%**, **0.0059%**, and **0.0424%** lower than the DOA, EMA, and SGA algorithms, respectively. On the other hand, according to the minimum cost values obtained by the PRO, DOA, EMA, and SGA algorithms, two TCSC devices were placed on branches (5, 4), (5, 27), (5, 4), and (2, 5), respectively.

TABLE IV
THE OPTIMAL VARIABLES OF CASE-1 AND CASE-2 OBTAINED FROM PRO ALGORITHM

Control variables	Max	Min	Case-1	Case-2
P_{THG2} (MW)	80	20	48.7234	79.9958
P_{THG5} (MW)	50	15	21.3335	50.0000
P_{THG8} (MW)	35	10	21.2588	35.0000
P_{THG11} (MW)	30	10	11.9726	29.9720
P_{THG13} (MW)	40	12	12.0000	39.1631
V_1(p.u)	1.1	0.95	1.0835	1.0632
V_2(p.u)	1.1	0.95	1.0642	1.0567
V_5(p.u)	1.1	0.95	1.0332	1.0385
V_8(p.u)	1.1	0.95	1.0378	1.0439
V_{11}(p.u)	1.1	0.95	1.0999	1.0984
V_{13}(p.u)	1.1	0.95	1.0689	1.0700
T_{11}(p.u)	1.1	0.9	1.0200	0.9800
T_{12}(p.u)	1.1	0.9	0.9000	1.0400
T_{15}(p.u)	1.1	0.9	0.9800	1.0000
T_{36}(p.u)	1.1	0.9	0.9600	0.9600
τ_{TCSC1}(%)	50%	0	27.3608	50.0000
τ_{TCSC1}(%)	50%	0	49.9999	23.8130
$TCSC_1$ branch, (con. buses):			**5, (2-5)**	**14, (9-10)**
$TCSC_2$ branch, (con. buses):			**4, (3-4)**	**5, (2-5)**
Total Cost ($/h)			**800.7291**	965.8169
Power loss (MW)			9.0971	**3.2629**
State variables				
P_{THG1} (MW)	200	50	177.2088	52.5321
Q_{THG1} (MVAr)	200	50	3.3587	-0.1942
Q_{THG2} (MVAr)	60	-20	20.5794	3.0035
Q_{THG5} (MVAr)	62.5	-15	25.9925	23.3088
Q_{THG8} (MVAr)	48.7	-15	36.5747	33.7850
Q_{THG11} (MVAr)	40	-10	26.5171	26.9927
Q_{THG13} (MVAr)	44.7	-15	14.5216	20.3510

The voltage values of the load buses are presented in Figure 3. Accordingly, these values were within the upper and lower bounds of the voltage limit. On the other hand, the lower and upper limit value ranges of control variables and reactive powers were given in Table 4. It is seen that the optimal solution values obtained from the PRO algorithm remained within this lower and upper limit ranges.

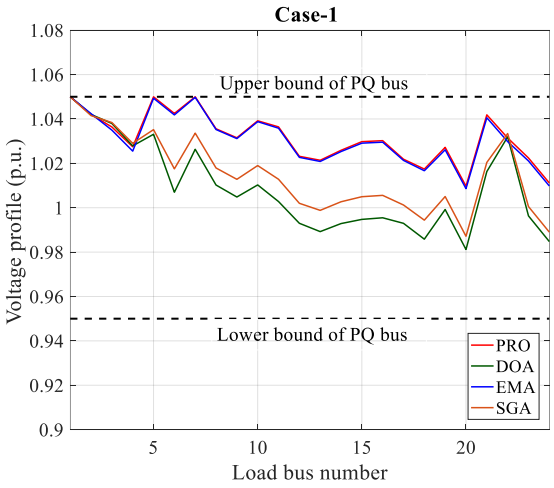


Figure 3. Voltage profiles of the load buses for all algorithms for Case-1

Case-2: *Minimization of the active power loss with TCSC*

The objective here was to place two TCSC devices in the most appropriate place while minimizing active power loss given in Eq. (20). The optimal variables obtained by the PRO algorithm for Case-2 are listed in Table 4. From Table 3, the

PRO algorithm yielded the minimum objective value of **3.2112 p.u.**, which was lower by **1.5863%**, **0.2814%**, and **1.1514%**, than the results of the DOA, EMA, and SGA algorithms, respectively. In this case, two TCSC devices were placed on the branches. Accordingly, TCSC devices were located on branches (14, 5), (25, 5), (7, 5), and (8, 5) according to the optimal results obtained from PRO, DOA, EMA and SGA algorithms, respectively. The voltage profiles of the load buses for all algorithms are shown in Figure 4, where these values remained within the specified voltage limits. Moreover, Table 4 provides the control variable and reactive power's lower and higher limit value ranges. According to these values, when the best solution values of the PRO algorithm were examined, it was seen that they remained within the valid limits.

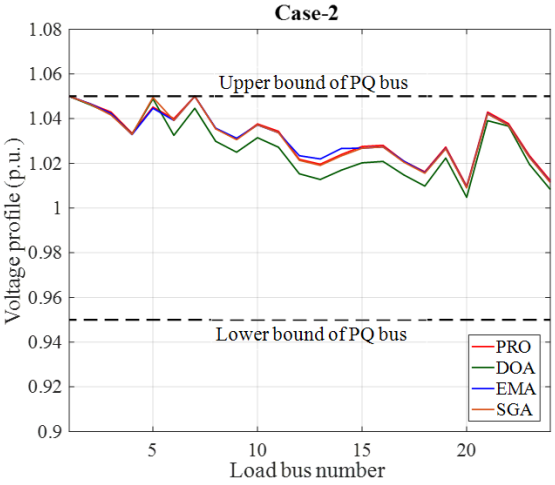


Figure 4. Voltage profiles of the load buses for all algorithms for Case-2

Case-3: *Minimization of the total cost with TCSC and TCPS*

The objective was to minimize the total cost given in Eq. (18), where two TCSC and two TCPS devices were included. The optimal variables identified by the PRO algorithm for Case-3 are tabulated in Table 5. It can be observed that all control variables were within the specified limits. When the results given in Table 3 were evaluated, the results obtained by PRO, DOA, EMA, and SGA were **800.6926 $/h**, **801.3040 $/h**, **800.7419 $/h**, and **801.0178 $/h,** respectively. These results clearly show that the result of PRO was **0.0763%**, **0.0062%**, and **0.0406%** lower than the DOA, EMA, and SGA algorithms, respectively. In this case, two TCSC and two TCPS devices were located on the branches. According to the optimal results obtained from PRO, DOA, EMA, and SGA algorithms, the TCSC devices were placed on branches (5, 13), (14, 13), (2, 1), and (39, 2), respectively. On the other hand, TCPS devices were located on branches (2, 6), (6, 8), (9, 1), and (5, 19) for the optimal results obtained from PRO, DOA, EMA, and SGA algorithms, respectively. The voltage profiles of the load buses for Case-3 are presented in Figure 5. It is demonstrated that the voltage values of the load buses remained within the specified upper and lower bounds. When examining whether the control variables given for Case-3 in Table 5 met the lower and upper limits, it was seen that all variables were within the specified limits.

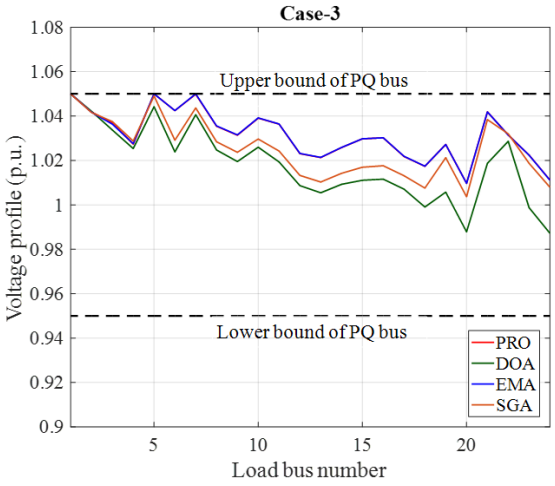


Figure 5. Voltage profiles of the load buses for all algorithms for Case-3

TABLE V
THE OPTIMAL VARIABLES OF CASE-3 AND CASE-4 OBTAINED FROM PRO ALGORITHM

Control variables	**Max.**	**Min.**	**Case-3**	**Case-4**
P_{THG2} (MW)	80	20	48.7512	42.6294
P_{THG5} (MW)	50	15	21.3364	19.3086
P_{THG8} (MW)	35	10	21.1918	10.1692
P_{THG11} (MW)	30	10	11.8607	10.0016
P_{THG13} (MW)	40	12	12.0001	12.0000
V_1 (p.u)	1.1	0.95	1.0827	1.0867
V_2 (p.u)	1.1	0.95	1.0635	1.0649
V_5 (p.u)	1.1	0.95	1.0323	1.0330
V_8 (p.u)	1.1	0.95	1.0363	1.0353
V_{11} (p.u)	1.1	0.95	1.0898	1.1000
V_{13} (p.u)	1.1	0.95	1.0686	1.0687
T_{11} (p.u)	1.1	0.9	1.0400	1.0200
T_{12} (p.u)	1.1	0.9	0.9000	0.9000
T_{15} (p.u)	1.1	0.9	0.9800	0.9800
T_{36} (p.u)	1.1	0.9	0.9600	0.9600
τ_{TCSC1}(%)	50%	0	27.2731	49.5061
τ_{TCSC1}(%)	50%	0	42.2571	29.3302
Φ_{TCPS1}	5	-5	-0.8802	-1.9422
Φ_{TCPS2}	5	-5	0.1184	-0.0201
$TCSC_1$ branch, (con. buses):			**5, (2-5)**	**4, (3-4)**
$TCSC_2$ branch, (con. buses):			**13, (9-11)**	**25, (10-20)**
$TCPS_1$ branch, (con. buses):			**2, (1-3)**	**8, (5-7)**
$TCPS_2$ branch, (con. buses):			**6, (26)**	**27, (10-21)**
Total Cost ($/h)			**800.6926**	**831.6959**
Power loss (MW)			9.0992	10.7089
State variables				
P_{THG1} (MW)	200	50	177.3591	199.9999
Q_{THG1} (MVAr)	200	50	2.9051	4.6583
Q_{THG2} (MVAr)	60	-20	19.6260	23.9679
Q_{THG5} (MVAr)	62.5	-15	25.1286	28.0478
Q_{THG8} (MVAr)	48.7	-15	30.8433	36.8284
Q_{THG11} (MVAr)	40	-10	36.2065	27.0314
Q_{THG13} (MVAr)	44.7	-15	14.3994	14.6355

Case-4: *Minimization of the total fuel cost with valve-point loading effect with TCSC and TCPS*

In this case, the objective was to minimize the total cost with VPLE given in Eq. (19), where two TCSC and two TCPS devices were located on the test system. The optimal variables obtained by the PRO algorithm for Case-4 are presented in Table 5. It is seen that all control variables were within the upper and lower bounds. The results of the PRO, DOA, EMA, and SGA algorithms presented in Table 3 were **831.6959 $/h**, **832.9082 $/h**, **831.6988 $/h**, and **831.9242 $/h**, respectively. Accordingly, the best objective function value was obtained from the PRO algorithm which was lower **0.1455%**, **0.0003%**, and **0.0274%** than the DOA, EMA, and SGA algorithms. In this case, according to the optimal results obtained from PRO, DOA, EMA, and SGA algorithms, the

TCSC devices were located on branches (4, 25), (22, 7), (5, 13), and (24, 29), and TCPS devices were located on branches (8, 27), (22, 34), (2, 1), and (8, 1), respectively. Figure 6 presents the voltage profiles of the load buses for all algorithms, illustrating that the load bus voltage levels stayed within the designated higher and lower boundaries. When the control variables obtained from the PRO algorithm for Case-4 given in Table 5 were evaluated in terms of whether they met the lower and upper limits, it was seen that they were within the limits.

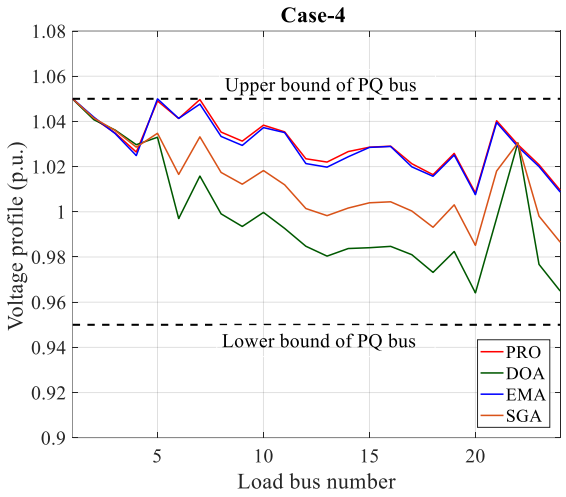


Figure 6. Voltage profiles of the load buses for all algorithms for Case-4

4.2. Convergence analysis

In order to evaluate the search performance of the PRO and its rival algorithms, the convergence graphs of them for all case studies were drawn. Since convergence graphs belong to the run in which the algorithms obtain the best solution, they do not fully reflect the search performance of the algorithms in solving the problem. Therefore, the box-plot graphs were drawn using the fitness values obtained by the algorithms as a result of 30 runs.

The convergence graphs and box-plots of all algorithms for Case-1 are shown in Figure 7. According to Figure 7 (a), the PRO algorithm demonstrated the best convergence performance compared to its competitors. Figure 7 (b) presents the box-plots for all algorithms. It is seen that the PRO algorithm had the smallest minimum, maximum, and median value, and therefore the mean value obtained as a result of 30 runs was the smallest among all algorithms. That is, the PRO algorithm obtained a stable search performance against its competitors.

The convergence graphs and box-plots of all algorithms for Case-2 are presented in Figure 8. From 8 (a), it is evident that the PRO algorithm converged to the lowest objective value among them. According to Figure 8 (b), it is clearly seen that the PRO algorithm had the smallest spread and therefore the smallest median and mean values among all algorithms. On the other hand, the DOA algorithm with the widest spread had the worst search performance.

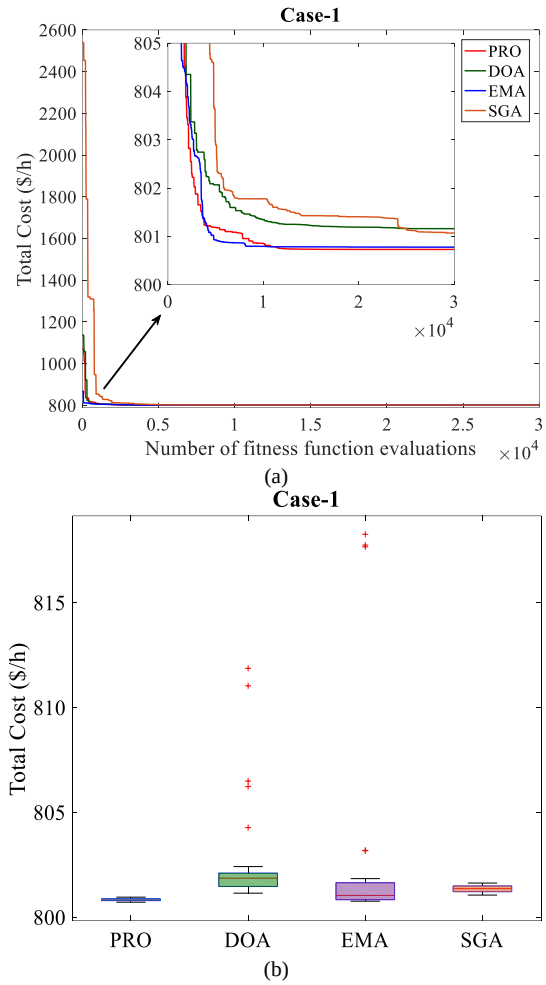


Figure 7. (a) Convergence graphs, (b) box-plots of all algorithms for Case-1

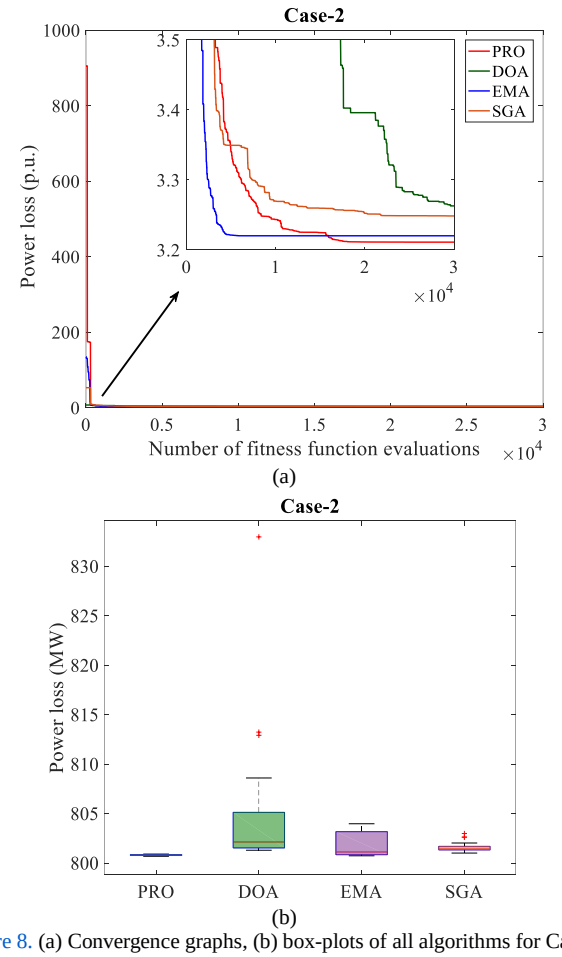


Figure 8. (a) Convergence graphs, (b) box-plots of all algorithms for Case-2

The convergence graphs of all algorithms for Case-3 are shown in Figure 9 (a). Accordingly, the PRO algorithm exhibited the best convergence performance in comparison to its competitors. The box-plots of all algorithms for Case-3 are shown in Figure 9 (b). While the PRO algorithm exhibited the best search performance, the DOA algorithm had the worst search performance among all algorithms.

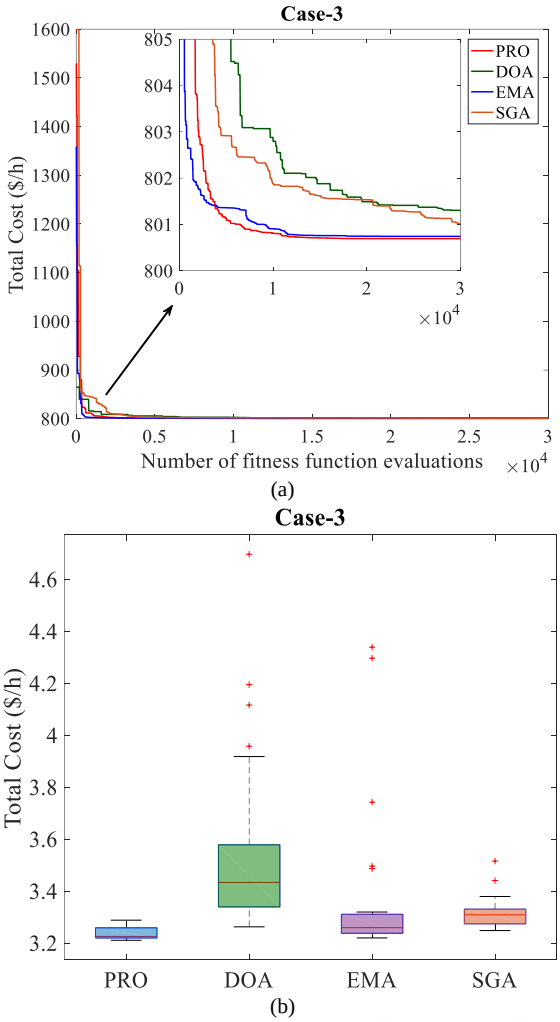
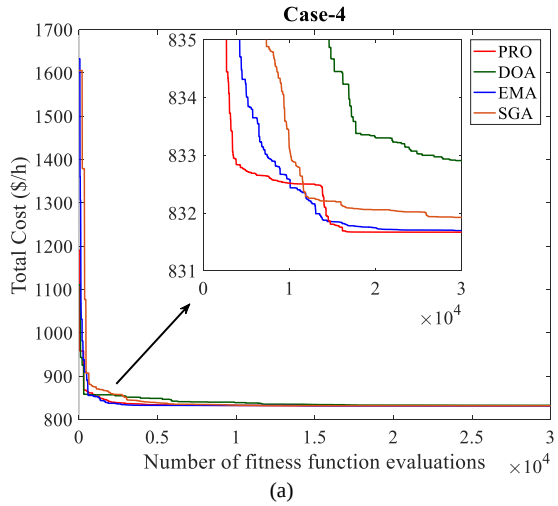


Figure 9. (a) Convergence graphs, (b) box-plots of all algorithms for Case-3



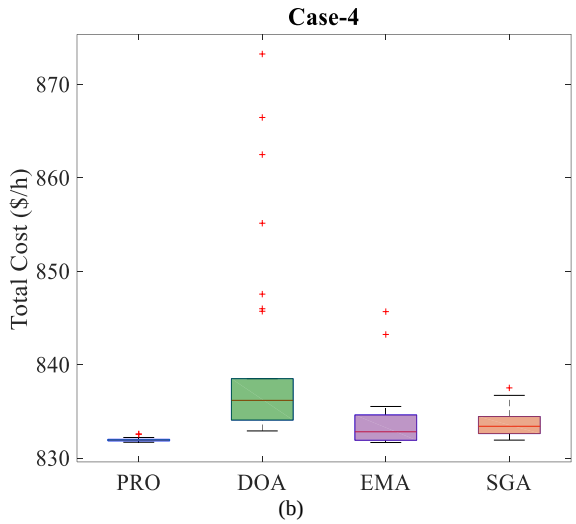


Figure 10. (a) Convergence graphs, (b) box-plots of all algorithms for Case-4

The convergence graphs of all algorithms for Case-4 are shown in Figure 10 (a). It can be seen that the PRO algorithm converged faster than its rivals. Figure 10 (b) presents the box-plots of all algorithms for Case-4. The PRO algorithm had the smallest minimum, maximum, median, and mean values for 30 independent runs. However, it is clear that DOA was the algorithm with the worst search performance due to its widest spread.

4.3. Statistical analysis

To analyze the performance of the algorithms, Friedman and Wilcoxon tests were applied on four case studies. The Friedman test provides broad information regarding the search performances of the algorithms, while the Wilcoxon test is used to determine the significance of the differences between the MHS algorithms, where the algorithms are compared pairwise. The Friedman score values of all algorithms are given in Figure 11. In the calculation of the score values given in Figure 11, the fitness function values obtained from the four case studies of the algorithms were used. From Figure 11, the score values of the PRO, DOA, EMA, and SGA algorithms were 1.2333, 3.6250, 2.3250, and 2.8167, respectively. That is, according to these results, the PRO algorithm ranked first, while the DOA algorithm ranked last.

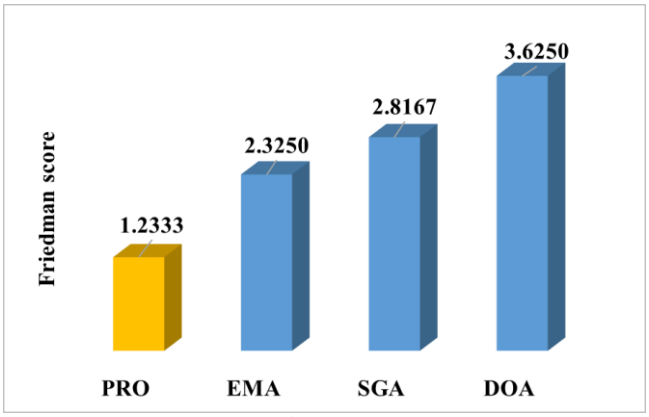


Figure 11. Friedman score values for all algorithms

Besides the Friedman test, the Wilcoxon test was performed to evaluate to performance of the algorithms in pairs. The Wilcoxon test was performed between the PRO algorithm and rival algorithms, and the results are presented

in Table 6. Here, the *Y*-value indicates the score based on the runs lost by the PRO algorithm, while the *Y*+ value represents the score based on the runs won by the PRO algorithm through pairwise comparisons of the results from 30 runs. Accordingly, the PRO algorithm outperformed its rivals in all case studies. To sum up, both Friedman and Wilcoxon test results proved the superior performance of the PRO algorithm over its competitors in four case studies.

TABLE VI
WILCOXON TEST RESULTS FOR ALL CASE STUDIES BETWEEN PRO AND OTHERS

Case(s)	*PRO vs. DOA*			*PRO vs. EMA*			*PRO vs. SGA*		
	Y-	*Y+*	*p-value*	*Y-*	*Y+*	*p-value*	*Y-*	*Y+*	*p-value*
Case-1	0	465	1.73e-06	25	440	1.97e-05	0	465	1.73e-06
Case-2	0	465	1.73e-06	32	433	3.72e-05	0	465	1.73e-06
Case-3	0	465	1.73e-06	51	414	1.89e-05	7	458	3.52e-06
Case-4	0	465	1.73e-06	61	404	1.29e-04	0	465	1.73e-06

5. CONCLUSION

This study presents the application of the partial reinforcement optimizer (PRO) algorithm for the solution of the OPF problem incorporating the optimal placement and sizing of the TCSC and TCPS devices. To solve the OPF problem, including the optimal placement of TCSC and TCPS devices, three objective functions were discussed. These were the minimization of total cost, the minimization of total cost with VPLED, and the minimization of active power loss. Using these objective functions and the TCSC and TCPS devices, four case studies were created. For the solution of these case studies, the performance of the PRO algorithm was compared with the up-to-date MHS algorithms, including DOA, EMA, and SGA. According to results of all algorithms, the PRO algorithm achieved a 0.0059%, 0.2814%, 0.0062 %, 0.0003% improvement over its closest competitor, the EMA algorithm, for Case-1, Case-2, Case-3, and Case-4, respectively. On the other hand, for the evaluation of the performance of all algorithms over the solution of these case studies, the statistical analysis methods were used. According to the results of the Friedman score, the PRO algorithm ranked first with 1.2333 score value against its competitors. Moreover, the Wilcoxon test results showed that the PRO algorithm outperformed its competitors in all case studies.

To sum up, the superiority of the PRO algorithm in solving the OPF problem incorporating the optimal placement and sizing of the TCSC and TCPS devices has been confirmed. In future studies, the PRO algorithm can be used to solve complex real-world problems.

REFERENCES

[1] S. Duman, "Solution of the optimal power flow problem considering FACTS devices by using lightning search algorithm", Iranian Journal of Science and Technology, Transactions of Electrical Engineering, vol. 43, pp. 969-997, Dec. 2019.

[2] M. H. Sulaiman and Z. Mustaffa, "Optimal placement and sizing of FACTS devices for optimal power flow using metaheuristic optimizers", Results in Control and Optimization, vol. 8, 100145, Sep. 2022.

[3] A. Mukherjee and V. Mukherjee, "Solution of optimal power flow with FACTS devices using a novel oppositional krill herd algorithm", International Journal of Electrical Power & Energy Systems, vol. 78, pp. 700-714, June 2016.

[4] E. Naderi, M. Pourakbari-Kasmaei, and H. Abdi, "An efficient particle swarm optimization algorithm to solve optimal power flow problem integrated with FACTS devices", Applied Soft Computing, vol. 80, pp. 243-262, July 2019.

[5] J. B. Edward, N. Rajasekar, K. Sathiyasekar, N. Senthilnathan, and R. Sarjila, "An enhanced bacterial foraging algorithm approach for optimal power flow problem including FACTS devices considering system loadability", ISA Transactions, vol. 52, no. 5, pp. 622-628, Sep. 2013.

[6] D. Prasad and V. Mukherjee, "A novel symbiotic organisms search algorithm for optimal power flow of power system with FACTS devices", Engineering Science and Technology, an International Journal, vol. 19, no. 1, pp. 79-89, March 2016.

[7] B. Mahdad, "Improvement optimal power flow solution considering SVC and TCSC controllers using new partitioned ant lion algorithm", Electrical Engineering, vol. 102, no. 4, pp. 2655-2672, Dec. 2020.

[8] M. Ebeed, A. Mostafa, M. M. Aly, F. Jurado, and S. Kamel, "Stochastic optimal power flow analysis of power systems with wind/PV/TCSC using a developed Runge Kutta optimizer", International Journal of Electrical Power & Energy Systems, vol. 152, 109250, Oct. 2023.

[9] M. A. Taher, S. Kamel, F. Jurado, and M. Ebeed, "Optimal power flow solution incorporating a simplified UPFC model using lightning attachment procedure optimization", International Transactions on Electrical Energy Systems, vol. 30, no. 1, e12170, Jan. 2020.

[10] P. P. Biswas, P. Arora, R. Mallipeddi, P. N. Suganthan, and B. K. Panigrahi, "Optimal placement and sizing of FACTS devices for optimal power flow in a wind power integrated electrical network", Neural Computing and Applications, vol. 33, pp. 6753-6774, June 2021.

[11] A. A. Ahmad and R. Sirjani, "Optimal placement and sizing of multi-type FACTS devices in power systems using metaheuristic optimisation techniques: An updated review", Ain Shams Engineering Journal, vol. 11, no. 3, pp. 611-628, Sep. 2020.

[12] S. R. Inkollu and V. R. Kota, "Optimal setting of FACTS devices for voltage stability improvement using PSO adaptive GSA hybrid algorithm", Engineering Science and Technology, An International Journal, vol. 19, no. 3, pp. 1166-1176, Sep. 2016.

[13] W. S. Sakr, R. A. El-Sehiemy, and A. M. Azmy, "Optimal allocation of TCSCs by adaptive DE algorithm", IET Generation, Transmission & Distribution, vol. 10, no. 15, pp. 3844-3854, Nov. 2016.

[14] S. R. Inkollu and V. R. Kota, "Optimal setting of FACTS devices for voltage stability improvement using PSO adaptive GSA hybrid algorithm", Engineering Science and Technology, An International Journal, vol. 19, no. 3, pp. 1166-1176, Sep. 2016.

[15] O. Ziaee and F. F. Choobineh, "Optimal location-allocation of TCSC devices on a transmission network", IEEE Transactions on Power Systems, vol. 32, no. 1, pp. 94-102, Jan. 2017.

[16] S. Raj and B. Bhattacharyya, "Optimal placement of TCSC and SVC for reactive power planning using Whale optimization algorithm", Swarm and Evolutionary Computation, vol. 40, pp. 131-143, June 2018.

[17] R. Agrawal, S. K. Bharadwaj, and D. P. Kothari, "Population based evolutionary optimization techniques for optimal allocation and sizing of Thyristor Controlled Series Capacitor", Journal of Electrical Systems and Information Technology, vol. 5, no. 3, pp. 484-501, Dec. 2018.

[18] N. H. Khan, Y. Wang, D. Tian, R. Jamal, S. Iqbal, M. A. A. Saif, and M. Ebeed, "A novel modified lightning attachment procedure optimization technique for optimal allocation of the FACTS devices in power systems", IEEE Access, vol. 9, pp. 47976-47997, Feb. 2021.

[19] K. Nusair, F. Alasali, A. Hayajneh, and W. Holderbaum, "Optimal placement of FACTS devices and power-flow solutions for a power network system integrated with stochastic renewable energy resources using new metaheuristic optimization techniques", International Journal of Energy Research, vol. 45, no. 13, pp. 18786-18809, Oct. 2021.

[20] A. A. Mohamed, S. Kamel, M. H. Hassan, M. I. Mosaad, and M. Aljohani, "Optimal power flow analysis based on hybrid gradient-based optimizer with moth–flame optimization algorithm considering optimal placement and sizing of FACTS/wind power", Mathematics, vol. 10, no. 3, pp. 361, Jan. 2022.

[21] M. H. Hassan, F. Daqaq, , S. Kamel, A. G. Hussien, and H. M. Zawbaa, "An enhanced hunter-prey optimization for optimal power flow with FACTS devices and wind power integration", IET Generation, Transmission & Distribution, vol. 17, no. 14, pp. 3115-3139, July 2023.

[22] A. Awad, S. Kamel, M. H. Hassan, and H. Zeinoddini-Meymand, "Optimal allocation of flexible AC transmission system (FACTS) for wind turbines integrated power system", Energy Science & Engineering, vol. 12, no. 1, pp. 181-200, Jan. 2024.

[23] A. A. Mohamed, S. Kamel, M. H. Hassan, and H. Zeinoddini-Meymand, "CAVOA: A chaotic optimization algorithm for optimal power flow with facts devices and stochastic wind power generation", IET Generation, Transmission & Distribution, vol. 18, no. 1, pp. 121-144, Jan. 2024.

[24] S. Mahapatra, N. Malik, S. Raj, and M. K. Srinivasan, "Constrained optimal power flow and optimal TCSC allocation using hybrid cuckoo search and ant lion optimizer", International Journal of System Assurance Engineering and Management, vol. 13, pp. 721-734, Apr. 2022.

[25] A. Taheri, K. RahimiZadeh, A. Beheshti, J. Baumbach, R. V. Rao, S. Mirjalili, and A. H. Gandomi, "Partial reinforcement optimizer: An evolutionary optimization algorithm", Expert Systems with Applications, vol. 238, 122070, March 2024.

[26] H. Peraza-Vázquez, A. F. Peña-Delgado, G. Echavarría-Castillo, A. B. Morales-Cepeda, J. Velasco-Álvarez, and F. Ruiz-Perez, "A bio-inspired method for engineering design optimization inspired by dingoes hunting strategies", Mathematical Problems in Engineering, vol. 2021, 9107547, 1-19, Sept. 2021.

[27] M. H. Sulaiman, Z. Mustaffa, M. M. Saari, H. Daniyal, and S. Mirjalili, "Evolutionary mating algorithm", Neural Computing and Applications, vol. 35, no. 1, pp. 487-516, Jan. 2023.

[28] A. Q. Tian, F. F. Liu, and H. X. Lv, "Snow Geese Algorithm: A novel migration-inspired meta-heuristic algorithm for constrained engineering optimization problems", Applied Mathematical Modelling, vol. 126, pp. 327-347, Feb. 2024.

[29] C. A. C. Coello, "Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art", Computer Methods in Applied Mechanics and Engineering, vol. 191, no. 11-12, pp. 1245-1287, Jan. 2002.

[30] C. B. Ferster and B. F. Skinner, Schedules of reinforcement. Appleton-Century-Crofts, 1957.

[31] The data of the IEEE 30-bus test system, http://labs.ece.uw.edu/pstca/pf30/pg_tca30bus.htm

[32] P. P. Biswas, P. N. Suganthan, R. Mallipeddi, and G. A. Amaratunga, "Optimal power flow solutions using differential evolution algorithm integrated with effective constraint handling techniques", Engineering Applications of Artificial Intelligence, vol. 68, pp. 81-100, Feb. 2018.

BIOGRAPHY

Burçin Özkaya obtained his BSc degree in Electrical and Electronics Engineering from Eskişehir Osmangazi University (ESOGU) in 2014. She received MSc. in Electrical and Electronics Engineering from the Süleyman Demirel University in 2018, and PhD degree in Electrical and Electronics Engineering from Düzce University in 2022. She worked as a Research Assistant with the Department of Electrical and Electronics Engineering at Suleyman Demirel University from 2015 to 2018, and with Department of Electrical and Electronics Engineering at Isparta University of Applied Sciences from 2018 to April 2023. She has been working as an Assistant Professor since April 2023 at Department of Electrical Engineering at Bandırma Onyedi Eylül University. Her research interests include power systems applications, optimization, meta-heuristic algorithms, and artificial intelligence.

European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol. 14, No.1, 2024

Research Article

Performance Comparison of Machine Learning Methods in Discovery of BACE-1 Inhibitors in Alzheimer's Disease Therapy

Bihter Das[1], **Suat Toraman**[2*]

[1]Firat University, Technology Faculty, Software Engineering Department, 23119, Elazig, Turkey. (e-mail: bihterdas@firat.edu.tr).
[2*] Firat University, School of Civil Aviation Department, 23119, Elazig, Turkey. (e-mail: storaman@firat.edu.tr). +904242370000-8978

ARTICLE INFO

Received: March, 20, 2024
Revised: May., 06. 2024
Accepted: May, 08. 2024

Keywords:
Drug Interaction, Alzheimer Disease, BACE-1 inhibitors, Machine Learning, Gradient Boosting Machine (GBM)

Corresponding author: *Suat Toraman*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1455786

ABSTRACT

Alzheimer's disease poses a significant challenge in the realm of neurodegenerative disorders, necessitating effective therapeutic interventions. One promising approach involves the discovery of β-secretase 1 (BACE-1) inhibitors, pivotal in mitigating amyloid-β peptide accumulation, a hallmark of AD pathology. In this study, we compare the performance of three prominent machine learning methods, namely Gradient Boosting Machine, Random Forest, and Support Vector Machine in the discovery of BACE-1 inhibitors. Leveraging the BACE dataset sourced from MoleculeNet, comprising quantitative and qualitative binding results of compounds, we explored the classification efficacy of these methods. Our experimental results reveal distinct precision, recall, and accuracy metrics for each method, showcasing RF with precision and accuracy scores of 1.00 and 99.67%, respectively, followed by GBM and SVM. Furthermore, feature importance analysis underscores pIC50 as the most influential attribute across all methods, emphasizing its pivotal role in classifying BACE-1 inhibitors. Additionally, RF prioritizes Estate as the second most important feature, while AlogP emerges as GBM's secondary significant attribute. These findings shed light on the efficacy of machine learning techniques in identifying potential therapeutics for AD, offering insights into feature importance variations among methods and highlighting the significance of diverse molecular descriptors in drug discovery.

1. INTRODUCTION

Alzheimer's disease (AD) is a prevalent chronic neurodegeneration-related condition affecting the elderly, characterized by the gradual decline of cognitive abilities [1]. The disease is characterized by symptoms such as memory loss, cognitive impairments, and personality changes, and seriously affects patients' quality of life. Alzheimer's disease, which affects millions of people worldwide, has been one of the priority areas of research for the medical and scientific community [2,3]. Abnormal accumulation of beta-amyloid peptide is a key component of the pathophysiology of AD. One of the enzymes enabling this peptide's formation is a protease known as human β-secretase 1 (BACE-1) [4]. BACE-1 performs one of the cuts on the amyloid precursor protein (APP), which is the starting material for the beta-amyloid peptide. Therefore, BACE-1 inhibitors have garnered considerable interest as a promising therapeutic target for addressing AD [5-8]. There are various studies in the literature regarding the use of BACE1 inhibitors for the Alzheimer's disease. Wang et al. studied how β-saron affects the proteins expression associated with AD. Findings indicate that β-Asarone can reduce Aβ accumulation and improve the autophagy process by regulating the protein expression of AD. This approach can be considered a potential strategy for treating AD [9]. Chang et al. devised a sensor to detect β-Secretase (BACE1), an enzyme pivotal in generating β-amyloid, a key factor in the onset of AD. The sensor has a design containing a ferrocene probe and an aldehyde group, thus producing an electrochemical signal by the N terminus released when the APP peptide is cleaved by BACE-1. This developed sensor can detect BACE1 activity by providing a sensitive, low-cost, and easy detection method [10]. Kalaimathi et al. revealed that four marine-derived compounds derived from cyanobacteria could be considered potential BACE1 inhibitors in treating Alzheimer's disease.

These findings may open a promising avenue for the development of new therapeutics [11]. Bhanukiran et. al. performed various physicochemical, reactivity, and stability assays to evaluate the vaccine derivative VA10, a potential BACE1 inhibitor that could be used to treat Alzheimer's disease. Their studies revealed that VA10 was effective as a BACE1 inhibitor and showed no toxicity in mice. These

findings support the evaluation of VA10 as a potential agent in the treatment of AD [12]. Nakano et al. discovered that smoking increases the risk of AD and that nicotinic acetylcholine receptor (nAChR) activation in neurons increases the production of amyloid β (Aβ) in their study. They found that nAChR activation triggers Aβ production by increasing the transcription of the BACE1 gene and increases Aβ accumulation [13]. Abraham et. al. used a multiscale modeling approach to identify minimal congenital characters β-secretase (BACE1) and α-secretase (ADAM10) transmembrane regions, which participate significantly in the amyloid cascade of AD. Our results showed that membrane composition affects the character of the transmembrane domains of BACE1 and ADAM10, providing support for speculations about the role of membrane domains in the etiology of Alzheimer's disease [14]. Vincent et al. investigated the potential link between AD and schizophrenia, highlighting the shared frontotemporal anomaly and increased risk of co-morbid dementia and psychosis. They focused on the molecular level, specifically, the metabolism of β-amyloid precursor protein and neuregulin 1 by β-site APP cleaving enzyme 1, proposing a model to explain the accompanying symptoms [15]. Dominko et al. investigated the proteolysis and distribution of BACE1 substrates, sex6 and Sex6L, in Alzheimer's disease (AD) using the 5xFAD mouse model. they found that while bace1 accumulation in ad brains did not affect the proteolysis of sez6 and sez6l, their distribution was altered in the peri-plaque regions. this suggests different localization and/or function of these substrates compared to app and bace1. The study highlights the potential role of aß-targeted therapies in mitigating the accumulation and modified distribution of BACE-1 and its substrates, along with APP in AD [16].

This study will examine how BACE-1 inhibitors can be evaluated and designed as potential drug candidates targeting Alzheimer's disease. First, we will describe the biochemical and structural properties of BACE-1 and its role in AD pathogenesis. Next, we discuss the various approaches used in the design and identification of BACE-1 inhibitor candidates. Finally, we will summarize current research findings and focus on future studies, highlighting the promise of BACE-1 inhibitors for AD therapy.

This paper aims to contribute to the development of innovative and effective therapeutic strategies for the treatment of Alzheimer's disease. Deepening current knowledge of the role of BACE-1 inhibitors in drug development and promoting progress in this field could be an important step in the fight against AD.

The primary contributions of this paper are as follows:

- Evaluation of BACE-1 inhibitors for identifying potential Alzheimer's disease treatment drug candidates.
- Discussion of various approaches used in the design and identification of BACE-1.
- Analysis of various molecular properties used in the classification of BACE-1 inhibitors and determination of which property has the most significant impact on classification.
- Comparison of the results of classification using machine learning methods such as Gradient Boosting Machine (GBM), Random Forest (RF), and Support Vector Machine (SVM).
- The aim of the study to contribute to the development of innovative therapeutic strategies for Alzheimer's disease treatment and to deepen the understanding of the role of BACE-1 inhibitors in the drug development process.

The remainder of the paper is structured as follows: Section 2 elaborates on the Materials and Methods employed in the experiment. Section 3 delineates the experimental results and engages in discussion. Finally, section 4 provides concluding remarks in the form of the Conclusion.

2. MATERIALS AND METHODS

This paper conducts a study on the evaluation and classification of potential drug candidates for the in the treatment of AD. The performances of methods such as GBM, RF, and SVM are discussed for the design of BACE-1 inhibitors that can be used in the treatment of this disease. Ultimately, the study summarizes current research findings, aiming to underscore the potential of BACE-1 inhibitors in treatment of AD. The steps of the experimental study to achieve this goal will be explained in this section.

2.1. Dataset

This study utilizes a dataset containing quantitative (IC50) and qualitative (binary label) binding results of inhibitors aimed at human β-secretase 1 (BACE-1).In addition to molecular properties, the data set also includes pIC50 values, which indicate the binding strength of inhibitors. The BACE dataset was retrieved from the MolecularNet database and contains 1513 compounds [17]. The dataset was split 80/10/10 to create training, validation, and testing subsets. This division is made to ensure that the model is evaluated correctly during the training, validation and testing phases. Table 1 illustrates a sample of the dataset utilized in the study.

2.2. Machine Learning Methods for Classification

This section will introduce the machine learning (ML) methods used to solve the classification problem on the BACE dataset. The data set contains the properties of each compound, along with a label indicating whether each compound is a BACE-1 inhibitor. This is considered a classification problem and various ML algorithms can be used to solve this problem. In this section, we will explain the basic principles and applications of popular classification algorithms such as SVM, RF and GBM. We will focus on how

2.2. Machine Learning Methods for Classification

This section will introduce the machine learning (ML) methods used to solve the classification problem on the BACE dataset. The data set contains the properties of each compound, along with a label indicating whether each compound is a BACE-1 inhibitor. This is considered a classification problem and various ML algorithms can be used to solve this problem. In this section, we will explain the basic principles and applications of popular classification algorithms such as SVM, RF and GBM. We will focus on how each algorithm works, what types of data it performs best with, and the metrics used to evaluate classification results.

TABLE I
A SAMPLE OF THE USED DATASET

Mol	CID	Class	pIC50	MW	AlogP	HBA	HBD	RB	Heavy Atom Count	Chiral Center Count	Chiral Center CountAll Possible	Ring Count	PSA	Estate	MR	Polar
O1CC[C@@H](NC(=O)[C@@H](Cc2cc3cc(ccc3nc2N)-c2ccccc2C)C)CC1(C)C	BACE_1	1	9.1549015	431.56979	4.4014001	3	2	5	32	2	2	4	77.239998	67.251999	129.9039	58.397999
Fc1cc(cc(F)c1)C[C@H](NC(=O)[C@@H](N1CC[C@](NC(=O)C)(CC(C)C)C1=O)CCc1ccccc1)[C@H](O)[C@@H]1[NH2+]C[C@H](OCCC)C1	BACE_2	1	8.8538723	657.81073	2.6412001	5	4	16	47	6	6	4	124.58	115.417	173.6176	76.254997
S1(=O)(=O)N(c2cc(cc3c2n(cc3CC)CC1)C(=O)N[C@H]([C@H](O)C[NH2+]Cc1cc(OC)ccc1)Cc1ccccc1)C	BACE_3	1	8.6989698	591.74091	2.5499001	4	3	11	42	2	3	5	125.86	96.585999	160.12421	75.639
O=C(NCC1CCCCC1)[C@@H](Cc1cc2cc(ccc2nc1N)-c1ccccc1C)CCC	BACE_4	1	7.9586072	443.6236	7.0788999	2	2	8	33	1	1	4	68.010002	66.001999	137.3194	61.431
S1(=O)(=O)C[C@@H](Cc2cc3c([nH]cc3CC(F)F)cc2)[C@H](O)[C@@H]([NH2+]Cc2cc(ccc2)C(C)(C)C)C1	BACE_5	1	7.2596374	505.6402	2.7595999	2	3	8	35	3	4	4	95.150002	87.500999	132.3071	60.943001
O1[C@@H]2COCC[C@@]2(N=C1N)c1cc(ccc1)-c1cncnc1	BACE_6	0	4.0030508	296.32379	0.87709999	5	0	2	22	2	2	4	82.620003	47.750999	80.689301	37.137001
[NH+]=1[C@](N=C(c2ccccc2)C=1N)(C)c1cc(ccc1)-c1cccnc1	BACE_7	0	4	327.4024	1.4339	2	0	3	25	1	1	4	65.239998	49.251999	97.589897	48.525002
n1cccc(NCc2cc(ccc2)-c2cccnc2)c1N	BACE_8	0	4	276.33569	2.2346001	2	2	4	21	0	0	3	63.830002	44.334999	86.700699	41.577
O1CCC(OC(=O)[C@@H]2[NH2+]C[C@]3(C2)c2c(NC3=O)cccc2)CC1	BACE_9	0	3.9430952	317.35959	-1.0807	4	2	3	23	2	2	4	81.239998	53.084	80.653801	36.096001
O=C1NC(=NC(=C1)CCc1cc2[nH]ccc2cc1)N	BACE_10	0	3.8860567	254.2872	1.7381001	2	3	3	19	0	0	3	87.559998	44.001999	75.263802	34.915001

2.2. Machine Learning Methods for Classification

This section will introduce the machine learning (ML) methods used to solve the classification problem on the BACE dataset. The data set contains the properties of each compound, along with a label indicating whether each compound is a BACE-1 inhibitor. This is considered a classification problem and various ML algorithms can be used to solve this problem. In this section, we will explain the basic principles and applications of popular classification algorithms such as SVM, RF and GBM. We will focus on how each algorithm works, what types of data it performs best with, and the metrics used to evaluate classification results. We will also cover important

topics such as the advantages and disadvantages of each algorithm, their application areas, and parameter tuning. Finally, we will compare the performance of these algorithms on the BACE-1 inhibitor classification problem and analyze the results. This will summarize the paper's main findings and evaluate the effectiveness of ML methods for BACE-1 inhibitor detection.

2.2.1 Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a classification and regression machine learning algorithm utilized for solving classification and regression tasks [18]. SVM creates a decision boundary to classify data points, and this provides the best separation of data points (Figure 1). The main purpose of SVM is to accurately classify new observations by creating a decision boundary between different classes. SVM can solve linear and nonlinear classification problems and generally performs well with high-dimensional datasets [19]. SVM has various important applications in drug discovery and development, such as biological activity prediction, drug mechanism of action analysis, drug toxicity prediction, and drug resistance analysis. Models based on this algorithm can be used for important decisions such as classification and prioritization of potential drug candidates [20]. Additionally, it has been shown that SVM can be used effectively in critical areas such as determining the mechanisms of action of drugs on target proteins and evaluating the safety profiles of drugs. Finally, SVM can also be used as an important tool for predicting the treatment response of medically resistant diseases, which can help develop personalized medicine approaches and optimize treatment strategies [21].

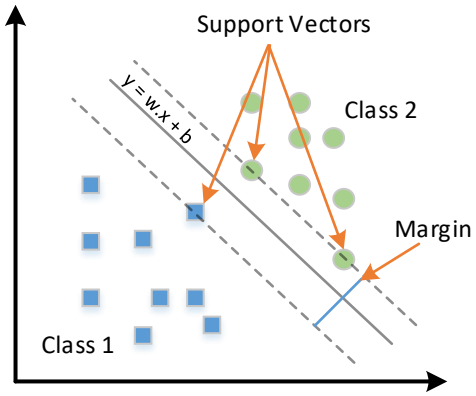


Figure 1. Support Vector Machine

2.2.2 Random Forest (RF)

RF: Random Forest (RF) is a popular ensemble learning technique employed for classification and regression tasks by employing a group of decision trees, known as a forest of trees [22] (Figure 2). Each decision tree splits the dataset by building trees on randomly selected features. Then, each tree makes a prediction and the results are combined to produce the final prediction. RF combats overfitting, works effectively with high-dimensional datasets and generally provides high accuracy [23]. Random Forest (RF) is a potent machine-learning technique extensively employed in drug discovery and development. RF has a variety of important applications, from predicting the biological activity of a given compound to drug discovery and design, drug mechanism of action analysis, and drug toxicity prediction. Therefore, RF is a valuable tool for researchers in the pharmaceutical industry and plays an important role in discovering and evaluating new drug candidates [24,25].

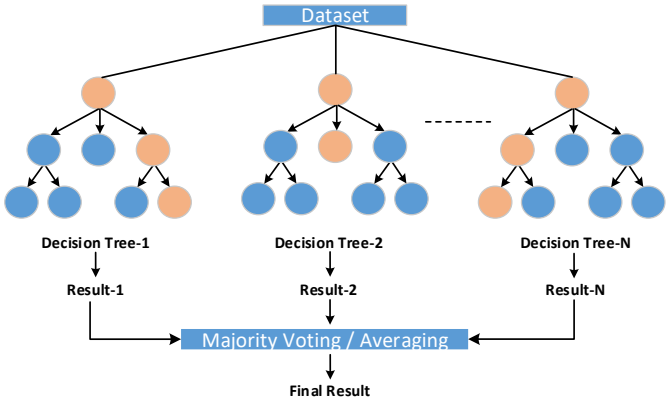


Figure 2. Random Forest

2.2.3 Gradient Boosting Machine (GBM)

Gradient Boosting Machine (GBM) is an ensemble learning technique that creates a strong predictor by combining many weak predictors (commonly known as decision trees) (Figure 3). GBM successively adds weak predictors, and each predictor focuses on correcting the errors of previous predictors. In this way, GBM can adapt to the complexity of the dataset and provide high accuracy. GBM also reduces overfitting and generally produces competitive performance [26,27]. GBM is less known or less used in drug studies than some other algorithms.

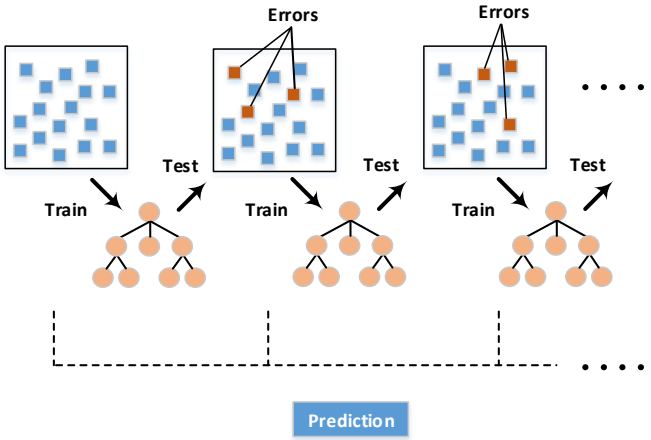


Figure 3. Gradient Boosting Machine

3. RESULTS AND DISCUSSION

This section presents the experimental results of three different machine learning methods used for the classification of the BACE-1 drug dataset and a discussion of these results. Classification results obtained using RF, SVM, and GBM methods were examined. The performance of each method was evaluated in terms of precision, sensitivity (Recall), and accuracy (Accuracy) criteria. Additionally, an in-depth discussion of possible factors that may affect the success of these methods is presented. This discussion will help us better understand the results obtained and their impact on drug discovery and development processes. In this study, a significant analysis was also performed to determine the most important features in classifying BACE-1 drug data. This analysis focused on identifying the features that contribute most to the classification performance of each machine learning method. The results obtained evaluate the contribution

of each feature to the classification process and reveal which features are more decisive in predicting drug effectiveness.

TABLE II.
HYPERPARAMETERS FOR ALL THREE METHODS

RF	"max_depth": [5, 8, 10], "max_features" : [2, 5, 10], "n_estimators" : [200, 500, 1000], "min_samples_split" : [2, 10, 80]
SVM	"C": [0.1, 10, 100, 1000], "kernel": ["linear", "rbf", "polynomial"]
GBM	"max_depth" : [3, 5, 8], "n_estimators" : [50, 100, 200], "subsample" : [.5, 1, 3]

Identifying these important properties can help identify priority targets and identify more effective drug candidates in drug discovery and design processes. Table 2 gives the parameters used and the ranges examined for all three models. Table 3 presents the performance results of all three methods in classifying BACE-1 data. As seen in Table 3, the RF method shows a very effective performance with high precision and high recall for classification. The SVM method also performs quite well, with high precision and recall, but slightly lower compared to the RF method. The GBM method shows a very successful performance with high sensitivity, recall, and accuracy for classification, and achieved slightly higher accuracy than the other two methods. These results show that machine learning methods such as RF, SVM, and GBM are effective in BACE drug data classification. Figure 4 shows the results of the confusion matrix for three methods.

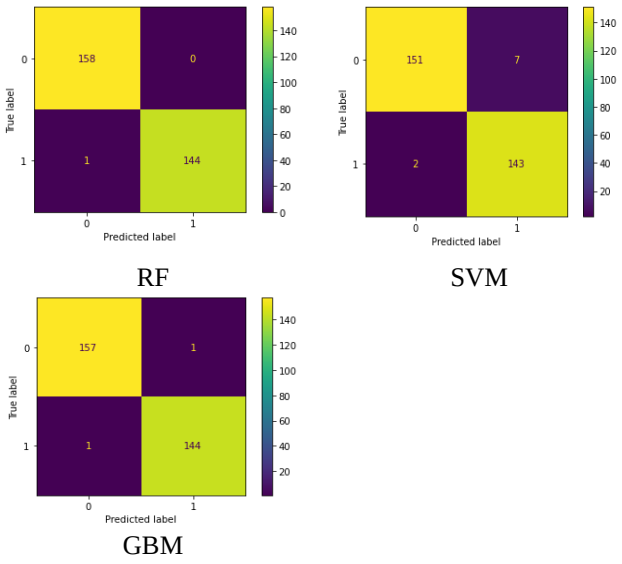


Figure 4. Results of the confusion matrix for three methods

TABLE III.
THE PERFORMANCE RESULTS FOR THREE METHODS

	Precision (%)	Recall (%)	Accuracy (%)
RF	100.00	99.31	99.67
SVM	95.33	98.62	97.03
GBM	99.31	99.31	99.34

Figures 5 and 6 give the most important chemical properties in classification for GBM and RF, respectively. These results show that it is meaningful to determine the PlC50 feature as the most important feature in classifying BACE drug data. Since PlC50 is a metric that measures the biological activity of a compound, this property is thought to directly reflect drug efficacy. Therefore, PlC50 emerged as the most important feature, indicating that it is a determining factor in the classification of drug candidates. The second important feature for RF was identified as Estate, indicating that this feature may play a decisive role in evaluating drug effectiveness. Estate is a property that represents the electronic properties of a compound, and this property can have an impact on the biological activity of the compound. For GBM, the second important feature is determined as AlogP, a feature that reflects the hydrophobicity and lipophilicity of the compound. The emergence of AlogP as a significant feature indicates that it may influence critical factors such as the cellular penetration of the compound and its ability to bind to the target protein. These results suggest that different machine learning methods may emphasize different features and that different mechanisms may be important in determining drug efficacy. Weights assigned to features can only be used in the case of linear kernel in SVM. This cannot be used in other kernel functions (rbf, polynomial). For this reason, feature importance could not be calculated for SVM.

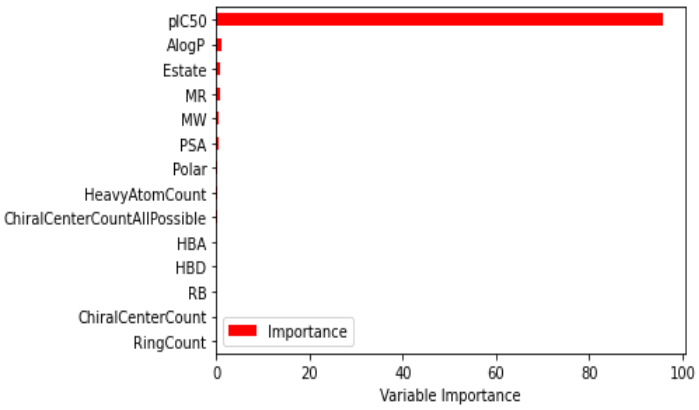


Figure 5. GBM feature importance

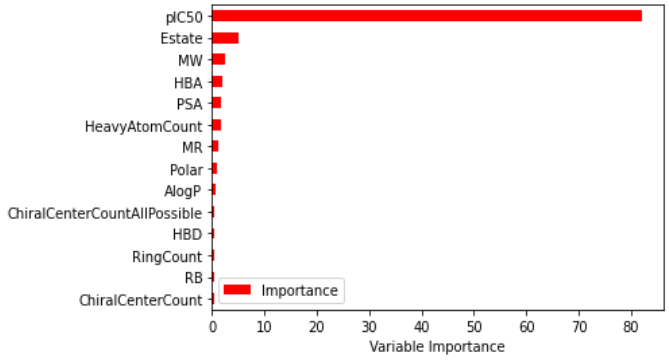


Figure 6. RF feature importance

4. CONCLUSION

This study thoroughly assessed the effectiveness of Gradient Boosting Machine (GBM), Random Forest (RF), and Support Vector Machine (SVM) algorithms in identifying potential BACE-1 inhibitors for the treatment of Alzheimer's disease. Through the analysis of precision, recall, and accuracy metrics, we observed that RF exhibited the highest classification

efficacy among the three methods, followed closely by GBM and SVM. Feature importance analysis revealed pIC50 as the most influential attribute across all methods, underscoring its significance in classifying BACE-1 inhibitors. . Moreover, RF prioritized Estate as the second most important feature, while AlogP emerged as GBM's secondary significant attribute. These findings contribute to our understanding of machine learning methods' utility in drug discovery and emphasize the importance of diverse molecular descriptors in identifying potential therapeutics for Alzheimer's disease. Future studies are planned to explore additional machine learning algorithms and feature selection strategies to increase the effectiveness of drug discovery efforts targeting Alzheimer's disease.

ACKNOWLEDGEMENT

This work is supported by the Turkish Scientific and Technical Research Council (TUBITAK) project named "Development of a New Model for the Discovery of Anti-HIV Effective Molecules Using Geometric Deep Network Approaches" and project code 123E098.

REFERENCES

[1] M. Thambisetty, L. Beason-Held, M. Kraut, R. Desikan, S. Resnick, & Y. “An. The Entorhinal Cortex-Hippocampal System Is An Early Target Of Clusterin-Related Neurodegeneration In Alzheimer’s Disease”. Alzheimer’s & Dementia, 10(4), Supplement, P160. doi: 10.1016/j.jalz.2014.04.145.

[2] E. Ruether, H. Moessler, & M. Windisch. “The MAD-B study — A randomized, double-blind, placebo-controlled trial with cerebrolysin in Alzheimer’s disease”. European Neuropsychopharmacology, 10, 355, 2000, doi: 10.1016/S0924-977X(00)80463-0.

[3] D. Al-Jumeily, S. Iram, F. Vialatte, & P. Fergus. “A novel method to analyze EEG synchrony for the early diagnosis of Alzheimer's disease in optimized frequency bands”. In 2014 IEEE 11th Consumer Communications and Networking Conference (CCNC) (pp. 1-4). Las Vegas, NV, USA. doi: 10.1109/CCNC.2014.6866646.

[4] M. M. Mishra & P. Kumar. “Crocin: A Potent Secondary Metabolite As BACE1 Inhibitor In Alzheimer’s Disease”. 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), s. 1486-1490. Coimbatore, Hindistan. doi: 10.1109/ICACCS57279.2023.10112776.

[5] S. Sharma ve Y. Hasija. “Identification and screening of BACE1 inhibitors using Drug Repurposing: A Computational Approach”. 2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT), pp. 1-5. Dehradun, Hindistan. doi: 10.1109/CISCT57197.2023.10351386.

[6] A. F. Nugroho, R. R. Septiawan, ve I. Kurniawan. “Prediction of Human β-secretase 1 (BACE-1) Inhibitors for Alzheimer Therapeutic Agent by Using Fingerprint-based Neural Network Optimized by Bat Algorithm”. 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), pp. 257-261. Jakarta, Endonezya. doi: 10.1109/ICCoSITE57641.2023.10127718.

[7] Das, B., Yan, R. A Close Look at BACE1 Inhibitors for Alzheimer’s Disease Treatment. *CNS Drugs* **33**, 251–263 (2019). https://doi.org/10.1007/s40263-019-00613-7

[8] F. H. Bazzari and A. H. Bazzari, “BACE1 Inhibitors for Alzheimer’s Disease: The Past, Present and Any Future?,” *Molecules*, vol. 27, no. 24, Art. no. 24, Jan. 2022, doi: 10.3390/molecules27248823.

[9] Z. Wang et al. “Inhibitory effects of β-asarone on lncRNA BACE1-mediated induction of autophagy in a model of Alzheimer’s disease. Behavioural Brain Research, vol 463, pp. 114896, 2024.

[10] Z. Chang et al. “Signal-on' electrochemical detection of BACE1 for early detection of Alzheimer’s disease”. Cell Reports Physical Science, pp. 101632, 2023.

[11] K. Kalaimathi et al. “Cyanobacterial metabolites as novel inhibitors of BACE1 implicated in Alzheimer’s disease through in silico approaches”. Intelligent Pharmacy, vol. 2, no. 1, pp. 144–149, 2024.

[12] B. Louis, V. K. Agrawal, ve P. V. Khadikar. “Single crystal X-ray, DFT, molecular dynamic simulations, and biological evaluation of 3-OH pyrrolidine derivative VA10 from alkaloid vasicine for BACE1 inhibition”. Journal of Molecular Structure, vol. 1300, pp. 137196, 2024.

[13] M. Nakano, T. Tsuchida, Y. Mitsuishi, ve M. Nishimura. “Nicotinic acetylcholine receptor activation induces BACE1 transcription via the phosphorylation and stabilization of nuclear SP1”. Neuroscience Research, 2023.

[14] C. B. Abraham, L. Xu, G. A. Pantelopulos, ve J. E. Straub. “Characterizing the transmembrane domains of ADAM10 and BACE1 and the impact of membrane composition”. Biophysical Journal, cilt. 122, no. 19, pp. 3999–4010,2023.

[15] B. Vincent ve S. Maitra. “BACE1-dependent metabolism of neuregulin 1: Bridging the gap in explaining the occurrence of schizophrenia-like symptoms in Alzheimer’s disease with psychosis?”, Ageing Research Reviews, vol. 89, pp. 101988. doi: 10.1016/j.arr.2023.101988, 2023.

[16] K. Dominko et. al. “Amyloid-ß plaque formation and BACE1 accumulation in the brains of a 5xFAD Alzheimer’s disease mouse model is associated with altered distribution and not proteolysis of BACE1 substrates Sez6 and Sez6L”. Mechanisms of Ageing and Development, vol. 207, pp. 111726. doi: 10.1016/j.mad.2022.111726, 2022.

[17] Z. Wu et. al. “MoleculeNet: A Benchmark for Molecular Machine Learning”. arXiv preprint, arXiv: 1703.00564, 2017.

[18] M. M. Inuwa and R. Das, “A comparative analysis of various machine learning methods for anomaly detection in cyber attacks on IoT networks,” *Internet of Things*, vol. 26, p. 101162, Jul. 2024, doi: 10.1016/j.iot.2024.101162.

[19] J. Lin et. al. “Accurate prediction of potential druggable proteins based on genetic algorithm and Bagging-SVM ensemble classifier”. Artificial Intelligence in Medicine, vol. 98, pp. 35–47. doi: 10.1016/j.artmed.2019.07.005, 2019

[20] J. I. B. Janairo. “Chapter 6 - Support vector machine in drug design. In Cheminformatics, QSAR and Machine Learning Applications for Novel Drug Development”, K. Roy (Ed.), Academic Press, pp. 161–179. doi: 10.1016/B978-0-443-18638-7.00021-9, 2023.

[21] B. Das, S. Toraman, I. Turkoglu, "A novel genome analysis method with the entropy-based numerical techniqueusing pretrained convolutional neural networks," *Turkish Journal of Electrical Engineering and Computer Sciences*: Vol. 28: No. 4, Article 9, 2020. https://doi.org/10.3906/elk-1909-119

[22] A. F. Amiri, H. Oudira, A. Chouder, ve S. Kichou. “Faults detection and diagnosis of PV systems based on machine learning approach using random forest classifier”. Energy Conversion and Management, vol. 301, pp. 118076, 2024, doi: 10.1016/j.enconman.2024.118076.

[23] A. Alloubani, B. Abuhaija, M. Almatari, G. Jaradat, ve B. Ihnaini. “Predicting vitamin D deficiency using optimized random forest classifier”. Clinical Nutrition ESPEN, vol. 60, pp. 1–10, 2024. doi: 10.1016/j.clnesp.2023.12.146.

[24] G. Cano et. al. “Automatic selection of molecular descriptors using random forest: Application to drug discovery”. Expert Systems with Applications, vol. 72, pp. 151–159, 2017. doi: 10.1016/j.eswa.2016.12.008.

[25] Das B. An implementation of a hybrid method based on machine learning to identify biomarkers in the Covid-19 diagnosis using DNA sequences. Chemometr Intell Lab Syst. 2022 Nov 15;230:104680. doi: 10.1016/j.chemolab.2022.104680.

[26] P. B. Yang, Y. J. Chan, S. K. Yazdi, and J. W. Lim, “Optimisation and economic analysis of industrial-scale anaerobic co-digestion (ACoD) of palm oil mill effluent (POME) and decanter cake (DC) using machine learning models: A comparative study of Gradient Boosting Machines (GBM), K-nearest neighbours (KNN), and random forest (RF),” *Journal of Water Process Engineering*, vol. 58, p. 104752, Feb. 2024, doi: 10.1016/j.jwpe.2023.104752.

[27] P. B. Yang, Y. J. Chan, S. K. Yazdi, and J. W. Lim, “Optimisation and economic analysis of industrial-scale anaerobic co-digestion (ACoD) of palm oil mill effluent (POME) and decanter cake (DC) using machine learning models: A comparative study of Gradient Boosting Machines (GBM), K-nearest neighbours (KNN), and random forest (RF),” *Journal of Water Process Engineering*, vol. 58, p. 104752, Feb. 2024, doi: 10.1016/j.jwpe.2023.104752.

BIOGRAPHIES

Bihter Das graduated B.S. and M.S. degrees from the Department of Computer Science at the Firat University in 2004 and 2007 respectively. Then she received Ph.D. degree at the Department of Software Engineering at the same university in 2018. She also worked between September 2017 and June 2018 as a visiting scholar at the Department of Computing Science at the University of Alberta, Edmonton, Canada. Her current research areas include data science,

big data, data analytics, bioinformatic, digital signal processing, genome data analysis.

Suat Toraman graduated B.S. and M.S. degrees from the Department of Computer Science at the Firat University in 2002 and 2006 respectively. Then she received Ph.D. degree at the Department of Software Engineering at the same university in 2016. Her current research areas include data science, big data, machine learning, deep learning, digital signal processing.

European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.14, No.1, 2024

Research Article

The Effect of Feed and Depth of Cut Parameters on Surface Roughness and Chip Morphology in X2CrNiMoN2253 Duplex Stainless Steel Materials

Fikret Sönmez*

[1*]Manisa Celal Bayar University, Hasan Ferdi Turgutlu Faculty of Technology, Mechanical Engineering Department, Manisa, Turkey. (e-mail: sonmezfikret@gmail.com).

ARTICLE INFO

Received: Mar., 20. 2024
Revised: May., 09. 2024
Accepted: June., 21. 2024

Keywords:
Surface roughness
Turning
Stainless steel
Tool wear
Machinability

Corresponding author: *Fikret Sönmez*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI:
https://doi.org/10.36222/ejt.1456172

ABSTRACT

Machining is one of the primary methods used to produce precision machine parts. Machining operations are influenced by many parameters. The most important of these parameters are feed and depth of cut. In addition, as a function of these parameters, many technical characteristics, particularly surface roughness, of the parts produced by machining change. For this reason, it is very important to monitor the machining parameters of the workpieces to be machined and to change them if necessary. Workpieces from many different groups of materials can be machined, but it is more important to monitor and control the machining parameters in stainless steels, which are classified as difficult-to-machine materials. In this study, the optimum machining conditions were investigated to obtain the lowest surface roughness values on stainless steel materials (X2CrNiMoN2253). The experiments were started with two different depths of cut (0.8 mm -1 mm) and three different feeds (0.1 mm/rev - 0.15 mm/rev and 0.2 mm/rev). It is understood that the feed is highly foremost on the surface roughness and changes the experimental results up to three times, however, the effect of cutting depth was limited. In the experiments, the best experimental results were obtained under 0.1 mm/rev feed. For a detailed analysis of the effect of the depths of cut values on the surface roughness, three additional depths of cut values (1.25 mm-1.5 mm-1.75 mm) were used under 0.1 mm/rev feed. In these additional experiments, it was observed that the effect of cutting depth was still limited, and the surface quality deteriorated with increasing cutting depth.

1. INTRODUCTION

Machining has significant superiority compared to numerous other manufacturing processes. These superior characteristics are directly affected by machining (cutting) circumstances involving cutting speed, feed, and depth of cut [1]. Machining is used to machine miscellaneous materials to meet the specific requirements of many industrial sectors, particularly the automotive and aerospace industries. However, some materials, such as stainless steel, are more challenging to machine as compared to other materials. The primary reasons for this difficulty are the weak thermal conductivity and the strong strain-hardening tendency of stainless-steel materials. Determining the appropriate cutting conditions is critical to machining these difficult-to-machine materials [2, 3].

The difficulties encountered in machining stainless steel have attracted the attention of many researchers. Three different cutting parameters (depths of cut, cutting speeds, and feeds) were used by Subbaiah et al. [4] to machine AISI 4340 steel material. In their investigation, feed was the dominant influence on surface roughness, affecting Ra by 68.5% and Rz by 31.2%. They found that tool wear increased as material hardness and cutting speed increased. According to Santhanakumar et al. [5], the surface roughness values improved as the cutting speed rose and deteriorated as the feed increased. The investigators also stated that a raise in the cutting speed accelerated the wear of the tool and that with intensified tool wear, the surface roughness deteriorated. Rashid et al. [6] found that 99.16% of the surface finish depended on the feed values and that tool wear increased at lower feed values. The tool wear investigation on ferritic stainless steel (AISI 430) was performed by Pekşen and Kalyon [7]. The analysis involved three different feeds and cutting speeds. After analyzing the results, the investigators concluded that feed and cutting speed dramatically affect tool life. The investigators also monitored the types of wear (notch, crater, and flank) on the cutting tool. Ebrahimi et al. [8] performed machining experiments on AISI 360 (17-4 PH) martensitic stainless steel. The investigators studied both surface roughness and tool wear values during the machining operation. It was noted that both feed and cutting speed have a foremost impact on tool wear. The researchers also found

that with properly selected parameters, the surface roughness values decreased by 23%.

Del Risco-Alfonso et al. [9] studied austenitic stainless steel (AISI 316L) at three cutting speeds and concluded that cutting speed had a direct influence on tool life. The main influence on tool wear was dependent on cutting speed although feed endured a negative outcome on tool wear. The machinability of austenitic stainless steel materials (AISI 304) using various cutting speeds, feeds, and depths of cut (under dry conditions) was investigated by Chen et al. [10]. The tool used in their investigation showed significant wear at the beginning, the wear increased at a later stage, and then severe tool wear was detected on the cutting tool. The investigators noticed that this wear caused the surface roughness values to deteriorate from 1.4 µm to 19.4 µm. Similarly, He et al. [11] showed that the rate of increase in tool wear changes step by step. In the first stage, rapid tool wear was observed, then it increased gradually, and finally, it increased rapidly. Surface roughness and tool wear experiments on AISI 316 stainless steel (austenitic) were performed by Derani et al. [12]. The investigators studied both nose and flank wear. Their investigation showed that although tool wear increased, the surface roughness was nearly constant. Szczotkarz et al. [13] studied tool wear on stainless steel (AISI 316) material underneath different machining conditions. The researchers point out that improper cutting conditions can lead to accelerated tool wear. In their study, Asiltürk et al. [14] analyzed the feed and depth of cut parameters on AISI 4140 material. They found that feed was the most dominant factor (87.7%) related to surface roughness. They emphasized that the depth of cut has a significant effect (28.6%) on tool vibration. Demirpolat et al [15] in their study analyzed the cutting speed, feed and depth of cut parameters on AISI 52100 material. The researchers emphasized that the surface roughness deteriorated as the feed and depth of cut values increased. The researchers also found that the chip shapes were significantly affected by the feed values.

In conducting the literature review, it was comprehended that there are countless investigations on stainless steel. However, only a limited amount of research has been found in the literature on X2CrNiMoN2253. In the current study, the reaction of different feed and depth of cut parameters on X2CrNiMoN2253 duplex stainless-steel material, which belongs to the group of difficult-to-machine materials, was investigated. For this purpose, the change in tool wear and surface roughness was investigated.

2. MATERIALS AND METHOD

A duplex stainless-steel material X2CrNiMoN2253 was employed in the machining investigations. Table 1 displays the chemical composition of this investigated material.

TABLE I
The chemical composition of X2CrNiMoN2253 materials

Element	Cr	Ni	Mo	Mn	C
wt.%	22.69	4.84	3.11	1.40	0.015

Machining operations were performed on a CNC turning center (Goodway GLS1500M) under dry machining conditions (without coolant) using a TNMG160404-HA insert (Korloy) and a PTGNR2020K16 tool holder (Seco). In the experiments, the cutting speed (V) was preferred as 180 m/min. Two different depth of cut values (0.8 mm and 1.00 mm) and three different feeds (0.10 mm/rev, 0.15 mm/rev, and 0.20 mm/rev) were used in the experiments (Table 2). The formula specified in Equation 1 was used to calculate the machining time [1].

$$T = \frac{\pi \times d \times l}{f \times 1000 \times V} \quad (1)$$

Where: T is the machining time, d is the workpiece diameter, l is the cutting length, f is the feed, V is the cutting speed.

TABLE II
The machining parameters

Depth of cut a_p (mm)	*Feed* f (mm/dev)	Machined diameter (mm)	Machining length (mm)	Machining time (min)
0.80	0.10	47.0	25	0.2051
0.80	0.15	45.4	25	0.1321
0.80	0.20	43.8	25	0.0956
1.00	0.10	41.8	25	0.1824
1.00	0.15	39.8	25	0.1158
1.00	0.20	37.8	25	0.0825

The workpieces were machined in nine passes, starting from Ø47 mm and going down to Ø28.8 mm. In each pass, the machining operation was conducted with a 25 mm longitudinal cutting length (Table 2). Although the cutting tool removes chips with a cutting length of 25 mm each time, the machining time has changed due to the change in the workpiece diameter value. The machining time, which was 0.2051 minutes in the first pass, was calculated as 0.0825 minutes in the last pass. Thus, the cutting tool performed a total of 0.8135 minutes of machining.

All machining operations were carried out under dry machining conditions by mounting the workpiece in the chuck between the tailstocks. The surface roughness results were recorded after each turning pass. The surface roughness average (Ra), surface roughness maximum average (Rz), and surface roughness maximum (Rt) values were utilized for the surface roughness measurements.

The arithmetic mean of the measurements taken five times at equal angular intervals (72°) around the workpiece was used to calculate each surface roughness parameter. A measuring terminal (Mahr M300) and a measuring device (Mahr Marsurf RD18) were employed for the measurements. In addition, all measurements were taken on the CNC turning center without the need to remove the workpieces from the chuck thanks to the Bluetooth connection. The measurements are explained in Figure 1.

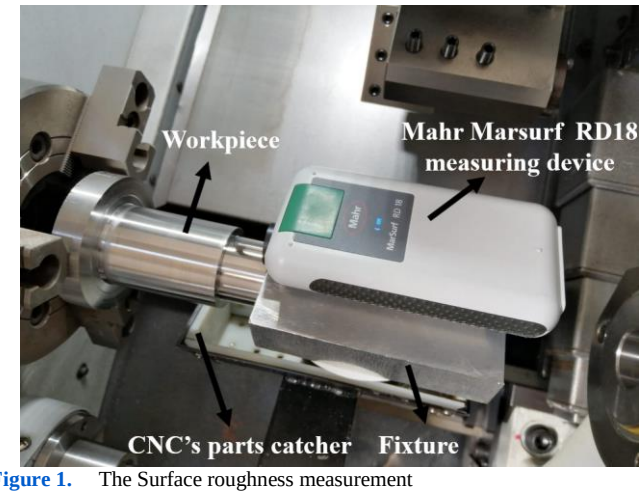


Figure 1. The Surface roughness measurement

Tool wear measurements were performed after all machining operations were completed. A toolmaker's

microscope (Mahr MM200) combined with a camera (M-shot MD30) was utilized for this purpose. The tool wear measurements performed with the toolmaker's microscope are pointed in Figure 2.

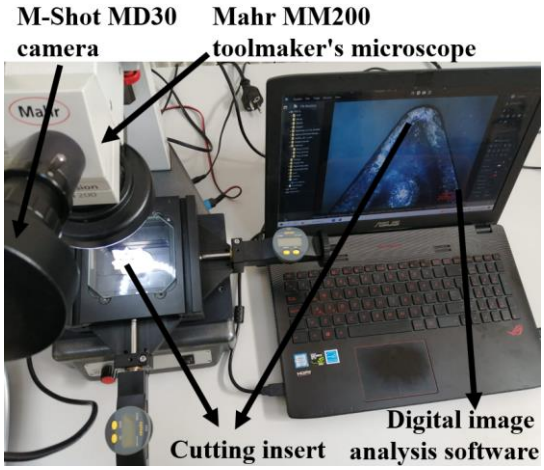


Figure 2. The tool wear measurements

3. RESULT AND DISCUSSION

3.1. Surface Roughness

After each machining operation on the CNC turning center, the surface roughness values were recorded without removing the workpiece from the CNC turning center using a surface roughness measuring device with a Bluetooth connection. The surface roughness values Ra, Rz, and Rt derived from the measurements are indicated in Table 3.

TABLE III
The machining parameters

Depth of cut a_p (mm)	*Feed* f (mm/rev)	Ra (µm)	Rz (µm)	Rt (µm)
0.8	0.1	1.528	7.142	8.628
0.8	0.15	2.471	10.35	11.64
0.8	0.2	4.147	16.47	17.09
1.0	0.1	1.414	8.552	11.50
1.0	0.15	2.801	13.60	16.87
1.0	0.2	4.693	17.59	18.20

In order to better understand the three surface roughness values (Ra, Rz, and Rt) depicted in Table 3, the impact of the feed and depth of cut on the surface roughness of the workpiece was analyzed. The figure created by using data including depth of cut and feed values from experimental data is depicted in Figure 3.

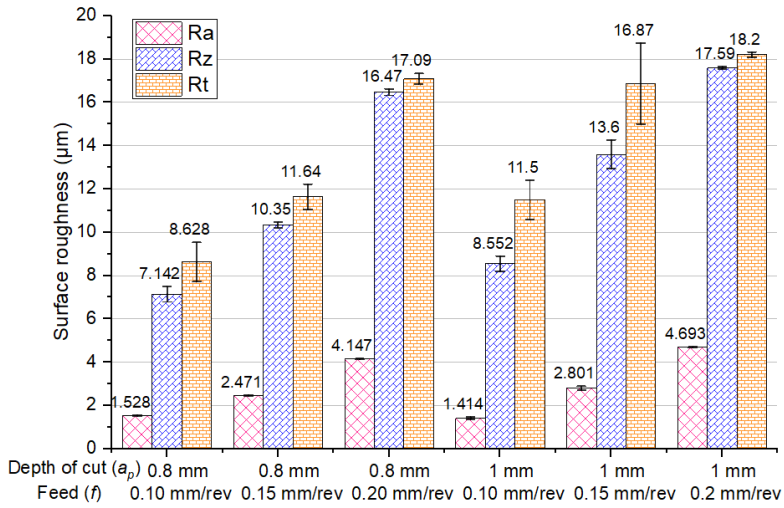


Figure 3. The surface roughness measurements depend on depth of cut and feed values.

Numerous cutting parameters directly affect machining operations. Each of these parameters has a different effect on the workpiece's features. Many studies have emphasized that the most noticeable impact on workpiece surface roughness values depends on the feed [6-7]. In this study, in parallel with the literature, it is clearly seen that the most impactful parameter in terms of average surface roughness is the feed value. It is clear that increasing the feed for both depths of cut (0.8 mm and 1.0 mm) causes serious deterioration in the surface quality (Figure 3). In the trials, the best surface roughness values were acquired at the lowest feed (0.10 mm/rev). In the experiments with a 0.8 mm depth of cut, when the feed was increased, it was determined that the surface roughness deterioration was 1.58 times and then 1.68 times, respectively. Similarly, in the experiments performed with a depth of cut (1.0 mm), depending on the increase in the feed, first 1.98 times distortion and then 1.68 times deterioration were observed. Figure 3 also shows the deviation observed in the surface roughness measurements. During the experiments, the chip control problem occurred in the experiments with ap=1 and f=0.15. Therefore, a significant deviation was observed in this experiment compared to other machining experiments.

The ANOVA analysis was implemented to investigate the effect of the depth of cut (0.8 mm and 1 mm) and the feed (0.1 mm/rev, 0.15 mm/rev, and 0.2 mm/rev), which are the experimental parameters, on the surface roughness of the workpiece. The ANOVA analysis of the surface roughness values are shown in Table 4 for Ra, Table 5 for Rz, and Table 6 for Rt.

TABLE IV
ANOVA analysis of surface roughness values (Ra)

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Depth of cut a_p (mm)	1	0.09703	0.09703	1.71	0.321
Feed f (mm/rev)	2	8.82525	4.41263	77.86	0.013
Error	2	0.11335	0.05667		
Total	5	9.03563			

TABLE V
ANOVA analysis of surface roughness values (Rz)

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Depth of cut a_p (mm)	1	5.568	5.5681	8.35	0.102
Feed f (mm/rev)	2	84.614	42.3070	63.41	0.016
Error	2	1.334	0.6672		
Total	5	91.516			

TABLE VI
ANOVA analysis of surface roughness values (Rt)

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Depth of cut a_p (mm)	1	14.143	14.143	6.62	0.124
Feed f (mm/rev)	2	57.685	28.843	13.50	0.069
Error	2	4.273	2.137		
Total	5	76.102			

The ANOVA analysis (Tables IV, V, and VI) examined the weightiness of both process parameters on the workpiece surface roughness. The efficiencies of the cutting parameters were calculated according to the Adj SS values of the parameters and are shown in Figure 4 for Ra, Figure 5 for Rz, and Figure 6 for Rt.

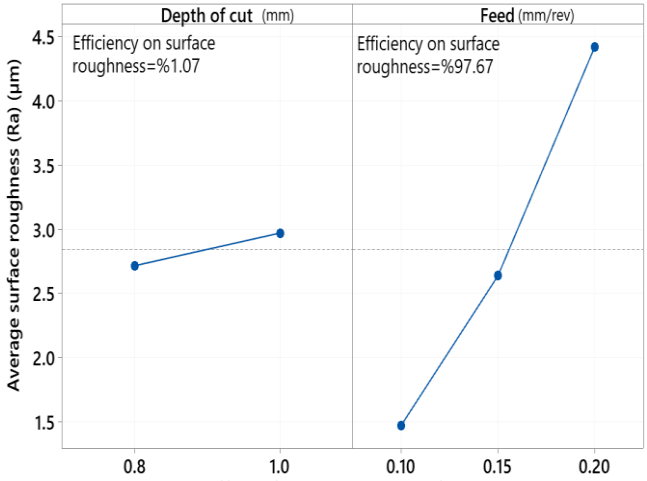


Figure 4. The effect of parameters on surface roughness (Ra)

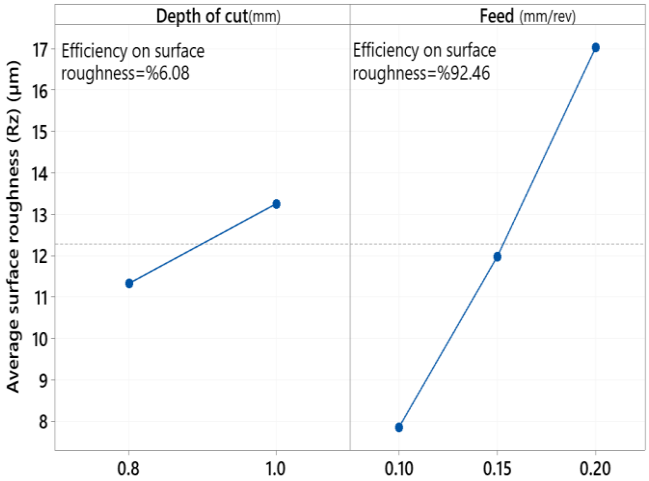


Figure 5. The effect of parameters on surface roughness (Rz)

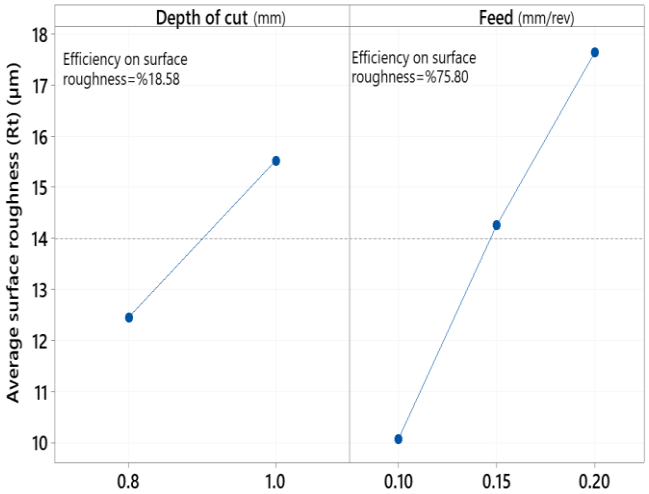


Figure 6. The effect of parameters on surface roughness (Rt)

As can be clearly seen in Table 4, the feed parameter is quite significant in this data set, as the p-value of the feed parameter is less than 0.05. In the R-sq analysis, the R-sq(adj) value was calculated as 98.75% and the R-sq(adj) value was calculated as 96.86%. Figure 4 clearly shows the obvious effect of feed (97.67%) on the machined part surface roughness. It is well known that the feed parameter has an obvious effect on the surface roughness [14-16].

It is interesting to note that despite the obvious effect of feed on surface roughness, the effect of depth of cut is very limited (1%). Although feed is the most dominant parameter in the ANOVA analysis conducted for the Rz and Rt values, it is interesting to note that the effect of the depth of cut is 6% for Rz and 18% for Rt. In this case, it can be argued that it is important to perform ANOVA analysis within different surface roughness parameters. The effect of the other process parameter, the depth of cut parameter, on the surface roughness should also be analyzed. To further investigate the effect of depth of cut on surface roughness, three additional depths of cut (1.25 mm-1.5 mm-1.75 mm) were used at a feed of 0.1 mm/rev. In addition to the existing two experiments conducted at a feed of 0.1 mm/rev, the results of three additional experiments at different depths of cut are shown in Table 5. With these experiments, the cutting tool, which was previously used for 0.8135 minutes, was used for a total of 1.2341 minutes.

TABLE V
The existing and new experiments

Depth of cut a_p (mm)	*Feed* f (mm/rev)	Ra (µm)	Rz (µm)	Rt (µm)	Experiments info
0.80	0.10	1.528	7.142	8.628	Existing
1.00	0.10	1.414	8.552	11.50	Existing
1.25	0.10	1.435	8.495	11.98	New
1.50	0.10	1.889	10.60	15.10	New
1.75	0.10	1.757	10.69	12.84	New

The values from Table 5, obtained for a total of five different cutting depths, are shown in Figure 5.

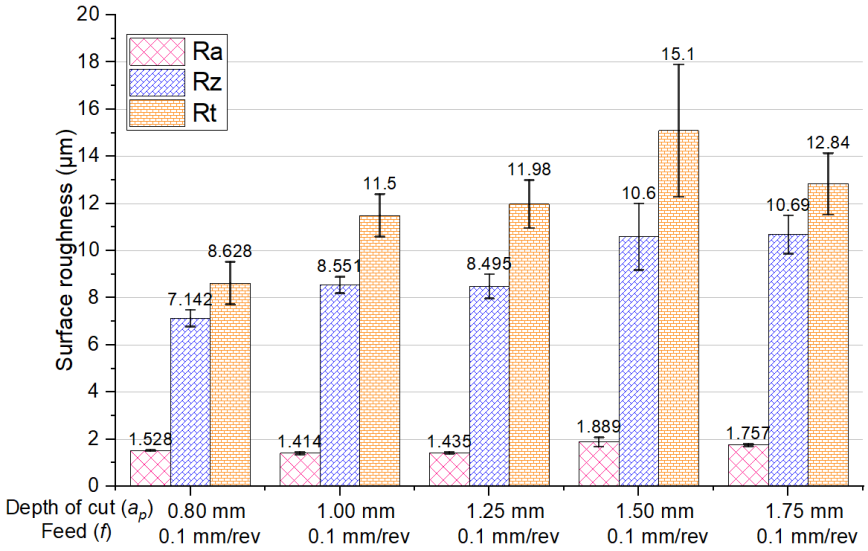


Figure 5. The surface roughness measurements depends on depth of cut values

As depth of cut becomes deeper, the surface roughness of the workpiece deteriorates. However, in the experiments with a 1.5 mm depth of cut, the surface roughness deterioration was greater than expected. In this experiment, large deviations are observed in the surface roughness obtained in this experiment.

Although surface roughness is affected by many factors, the most dominant effect is generally due to feed. The following equation (Equation 2) is commonly used to calculate theoretical surface roughness values [1].

$$R_{th} = \frac{f^2}{8 \times r_\varepsilon} \tag{2}$$

Where: R_{th} is the theoretical surface roughness values, r_ε is the nose radius of the cutting tool radius.

As can be seen from Equation 2, the feed and the nose radius value of the cutting tool are critical in calculating the theoretical surface roughness. The nose radius value of the cutting tool was held constant in the experiments, therefore the theoretical surface roughness changes only as a function of feed. Since increasing the depth of cut values increases the cutting forces [15], vibrations [14] the deflection of the cutting tool increases, and the surface roughness deteriorates. In addition, the cutting mechanics change due to the increased chip cross-section, and the surface roughness deteriorates [2].

3.2. Tool wear

The duplex stainless steel workpiece (X2CrNiMoN2253) was machined in a total of nine passes., with six initial experiments and three additional experiments. The cutting tool was utilized for a total of 1.2341 minutes during these studies, compared to its previous usage of 0.8135 minutes. After these operations, the cutting tool was inspected for tool wear (Figure 6). The images of the top surface of the cutting tool (Figure 6b), the flank surface (Figure 6c), the nose surface (Figure 6d), and the top surface at an angle of 10° (Figure 6e) were examined. Tool wear inspections were performed using the measurement procedures specified in ISO 3685:1993 (Tool-life testing with single-point turning tools) were used [16].

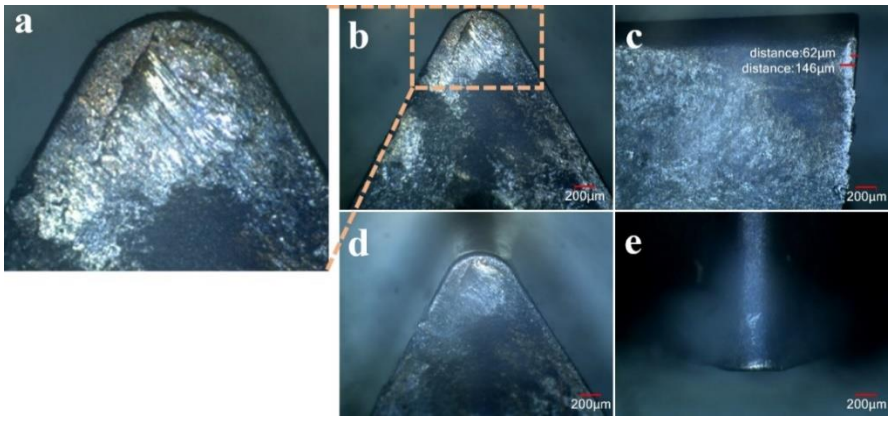


Figure 6. The tool wear inspection

It is a well-established fact that cutting tool wear increases over time. Since the total cutting time of the cutting tool is very short, the amount of wear was also less. A maximum of 146 µm of flank wear was recorded (Figure 6c). In addition, no significant notch wear was observed. The top surface of the tool (Figure 6b) was examined in detail (Figure 6a). It is evident that the coating on the top surface of the cutting tool was severely worn, and crater wear had begun. It is also clear that the nose wear is quite low. Although the cutting speed in the experiments was slightly higher than the tool manufacturer's recommendation, tool wear was limited due to the short cutting time.

3.2. Chip Morphology

It is well known that chips with long morphologies should be avoided in machining processes. Since chip evacuation is crucial in industrial applications, it is important to choose machining conditions that aim to produce short chips. For this reason, it is necessary to select the optimum experimental conditions both scientifically and industrially. The chips obtained as a result of the machining process are directly related to the cutting parameters. Different chip morphologies were obtained depending on the cutting depth and feed used in the experiments (Figure 7).

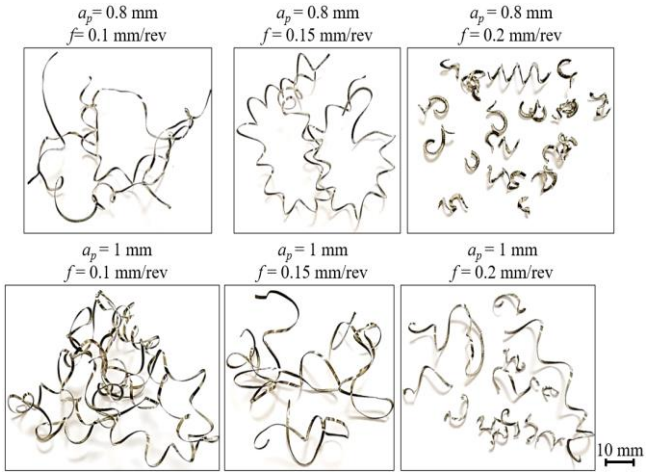


Figure 7. The chip morphologies depends on depth of cut and feed values.

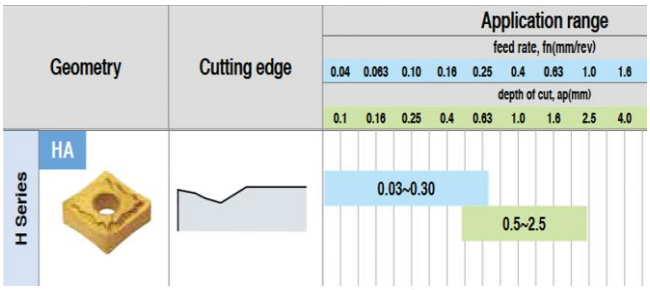


Figure 8. HA Chip breaker form [15]

Chip morphology stands a function of multiple cutting parameters. Depth of cut, feed, and cutting speed are the most important. However, cutting tools often have special chip breaker forms designed for specific machining conditions (Figure 8). The values recommended by the tool manufacturer are used to fulfill the functions of these chip breaker forms. The depth of cut and feed values used in the experiments also have a significant effect on chip morphology (Figure 7). Moreover, Figure 7 clearly shows that increasing the feed shortens the chip lengths. It is also clear that the raise in the value of the depth of cut shortens the chip lengths, although not as much as the feed.

It is known that chip breaking becomes difficult when applications require elevated cutting speeds, low feed, and depths of cut. In these experiments, this situation occurred generally, and the chip breaker shape was not effective in general, except for 0.2 mm/rev feed. The chip shapes were examined in Figure 9 by keeping the feed constant and depending only on the depth of cut.

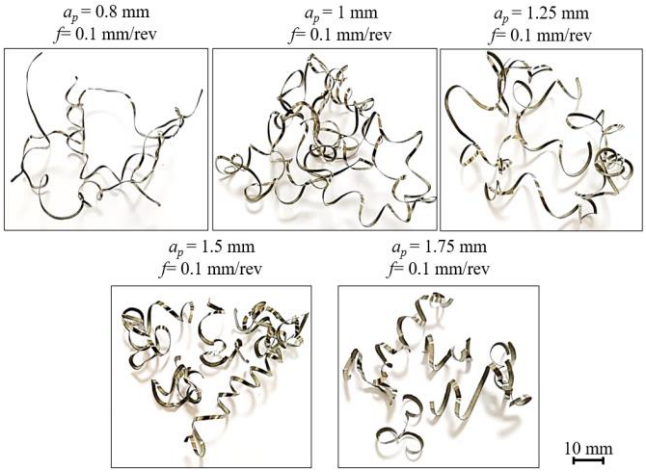


Figure 9. The chip morphologies depends on depth of cut

Figure 9 confirms that the chip shapes are affected by the depth of cut. However, this effect is not obvious. Relatively short chips were obtained at only 1.5 mm and 1.75 mm.

The cross-sectional area of the chips varies as a function of the approach angle, feed, and depth of cut values. Since the approach angle and feed were constant under the experimental conditions shown in Figure 9, only the cross-sectional area of the chips changed depending on the feed value. In Figure 9, it can be clearly seen that the chip cross-section has widened, but the long chip morphologies do not change in the experiments. It is well known that the feed effects the chip morphologies [15] and must be increased to obtain short chips in cutting tools. In this case, it can be proposed that there is no change in chip length due to the low feed value.

In this case, it is clear that a high feed (0.2 mm/rev) should be used to obtain the most appropriate chip shapes according to the current experimental parameters. The fact that the chip breaker shape is not active is a situation that should be avoided, as it makes chip control difficult and affects the surface roughness.

4. CONCLUSION

In the current study, the most favorable machining conditions were investigated to obtain the lowest surface roughness values on stainless steel materials (X2CrNiMoN2253) using different cutting depth of cuts values (0.8 mm -1 mm) and feed values (0.1 mm/rev - 0.15 mm/rev and 0.20 mm/rev). The following conclusions were obtained as a result of this study.

The feed and the depth of cut values have a distinct change on the surface roughness (Ra). However, it is comprehended that the feed has a rather dominant effect (97.67%) on the surface roughness values. Compared to Ra values, it is understood that the effect of feed on other surface roughness parameters (Rz, Rt) decreases.

It has been determined that the depth of cut has a certain effect on the surface roughness. In general, the surface roughness deteriorates as the depth of cut becomes deeper. Unlike Ra values the effect of the depth of cut parameter has become evident in Rz and Rt values. The effect of the cutting depth was calculated as 1.07% for Ra, 6.08% for Rz and 18.58% for Rt.

The tool wear value is also very limited since the total cutting time of the cutting tool is very short. In the examinations, the utmost value of the flank wear was determined as 146 µm.

In terms of chip morphology, it has been observed that the chip morphology changes significantly depending on the feed. It was found that a high feed and a high depth of cut values contribute to the formation of better chip forms. Chip control could not be partially achieved at low feeds.

In present study, the effectiveness of feed and depth of cut on the workpiece surface roughness and cutting tool wear was investigated. In future studies, the effects of different cutting parameters on cutting forces and cutting tool vibrations can be investigated. Also, different coolant and MQL cutting conditions can be investigated.

REFERENCES

[1] Stephenson DA., Agapiou JS. *Metal cutting theory and practice*. 2nd ed. Florida, USA, CRC press, 2018.

[2] Tzotzis, A., García-Hernández, C., Huertas-Talón, J.L., Kyratsis, P., 3D FE Modelling of Machining Forces during AISI 4140 Hard Turning. *Strojniški vestnik – Journal of Mechanical Engineering, 66 (7-8)*, pp. 467-478, (2020), doi: 10.5545/sv-jme.2020.6784.

[3] Valić GS, Kostadin T, Cukor G, Fabić M. Sustainable Machining: MQL Technique Combined with the Vortex Tube Cooling When Turning Martensitic Stainless Steel X20Cr13. *Machines, 11 (3)*, pp 1-22, (2023), doi: 10.3390/machines11030336.

[4] Subbaiah, K.V., Raju, C., Suresh, C., Parametric analysis and optimization of hard turning at different levels of hardness using wiper ceramic insert. *Measurement, 158*, pp. 1-13 (2020), doi: 10.1016/j.measurement.2020.107712.

[5] Santhanakumar, M., Adalarasan, R., Siddharth, S., Velayudham, A., An investigation on surface finish and flank wear in hard machining of solution treated and aged 18 % Ni maraging steel. *Journal of the Brazilian Society of Mechanical Sciences and Engineering, 39 (6)*, pp. 2071-2084, (2016), doi: 10.1007/s40430-016-0572-0.

[6] Rashid, W.B., Goel, S., Davim, J.P., Joshi, S.N., Parametric design optimization of hard turning of AISI 4340 steel (69 HRC). *The International Journal of Advanced Manufacturing Technology, 82 (1-4)*, pp. 451-462, (2015), doi: 10.1007/s00170-015-7337-2.

[7] Pekşen, H., Kalyon, A., Optimization and measurement of flank wear and surface roughness via Taguchi based grey relational analysis. *Materials and Manufacturing Processes, 36 (16)*, pp. 1865-1874, (2021), doi: 10.1080/10426914.2021.1926497.

[8] Ebrahimi, S.M., Hadad, M., Araee, A., Sustainable machining of hardened AISI630 stainless steel using thermally enhanced turning technique. *Machining Science and Technology, 25 (4)*, pp. 608-636, (2021), doi: 10.1080/10910344.2021.1903922.

[9] Risco-Alfonso, R., Pérez-Rodríguez, R., Robledo, P.D.C., Santana, M.R., Quiza, R., Optimization of the Cutting Regime in the Turning of the AISI 316L Steel for Biomedical Purposes Based on the Initial Progression of Tool Wear. *Metals, 11 (11)*, pp. 1968, (2021), doi: 10.3390/met11111698.

[10] Chen, J., Wang, Y., Zhang, Y., Yang, S., Zhang, X., Investigation on Tool Wear Mechanism during dry cutting 304 Stainless Steel, *Manufacturing Technology. 20 (1)*, pp. 36-44, (2020), doi: 10.21062/mft.2020.011.

[11] He, Q., DePaiva, J.M., Kohlscheen, J., Veldhuis, S.C., Analysis of the performance of PVD AlTiN coating with five different Al/Ti ratios during the high-speed turning of stainless steel 304 under dry and wet cooling conditions. *Wear, 492-493*, pp. 204-213, (2022), doi: 10.1016/j.wear.2021.204213.

[12] Derani, M.N., Ratnam, M.M., Nasir, R.M., Improved measure of workpiece surface deterioration during turning using non-contact vision method". *Precision Engineering, 68*, pp. 273-284, (2021), doi: 10.1016/j.precisioneng.2020.12.016.

[13] Szczotkarz, N., Mrugalski, R., Maruda, R.W., Królczyk, G.M., Legutko, S., Leksycki, K., Dębowski, D., Pruncu, C.I., Cutting tool wear in turning 316L stainless steel in the conditions of minimized lubrication. *Tribology International, 156*, pp. 1-11, (2021), doi: 10.1016/j.triboint.2020.106813.

[14] Asiltürk, İ., Kuntoğlu, M., Binali, R., Akkuş, H., & Salur, E., A comprehensive analysis of surface roughness, vibration, and acoustic emissions based on machine learning during hard turning of AISI 4140 steel. *Metals 13.2* (2023): 437.

[15] Demirpolat, H., Binali, R., Patange, A. D., Pardeshi, S.S., & Gnanasekaran, S., Comparison of tool wear, surface roughness, cutting forces, tool tip temperature, and chip shape during sustainable turning of bearing steel. *Materials 16.12* (2023): 4408.

[16] Başak, H., & Baday, Ş., Küreselleştirilmiş orta karbonlu bir çeliğin işlenmesinde, kesme parametrelerinin kesme kuvvetleri ve yüzey pürüzlülüğüne etkilerinin regresyon analizi ile modellenmesi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 22(4)*, (2016) pp.253-258.

[17] Korloy Cutting Tool Catalog, 2020-2021 ed., 2020, p. 53.

[18] ISO. Tool-life testing with single-point turning tools. 1993.

BIOGRAPHIES

Fikret SÖNMEZ received his MSc degree in 2009 and his PhD degree in 2016, from the Department of Mechanical Education at Gazi University, Ankara, Turkey. He is an Associate Professor at the Department of Mechanical Engineering, Hasan Ferdi Turgutlu Faculty of Engineering, Manisa Celal Bayar University, Manisa, Turkey. Assoc. Prof. SÖNMEZ's fields of interest are are machining, machinability, surface integrity, mechanics of machining, optimization and and MQL.

European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.14, No.1, 2024

Research Article

An Innovative Design of a 3D Game-Supported Physical Therapy Device for Wrist Disorders

Pınar Aslan Yıldırım[1*], **Adnan AKKURT**[2]

[1*]Gazi University, Industrial Design Engineering Department, 06500, Yenimahalle, Ankara, Turkey. (e-mail: pinaryildirim@gazi.edu.tr).

[2] Gazi University, Industrial Design Engineering Department, 06500, Yenimahalle, Ankara, Turkey. (e-mail: aakkurt@gazi.edu.tr)

ARTICLE INFO

Received: Dec., 18. 2023
Revised: Feb., 20, 2024
Accepted: April, 02, 2024

Keywords:
Industrial Technology Engineering
Unity
Arduino Uno
Wrist Dysfunction

Corresponding author: *Name Surname*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1406396

ABSTRACT

Functional disorders can severely affect a person's quality of life, making even simple daily activities a challenge. Especially finger, hand, and wrist disorders can cause problems even in the daily necessities, which are very simple. Patients with reduced mobility should go to physical therapy centers and receive treatment to regain their lost mobility. Physical therapy and rehabilitation treatments are processes that give results in long-term and require patience and persistence. In recent years with the effect of growing computer science technologies, there has been a growing interest in the use of game-supported treatments for hand and wrist functional disorders like in other fields. These new-generation treatments use interactive games to deliver sustainable treatments that make the process fun and engaging. These games are designed and developed to make patients do some specific movements to gain their functions again by making their muscles stronger. With the infrastructure described in this study, it aims to make a prototype that includes a microprocessor, wireless communication capability, and gyro sensor so that patients can perform physical therapy movements while playing games. Additionally, in this study, the presented prototype is compared in terms of cost with the high-budget products preferred in the market. With this study, we aimed to raise awareness that producing such innovative technologies may be less costly than transferring them from outside.

1. INTRODUCTION

With Industry 4.0, the technology is progressing quickly. With the automated systems, physical effort needed works are getting easier for humans. However, it causes them not to move during the day, and benefit from the sun, and fresh air less. Due to these disadvantages, humans lose their physical resistance and can faced with muscle diseases easily. It is also possible to encounter physical damage that cannot be compensated in unexpected accidents and some issues came from innate or traffic events. According to a guide about the global strategy for improving musculoskeletal health; approximately 1.71 billion people globally live with musculoskeletal conditions, affecting people of all ages. This report also shows us that high-income countries have the highest number of affected individuals (441 million) [1].

Additionally, physical disorders can be seen because of birth defects. "Birth defect" is a medically accurate term to describe structural changes to a person's body that happen during fetal development. About 3% of children born in the United States (US) have a major birth defect; birth defects account for about 20% of all infant deaths [2]. Every four and a half minutes, a baby in the US is born with a birth defect. This equals 1 out of every 33 babies born or 120,000 babies each year [3]. These statistics show us how common birth defects are. The humans are born with birth defects will probably need some physical support equipment or physical therapies.

Some jobs like dentistry, software engineering, and hairdressing, besides some issues that come from innate, accident, or chronic diseases, or arising from intensive computer use in daily life may cause physical disorders.

For patients with limited mobility due to many different reasons, it may be necessary to be treated in physical therapy centers to restore the lost movement capabilities. Physical therapy and rehabilitation processes are treatments that give results over a long period and require patience and stability. Treatments can be painful and challenging at times, depending on the place and severity of the condition.

This study presents a design that will help patients who need physical therapy for wrist disorders to perform painful and sometimes boring rehabilitation treatments more easily and funny. In traditional treatments for wrist disorders, the physical therapy specialist teaches the patients flexion, extension, radial-ulnar deviation and wants them to make these exercises in sets. The specialists sometimes use some toys like play dough, stress balls, and hand bows to simulate the daily movement requirements. During these treatments, the patients need only a table and a towel, which is shown in Figure 1.

Figure 1. Traditional treatment choices for wrist disorders

These types of traditional treatments sometimes come boring to the patients and they don't prefer to complete all treatment periods. At this point, we prepared a system to encourage patients with wrist function disorder to perform physical therapy movements. They wear a lightweight box on their hand surface and fasten it with velcro straps. After that, they can make their treatment while playing a game. The electronic infrastructure in the box can take the hand movement angles via Bluetooth and gives the movement to the game objects. The patient and the specialist can view the improvement of the patient from the reports. This feature will be planned for the last version of this study.

In this paper, we studied to take the data from a wire connection and gave the movement to the game objects with the data of the users' wrist movement angles. We used an Arduino Uno, which includes an Atmega 328P microprocessor it, an MPU6050 IMU 3-axis gyro sensor, and cables for communication. The angle data was collected from the users while they were doing the physical therapy exercises with 3D games. This data was read by a gyro sensor with a developed script file. By transferring the movement data to the game platform, Unity, we gave directions to the game object. We designed a lightweight box to locate the circuit on user's hand. We developed a 3D game, which allows users collecting gold while driving a car, and a car selection with collected gold. We didn't use any commercial equipment or games for developing this prototype. We are still studying this prototype to provide it with a wireless connection, and improving its reports.

It will be possible to get more efficient results by improving this prototype in terms of game technologies, motion sensitivity, and reporting features. This design offers patients to be able to perform the necessary movements at home. It is also suitable for clinics because it can be cleaned easily and it can fit many patients' hands. This product will be suitable for both adult and pediatric patients.

This kind of technologies are getting popular in health treatments. Unfortunately, most of the physical therapy centers cannot afford this kind of trend technologies' budgets. In this study, we propose a prototype that has a big price difference with professional products. It is evaluated that by supporting the mass production of similar prototypes, cost-effective and functional products that compete with the global market can be developed. It aims to raise awareness on the possibility of a domestic product alternative to centers and patients currently using or planning to use computer game-based treatment tools which are imported with high budgets.

2. OVERVIEW

Commercial products that are especially preferred for hand and wrist physical therapy and rehabilitation processes were studied. Then, we searched for a literature review and academic studies related to our study. You can find the details below.

2.1. Commercial Products and Academic Literature Review Related to Smart Tools Suggested For Hand and Wrist Disorders' Physical Therapy

There are many commercial products that provide treatments with gaming options.

"Rapael Smart Glove" is a wearable technology product. It has bending sensor, 9-axis motion and position sensor with 3 acceleration channels, 3 magnetic field channels, 3 angular velocity channels. With Wi-Fi connection, it sends the wrist movement angles to the computer. It uses 32-bit microcontrollers and low power consumption which makes the device lightweight, 132 g. After each exercise, the patients can see the performance report and share the results with the doctor [4].

"Hocoma/Armeo@Spring" is for neurological disorders such as multiple sclerosis (MS), spinal cord injuries, cerebral palsy, or subsequent brain damage. It has mechanical arm support, and relevant performance feedback [5].

"Armeo Senso" has a tool for the hand and 3 motion sensors on the wrist, elbow, and chest. It is appropriate for use at home with remote control, following the initial training provided by a therapist [5].

There are some academic literature that suggests treatments with gaming options in Türkiye.

Fidan U. et al. studied on "Design and Implementation of Measurement and Exercise System with Kinect Sensor for Neurological Rehabilitation" [6]. They used Kinect device for taking patients' movement data and sent these data to Unity 3D gaming motor. By this study, they were offering a sustainable solution to the patients who were suffering from stroke, parkinson's and cerebral palsy.

Erdogan H. et al. studied on "Game Design for Physical Therapy and Rehabilitation Using Kinect" [7]. They used Kinect device for taking patients' movement data and sent these data to Unity. They offered patients or users the opportunity to exercise without getting bored. They prepared one main game called "Maze" and the user can reach other 3 game choices through the main game.

Ozcan H. and his friends studied on "A Leap Motion Based Mobile Game Design for Developing Hand and Wrist Movement in Children" [8]. They used Leap Motion device for taking patients' movement data and sent these data to the computer and a game infrastructure. Unfortunately, they did not mention about the technology or tool that they used for developing games. They offered the children a motivated way to do the physical therapy they needed.

In all the researches that we reviewed during this study,

1. They offer a sustainable and funny way to do the mandatory physical therapies especially for neurological issues, stroke, parkinson's and cerebral palsy.
2. They claim that they provide more effective and successful treatment conclusions by the support of gaming.
3. They prefer to use commercial products for taking the patients' movement data. Most preferred commercial products are Nintendo's Wii remote [9], Sony's EyeToy

[10], Microsoft's Kinect [6, 7] and Ultrahaptics's Leap Motion [8].

4. They generally provide game options which were prepared with Unity game application.

2.2. Comparison

Below you can find a table where we compared those products and studies:

TABLE 1
COMPARISON OF TOOLS AND STUDIES

	Treatment Area	Mechanic Support	3D Game Support	Budget/ Subscription
Design and Implementation of Measurement and Exercise System with Kinect Sensor for Neurological Rehabilitation	Hand, hand wrist, arm	X	✓	Half-commercial, using Kinect device.
Game Design for Physical Therapy and Rehabilitation Using Kinect	Hand, hand wrist, arm	X	✓	Half-commercial, using Kinect device.
A Leap Motion Based Mobile Game Design for Developing Hand and Wrist Movement in Children	Hand, hand wrist, arm	X	✓	Half-commercial, using Leap Motion device.
Rapael Smart Glove	Hand, hand wrist	X	✓	£17,230 / £175 per month [11]
Armeo@Spring	Hand wrist, arm, shoulder	✓	✓	N/A
Armeo@Senso	Hand, hand wrist, arm	X	✓	N/A

To summarize, the results show us that there are some choices about using new generation technologies in wrist/hand physical therapy treatment. But they are limited or have high prices. They are generally commercial or half-commercial with using a commercial device's capabilities. This conclusion directed us to develop a low-cost and effective solution.

3. OUR DESIGN

3.1. Methodology

During the research and study, we noticed that physical therapy and rehabilitation treatments produce productive results only if the patients make the exercises regularly. This project started with the idea of making physical therapy and rehabilitation fun while entertaining the patients and improving them. Due to the needs and applications that vary according to the region where physical therapy will be applied, this study has developed against wrist disorders caused by any reason.

In the literature reviews, Sutanto and his friends shows us that a gaming physical therapy solution supported with a commercial device like Kinect, Leap Motion or Wii,etc. isn't a cost-effective choice [12]. We tried various development alternatives for a cost efficient solution with this idea. We have an awareness after these researches, we should develop not only the games but also the communication infrastructure to be able to provide the patients or users a sustainable and cost-effective solution.

The project needs a multidisciplinary study. For this reason, it is possible to itemize the components of the project, each of which touches on a separate discipline;

- Material research suitable for modeling and manufacturing,
- Determine the reference values of healthy people and implement them in the system,
- Electronic and communication infrastructure,
- 3D game design and implementation.

Figure 2 shows us that the design of electronic and communication infrastructure, 3D game development, and implementation. These titles will be explained in detailed in next subtitles since other components like modeling and manufacturing and health subtitles constitute a study subject belonging to a separate discipline.

Figure 2. The process steps examined in this study.

3.1.1. Electronic and Communication Infrastructure

This section includes the infrastructure components and functions that will collect the movement angles of the hand and wrist on the material to be positioned on the hand and transmit these data to the computer.

First of all, since the study aims to include hand and wrist movements, the circuit that plans to be put on the hand's surface should consist;

- Sensors to consistently detect motion angles,
- A microprocessor to collect this data,
- A serial port adapter for data communication,
- A Bluetooth connection adapter is suitable for communication from the serial port to provide wireless usage,
- Lightweight, not preventing movement, and having instantaneous data transmission.

Many researches and trials have been carried out within the scope of these requirements. First, we studied an Arduino Uno

microprocessor, a three-angle MPU6050 gyro sensor, a cable with USB A and USB B inputs, yellow, gray, and red male-male jumper cables, and a breadboard on which all these would locate. The wired connection has been used to make it faster.

The circuit that prepared for this study is shown in Figure 3. We continue to improve this study for making wireless connection for the next versions.

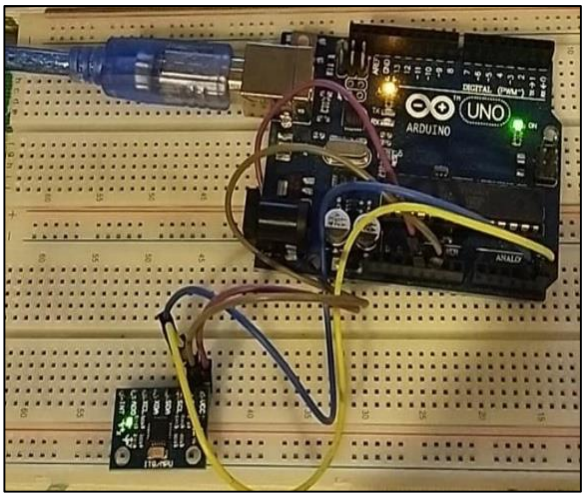


Figure 3. Circuit Elements

A *.ino file was created using the Arduino editor to read the angle values given to the breadboard with the use of the three-angle gyro sensor. In the Arduino code file;

- The MPU6050 IMU has both 3-Axis accelerometer and 3-Axis gyroscope integrated on a single chip,
- The gyroscope measures rotational velocity or rate of change of the angular position over time, along the X, Y and Z axis,
- The outputs of the gyroscope are in degrees per second, so in order to get the angular position we just need to integrate the angular velocity,
- To implement this, we used MPU6050_tockn.h and Wire.h libraries of the gyro sensor,
- The code block consists of two methods, setup, and loop,
- In the setup method; the serial communication channel opened, the port read rate was assigned as 9600 baud rate, and the assigned port was COM4. The MPU 6050 is automatically calibrated via the calcGyroOffsets function,
- In the loop method; It reads the three-angle data of the gyro sensor as X, Y, and Z angles from the serial port every 100 ms and writes it to the screen by putting comma between them,
- Because this study aims improvement of the patients' hand wrist movements, we don't need to use accelerometer data. For the further development, this capability can be integrated easily and provide a treatment alternative for the patients' arm disorders.

After the program is loaded on the Arduino card and opened at a 9600 baud rate, the angular values of the gyro sensor are read and written to the screen every 100 ms. You can find the detailed connection schema (Figure 4) and circuit design (Figure 5) below:

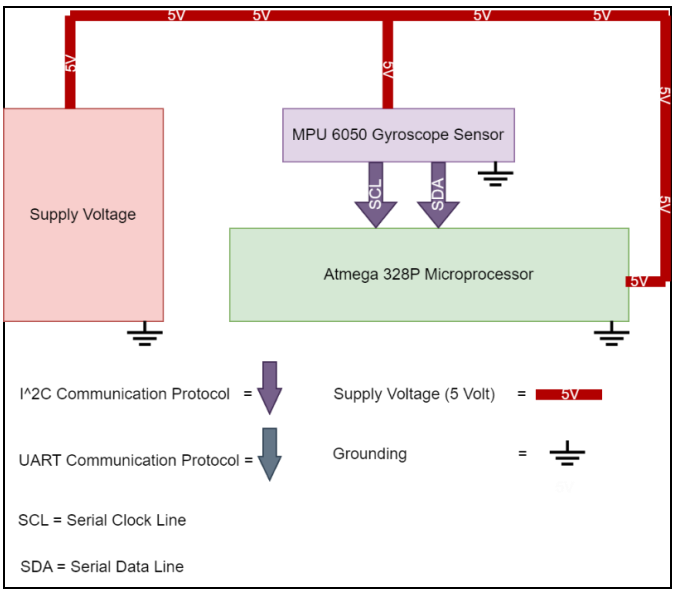


Figure 4. The circuit connection schema.

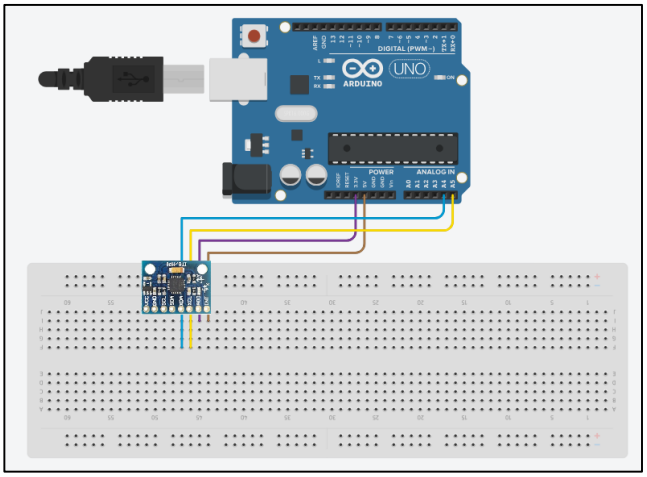


Figure 5. The circuit design.

These components are easily reachable and cost-effective items as shown in Table 2. Even we are still studying to improve this device to make it more comfort and user-friendly, we will keep the main components of this circuit. So there won't be more changes about the required budget for the next version.

TABLE 2
THE PREDICTED COSTS REQUIRED IN THIS STUDY

	The Item Used In The Circuit Design	Price (€)
1.	Arduino Uno Microprocessor	~32 € [13]
2.	MPU 6050 3 axis gyro sensor	~3 € [14]
3.	Breadboard (83.5mm x 54.5mm x 8.5mm)	~2 € [15]
4.	Cables (M-M Jumper, 200mm)	~1€ [16]
Total		~38€

3.1.2. 3D Game Development and Implementation

There are many alternative platforms for game development. Platforms such as Unity, Unreal Engine, Marmalade, and Buildbox are among the most preferred. According to Staiano and Flynn, using the videogames for balance, rehabilitation, and illness management could considered useful and funny [17]. According to Elor, et al., insights from the interviews, that they made around 130 interviews with therapists/physical therapists from mostly

California and New York, suggest that virtual reality offers great potential to improve physical therapists' skills and expand their scope of practice in rehabilitation, and health care through telemedicine[18]. Patients are more willing to complete the treatment. In our study, we decided to develop a game and use Unity since it is the most preferred platform, has many assets and source documents provided by the community, provides non-commercial opportunities, and has technical support forums used by a wide variety of audiences.

We used Windows 10 Enterprise operating system. As with many development platforms, the configurations you need to make during development may change depending on the version you use in Unity. The last consistent version at that time was 2021.2.3f1, and it has been preferred for development.

It is also possible to customize the development stages required for the game within itself. To develop a successful game, you need to create a consistent scenario of the game, design and draw the characters to use, gain the movement abilities of the characters both as drawing and code, create the game scene, design, and implement the game objects to be in this scene, give sound, light, and visual effects, etc. Many stages needed to be completed. In this study, since our aim is not to develop a game in a professional sense, free templates that are already offered at some resources were used. It is possible to access quite a lot of characters and environment design libraries through the asset store, which can also be accessed from the Unity editor, especially for character designs, and you can easily download these designs from the Package Manager and import them into the game.

First, we tried to develop an augmented-reality-supported animation-based game alternative shown in Figure 6. It was similar to a dart game. However, it didn't work synchronously with the circuit.

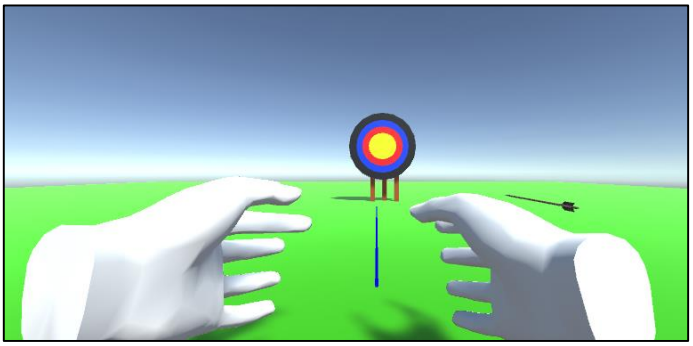


Figure 6. The developed augmented-reality-supported game trial

So we decided to develop a 3D game with serial port communication. A gold collection game with an interactive car with wrist movements has been prepared to treat wrist disorders.

The game starts with an object selection, a standard car option. When the user lowers his wrist while it is parallel to the ground, the car game object will move in the forward direction, by moving his wrist to the right and left, the car game object will change direction in the same way. With the gold collected in this way, the user will be able to choose other car options among four options. In this study, the basic functions and configurations that support physical therapy movements for wrist movements related to the developed 3D computer game will be emphasized. Visual details will not be entered.

During the preparation of the game, Arcade Free Racing Car, Kajaman Roads, Joystick Pack, and Gold Coins assets were downloaded from the asset store and imported into the project.

The most critical step of the game is to read the data coming from the serial port in a meaningful way in the game environment. For this task, Unity and Visual Studio work together. There are many programmatic activities in the game like controls, conditions, transitions between scenes, etc. For many steps, Unity needs scripts written in Visual Studio. For consistent data read from the serial port, System and System.IO.Ports libraries must be used in Visual Studio. To do this, the Player Settings screen opens via File-> Build Settings->Player Settings. There is an "Other Settings" title. The configuration must be set like Figure 7. Below the "Configuration" part, "Scripting Backend" should be selected "IL2CPP", and "Api Compatibility Level" should be selected ".NET Framework". The other settings are the same. Only in this way, the System.IO.Ports library can be activated on the Visual Studio platform.

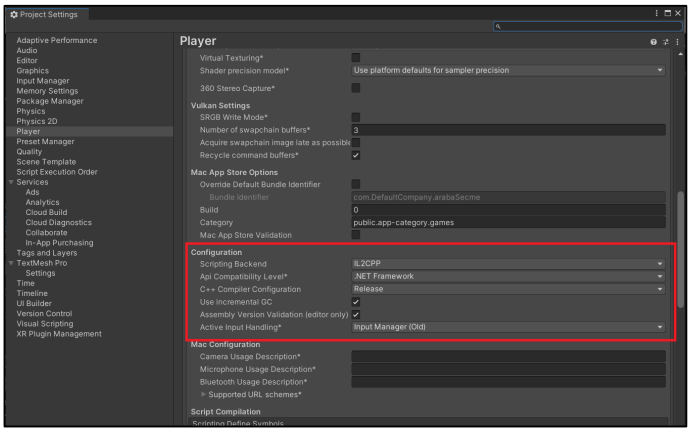


Figure 7. Player Settings Configuration Editor

Prefab allows you to save the current properties of a completed game object, use it later, change it, and automatically make changes at once at all the points you use prefab. Like working with templates. We can list prefabs in the Project tab in the Unity Editor. We can move the prefabs, listed in the Project tab, to the scene or hierarchy part by drag and drop.

There are different road, landscape, and car options in the selected asset packages. There is also a prefab in the car asset package where the colliders of the wheels are defined. Collider is a feature that allows us to use the physical features of the game object with the Rigidbody feature, together with the Unity physics engine capability. With this feature, game objects can pass through each other, gain mobility, etc. A collider unspecified prefab was chosen to provide more comfortable control. This asset consists of four wheels and a body. The body includes a Rigidbody property, and the wheel collider property is assigned to each wheel. With these essential configurations and the script to be created, we can give the wheels rotation and other movements.

The canvas component is used to show stable objects on the game screen. We used canvas to show a joystick, the amount of collected gold, and buttons for turning the main menu and exiting the game. To simulate the progress of the car, we desire to add a joystick to the game screen. The downloaded asset for the joystick should be imported into the project as in the previous objects. All these stable views should be sub-objects of the canvas.

The gold game object that should be in the game scene is also imported into the project. On the Inspector tab, the tab where the properties of the game objects are listed, the tag of the gold game object is defined as "gold". By the definition of all game objects will make the code and object management more controlled. Since the gold object will be created and deleted regularly in this game, it will be called by the code with the related game object tag. The components used in the game scene are seen in Figure 8.

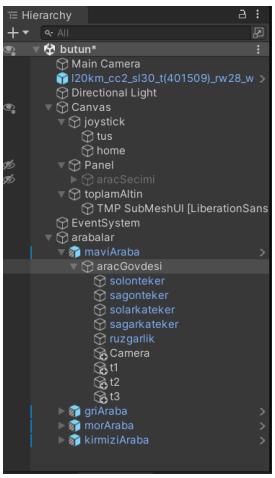


Figure 8. Main Scene Game Objects List in the 3D Game

The primary function defined in the game performs by the script file assigned to the car. Due to the car selection capability in the game, similar scripts have been defined in other car game objects. It can be seen in Figure 9 where Rigidbody, Box Collider, and Scripts are defined on the car body.

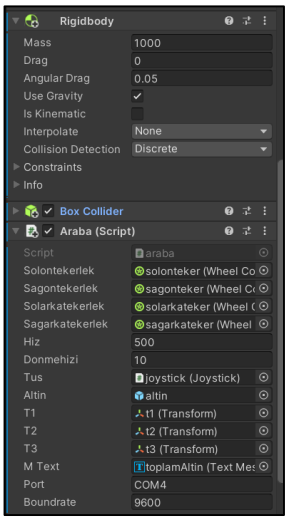


Figure 9. Car Game Object Configurations in Inspector Tab

Scripts can also be created by choosing a script file under the project folder view, Project tab, or by selecting the “Adding a New Script” choice in the Inspector tab, where the relevant object properties are listed. Double-clicking on the relevant file will switch to the Visual Studio platform. Coding is done using the C# programming language. The class is opened with two standard methods’ templates, “Start” and “Update”. In the Start() method, it is necessary to define the opening of the serial port, which will be done once. Since the codes related to reading data will be a recursive process, it should be written into the Update() method.

Libraries and variables related to reading data from the serial port should be defined in the "araba" script file. Another important point here is that the SerialPort(string port name, int baud rate) method should be entered with the values as SerialPort(“COM4”, 9600) that are compatible with the code written in Arduino. The flow charts for reading data from the serial port can be seen in Figure 10 and Figure 11.

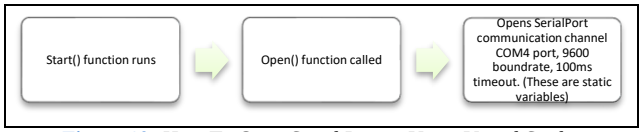


Figure 10. How To Open Serial Port in Unity-Visual Studio

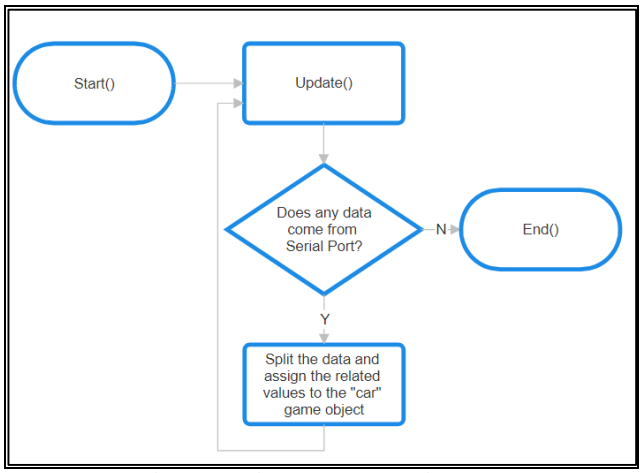


Figure 11. The flow chart of the receiving data from Serial Port process

With the car game object moving with the data from the three-angle gyro sensor, the user will make flexion, extension, and radial-ulnar deviation movements to treat wrist disorders while playing the gold collecting game. The car will move forward with flexion movement, back with extension movement, and right-left with radial-ulnar deviation which can be seen in Figure 12.

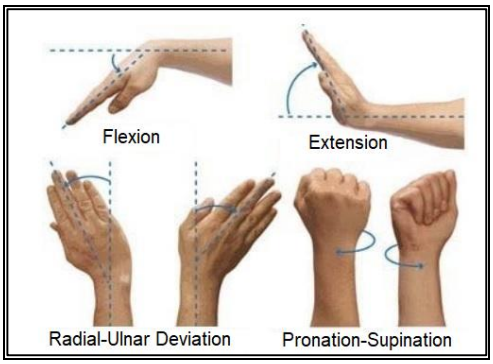


Figure 12. Physical Therapy Movements For Wrist Function Disorders [19]

The players start with the screen seen in Figure 13-a.

Figure 13-a. Start screen for the 3D Game

Figure 13-b. The main game screen of 3D Game

This study has a game view as seen in Figure 13-b. There is a car game object that can be moved by the user's wrist movements. By this way, the user can gain gold and can change the car game object choice. To switch the screens or change the car choice, the user has to use mouse control.

Figure 14. Car options screen.

After the users collected some gold, they can select the other car options, which can be seen in Figure 14.

4. CONCLUSION

There are many people that experienced various physical disorders. As the reviews show us, the number of people who needs physical therapy treatments is increasing day by day [1]. Also, we can see the results that the patients and the physical therapists have preferred game-based treatments or telehealth since Covid-19 according to Staiano and Flynn [17], and Elor, et al. [18]. The addition of game-based components to the tool enhances the interaction and enjoyment of physical treatment. Patients are more likely to remain motivated and involved throughout their rehabilitation process by having games in therapy sessions. Games offer patients a fascinating and engaging setting that diverts them from potential discomfort and boredom. This improves adherence to prescribed activities and makes therapy fun [20].

There is great potential for improving the rehabilitation process with the development of a device built with electronic infrastructure for Bluetooth communication with a computer and the ability to let patients play games while receiving physical therapy for wrist disorders. This technology improves patient engagement, motivation, and compliance while incorporating interactive games into therapeutic activities, resulting in more efficient and fun wrist rehabilitation. Besides these positive sides, unfortunately, these new-generation solutions require a high budget. The physical therapy clinics which decided to use this kind of tool have to pay purchasing budgets and monthly game renting prices. This study aims to show that the budget required to manufacture this tool is not as much as what clinics pay to own these technologies. Production and presentation of original, interesting game choices are essential.

In this study, we showed the current game-based treatment choices and the request for these new treatment methods. We compared their capabilities and budgets. After that, we described our development of a tool that provides physical therapists and wrist disorder patients with an enjoyable, effective, and affordable alternative to treatment. In this version of the study, we used communication with wire. By integrating electronic infrastructure, Bluetooth technology enables easy communication between the tool and a computer. Since real-time data transmission and remote monitoring are made possible by this connectivity, healthcare providers are better able to monitor the progress of their patients and promptly alter their treatment plans. The tool's flexibility and utility will be increased by the wireless and adaptable interface that Bluetooth connectivity offers [21]. As Karagedik et al. showed us in their study, there are different items available to make a connection both Bluetooth and wireless connections [22]. They are also showing these technologies performance measurement in their study. We have the awareness of the wireless usage effect so we are continuing the improvement of the tool. We are planning to develop more game choices, detailed reports for patients and therapists, and compact design for our study's further versions.

By developing these new-generation technologies inside the country, we won't have to spend more money for providing these capabilities from abroad. In this way, the clinics, the physical therapists, and the patients can reach these types of technologies more easily and it will be easier and cheaper to take technical support also it won't need high budgets as today to subscribe to the game offers. Thus, improvement and advancement in new technologies will motivate new developers and investors.

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to our fellow authors Prof.Dr. Mehmet Demirtaş, Assoc. Prof. Dr. Emel Sönmezer, for their invaluable contributions to this research endeavor. Their expertise, dedication, and collaborative spirit greatly enriched the quality and depth of this paper. We extend our sincere appreciation for their unwavering support and insightful discussions throughout the research process. Besides that, thanks to Yunus Emre Ertekin for his guidance

regarding electronical communication architecture. Together, we have achieved a synergy that has significantly advanced our understanding in this field.

REFERENCES

[1] World Health Organization International, "Musculoskeletal health," https://www.who.int, para. 5, July 14, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions#:~:text=A%20recent%20analysis%20of%20Global,and%20rheumatoid%20arthritis%20(1). [Accessed: Sept 01, 2023].

[2] Briggs, A. M. et al., "Health systems strengthening to arrest the global disability burden: empirical development of prioritised components for a global strategy for improving musculoskeletal health," BMJ Global Health, vol. 6, no. 6, p. e006045, Jun. 2021 , doi: https://doi.org/10.1136/bmjgh-2021-006045.

[3] Joyce A, M. et al. "Annual summary of vital statistics--2003." Pediatrics vol. 115,no. 3, p. 619-34, 2005, doi:10.1542/peds.2004-2695.

[4] Rapael Smart Glove Official Web Site, "Rapael Smart Glove Brochure", https://www.neofect.com, pg.4-5. [Online]. Available: https://content.neofect.com/en/wp-content/uploads/2016/03/RAPAEL-GLOVE-Brochure_Eng.pdf.[Accessed: Sept 01,2023].

[5] Hocoma: Advanced Technologies for Movement Rehabilitation, "Hocoma Armeo Brochure", https://www.hocoma.com, pg.23, February, 2019. [Online]. Available: https://knowledge.hocoma.com/wp-content/uploads/2018/12/bro_ArmeoTC_ResearchBrochure_20190218_WEB-1.pdf. [Accessed: Sept 01, 2023].

[6] Fidan, U., and Özkan N.. "Nörolojik rehabilitasyon için kinect sensörlü ölçüm ve egzersiz sisteminin tasarımı ve gerçekleştirilmesi." Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi, vol.18, no. 2, p. 727-733 (2018), doi: 10.5578/fmbd.67249.

[7] Erdoğan, H., and Ekenel, H.K.. "Kinect Kullanarak Fizik Tedavi ve Rehabilitasyon Amaçlı Oyun Tasarımı Game Design for Physical Therapy and Rehabilitation Using Kinect.", TıpTekno'15 Tıp Teknolojileri Ulusal Kongresi vol.6, no.4, p. 288-291, 2015.

[8] Özcan, H., et al. "Çocuklarda El ve Bilek Hareketlerini Geliştirmeye Yönelik Leap Motion Temelli Bir Mobil Oyun Tasarımı A Leap Motion Based Mobile Game Design for Developing Hand and Wrist Movement in Children.", 4th International Symposium on Innovative Approaches in Engineering and Natural Sciences, vol.4, no. 6, pg. 211-219 (2019), doi: 10.36287/setsci.4.6.061.

[9] Okkesim, Ş., et al. "Wii Oyun Konsolunun Tremor Karakteristiğinde ki Etkisi Effect of Wii Game Console on Tremor Characteristics.", Tıptekno'14 Tıp Teknolojileri Ulusal Kongresi, vol.5, no.3, p. 287-290, 2014.

[10] Rand, D., Rachel Kç, and Patrice T. L. W.. "The Sony PlayStation II EyeToy: low-cost virtual reality for use in rehabilitation." Journal of Neurologic Physical Therapy, vol.32, no.4, 155-163, 2008.

[11] Baffin Store, "NEOFECT Smart Glove for Clinic", baffin.co.uk. [Online]. Available: shop.baffin.co.uk/neurorehabilitation/32-669-rapael-smart-glove.html#/1-size-s/26-side-1_pair. [Accessed: Sept 01, 2023].

[12] Sutanto, Y. S., et al. "Videogame assisted exercise training in patients with chronic obstructive pulmonary disease: A preliminary study.", Pulmonology, vol. 25, no.5, 275-282, 2019.

[13] Robotistan Store, "Orijinal Arduino UNO R3", robotistan.com [Online]. Available: https://www.robotistan.com/orjinal-arduino-uno-r3-yeni-versiyon. [Accessed: Dec 07,2023].

[14] Robotistan Store, "MPU6050 6 Eksen İvme ve Gyro Sensörü - GY-521", robotistan.com [Online]. Available: https://www.robotistan.com/mpu6050-6-eksen-ivme-ve-gyro-sensoru-6-dof-3-axis-accelerometer-and-gyros. [Accessed: Dec 07,2023].

[15] Robotistan Store, "Orta Boy Breadboard", robotistan.com [Online]. Available: https://www.robotistan.com/orta-boy-breadboard. [Accessed: Dec 07,2023].

[16] Robotistan Store, "40 Pin Ayrılabilen Erkek-Erkek M-M Jumper Kablo-200 mm", robotistan.com [Online]. Available: https://www.robotistan.com/40-pin-ayrilabilen-erkek-erkek-m-m-jumper-kablo-200-mm. [Accessed: Dec 07,2023].

[17] Staiano, A. E., and Flynn, R.. "Therapeutic uses of active videogames: a systematic review." Games for health journal, vol. 3, no.6, p.351-365, Nov 25, 2014, doi:https://doi.org/10.1089/g4h.2013.0100.

[18] Elor, A., et al., "Physical therapist impressions of telehealth and virtual reality needs amidst a pandemic." Frontiers in Virtual Reality vol. 3, no. 915332, 2022, doi: https://doi.org/10.3389/frvir.2022.915332.

[19] KARABIYIK, Ö., and NAJAFYAN, H.. "Donatı Demiri Tel Bağlama El Aletlerine Ergonomik Bir Tasarım Yaklaşımı." Mühendislik Bilimleri ve Tasarım Dergisi, vol.3, issue.3, p. 245-250, Dec 12, 2015, Retrieved from https://dergipark.org.tr/en/pub/jesd/issue/20874/224032.

[20] Janssen, J. et al. "Gamification in Physical Therapy: More Than Using Games." Pediatric physical therapy: the official publication of the Section on Pediatrics of the American Physical Therapy Association, vol. 29, no.1, p: 95-99, Jan 19, 2017, doi:10.1097/PEP.0000000000000326.

[21] Alliance, Smart Card., "Bluetooth low energy (ble) 101: A technology primer with example use cases." Mobile & NFC Council, SAD, p:21-22, June, 2014, Publication Number: MNFCC-14001.

[22] KARAGEDİK, N., BAL, S., & YAYLA, A.. "Design and Testing of a Wireless Communication Enabled FPGA Development Board: A Comprehensive Education and Application Platform from IoT to Circuit Design." European Journal of Technique (EJT), vol.13, issue.2, p. 123-129, Dec 31, 2023, https://doi.org/10.36222/ejt.1399107.

European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.14, No.1, 2024

Research Article

Classification of filigree silver with Artificial Neural Networks according to production methods

Sabahattin Akgül[1], Hamit Adin[2*] and Hüseyin Ahmetoğlu[3]

[1]Mardin Artuklu University, Midyat Vocational School, 47060, Mardin, Turkey. (sabahattinakgul@artuklu.edu.tr).
[2*] Batman University, Department of Mechanical Engineering, Batman, Turkey. (hamit.adin@batman.edu.tr).
[3]Mardin Artuklu University, Graduate School of Science, Computer Technologies, Mardin, Turkey. (huseyinahmetoglu@artuklu.edu.tr).

ARTICLE INFO

Received: Aug., 02. 2023
Revised: Jan., 14, 2024
Accepted: Jan, 14, 2024

Keywords:
Artificial Neural Networks
Deep Learning
Filigree method
Jewelry

Corresponding author: **Hamit Adin**

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1336397

ABSTRACT

The jewelry industry uses precious stones and metals in various ways while ornaments and jewelry are made. One of the methods used is the filigree method. The most critical factor in the filigree method is human and craftsmanship. However, rapid technological developments make the machine use in filigree mandatory. As a result, filigree products produced by handwork can be created using serial molds in the factory environment. This study aims to classify the molded product filigree silver using artificial neural networks. Filigree products produced by filigree masters and as mold products were compared to distinguish the filigree products. The color of the silver jewelry, the state of the jewelry, the silver setting status, the brass metal used in the silver jewelry, the form of the inner filling motif, the shape of the roof wire, the smoothness of the structure, the proper placement of the inner filling, the symmetrical status of the motifs on the jewelry are trained in the system using Deep Learning, which is an artificial neural networks method through thehe data collected from features such as the use of valuable or worthless stones. The success of classifying filigree jewelry handcrafts or mold products using Deep Learning through artificial neural network methods was evaluated. As a result of the study, the classification with deep learning was conducted successfully.

1. INTRODUCTION

Today, there is an understanding among the public that handmade products have higher quality than ready-made products produced by mass production methods in the factory. For instance, homemade food products are of better quality and tastier than factory-produced foods. At the same time, an entirely handmade car is more durable than cars produced in a factory environment with the help of robots on the mass production line. It can be said that handcrafted jewelry is more durable and of higher quality than mass-manufactured jewelry made out of the mold and that it is preferred among people compared to fabricated products. It is the process of melting filigree silver or gold into rod-shaped molds, obtaining the desired wire thickness by rolling these rods and making the wires into jewelry or ornaments by giving them shapes and patterns with various hand tools. This art is also called vav work or filigree among the public. Historically, the first filigree handcraft works were made in early 3000 BC in Mesopotamia and Egypt. The filigree jewelry processing method is commonly used, especially in Mardin and Midyat regions [1, 2]. The manufacture of filigree jewelry and ornaments, which is a process that requires workforce, is made entirely by hand labor, and requires a lengthy production process. It takes great effort and patience. Even the smallest piece of jewelry can take hours to build. With the development of technology, filigree-looking jewelry can be produced quickly, and a considerable number by using machine mold, and this fabricated product can be launched as if they were real filigree products.

This study aimed to evaluate the filigree products using the distinguishing features of hand-made and mold-made methods with the help of artificial neural networks. The deep learning method of artificial neural networks was used for this application. Many handcrafted and fabricated filigree products were examined, and the filigree masters, who lived in Midyat region and have been carrying out this work for many years, were required to assist in determining the distinguishing features between these two production methods. The resulting data was processed by using the Python programming language. In Figure 1 and Figure 2, pictures of handcrafted filigree and mold work products are shown. Artificial neural networks are computer systems. These systems help to realize the learning function, one of the basic features of the human brain, which has a complex structure. Learning performs its process with the help of networks created by the obtained data. These networks are like the nerve cells in the human brain. Each connection to these networks has a weight value, and the information is stored at

these weights. Artificial neural networks develop a calculation method that differs from traditional calculation forms. These systems can achieve a successful result even when the data is insufficient. Artificial neural networks are one of the most powerful systems in which results are obtained, especially in classification, pattern recognition, filtering of signals, compression, and optimization of data [1-3].

Deep learning is a class of machine learning technique that utilizes many nonlinear hidden layers for supervised or unsupervised feature extraction, transformation, pattern analysis, and classification [3].

Figure 1. Handmade product filigree samples

Figure 2. Mold product filigree samples

Deep learning allows and enables calculation models consisting of multiple processing layers to learn data representation with numerous levels of abstraction. With these methods, speech recognition, visual object recognition, and object detection have vastly improved due to the state of technology in many other areas. Deep learning uses the backpropagation algorithm to show how a machine must change its internal parameters to calculate the representation at each layer from the presentation at the previous layer and explore the complex structure within complex datasets. Deep evolutionary networks are very effective at processing images, video, speech, and audio, while repetitive networks lead over sequential data such as text and address [4, 5].

In the design of artificial neural networks, perceptron structure was asked to be copied like the operational structure of human brain. The weights in the connections allow communication of the information processing elements. Each element has its memory. With all this, computer programs emulate biological neural networks. The Perceptron concept, which we can be called as the smallest module of this program, is essentially a single-layer processing unit [5].

The activation function processes the net input to the cell to determine the output that the cell will produce in response to that input. Since the derivative of the activation function is also used in feedback networks, attention is paid to selecting an easily calculated role to avoid slowing down the calculation. The structure of the artificial neural network is established based on a relational and differential realizable function. This function defines a linear neuron structure [4-6].

Fully connected layer is associated with all of the neurons in the previous layers in the fully connected layer diagram, each of the neurons in the respective layer. If a classification is to be made, the number of neurons in the last fully connected layer to be used must be the number of classes. Experience is related to the importance and effect of variables such as step size, dropout rate, batch size, and number of epochs (all train data is an epoch.) available by trial and error method [4-7].

2. APPLICATION AND METHOD

This study examined handcrafted jewelry produced by the filigree method and filigree-looking mold production jewelry. 100 filigree handmade and 100 filigree mold products were obtained from the market, and the distinguishing features between the two product groups were determined by examining them together with filigree Masters.

The distinguishing features identified are:

1. The color of the silver jewelry
2. Resource status in jewelry
3. Silver setting status
4. Brass used in silver jewelry
5. Internal filling form status
6. Roof wire form smoothness
7. Proper placement of inner filling
8. Symmetrical status of motifs on jewelry
9. Inner filling wire motif tightness condition
10. Use of precious or worthless stones

While filigree handcrafted silver is processed, blacked-out is also preferred in addition to the unique color of silver jewelry. Darkened charms both look beautiful in color and obtain an antique look. For this reason, in some cases, silver jewelry is blacked out by the Silver dimming (oxidation) method. This method is performed by dipping the lightly heated joint into liquid sulfur soap. In addition, with the rotating technique, also called rhodium-plated, some or all of the jewelry is coated with rhodium plum and colored. Although filigree products are produced from gold, it is also possible to come across gold-plated filigree products.

The welding process takes a significant place in handcrafted filigree products. Powder welding is used to boil the inner fillings into the roof. In other burning operations, Rod welding is used. The use of welding in fabricated products is almost non-existent. Only the products removed from the mold in pieces are boiled to be joined together. Also, to add visual value, any figure, shape, letter, or symbol produced from yellow brass material is burned or assembled into silver jewelry.

The silver has an international standard adjustment measure. The silver settings are defined within thousand systems, and the purest silver's value is considered 1000 settings. The silver used in the commercial field has a maximum purity setting of 999. Fabricated filigree products are generally 925carat. Ln the international market, 925-carat silver is also known as Sterling Silver. Handcrafted filigree products made in Mardin and Midyat regions are usually 950 carat.

In some cases, different metals may be boiled or assembled in addition to giving silver jewelry a more beautiful appearance. To add another image and beauty to silver jewelry, copper and zinc alloy brass metal is boiled or assembled in bright yellow color. Lt is mainly seen in filigree fabricated jewelry products.

In handcrafted filigree products, the thickness of the inner filling wire placed inside the roof is generally between 0.2 - 0.3 mm. Because the created inner fill is handcrafted, it may not be ideally placed. The thin wire appearance is obtained in fabricated mold products by designing the mold. A thicker internal filler wire thickness is designed. They can provide excellent processing because manufactured products are created in the mold.

Filigree products are manufactured jewelry. For this reason, there may be situations where the forms, such as smoothness and roundness in the roof parts, cannot be appropriately given. However, since machine mold is designed in filigree products, perfect forms, and shapes can be obtained.

Since human hands make handcrafted filigree products, defects may occur during the installation of the inner filler. In mold products, this possibility can only happen if there is a defect in the mold.

Since handcrafted filigrees are made with human hands, symmetrical defects may arise when the inner filling motifs are installed with roof wires. These errors are minimal in mold products.

Wire tightness distances of inner filling motifs in handcrafted filigree products are less. In fabricated products, the tightness rate is generally higher. Wires weave more frequently.

Precious or worthless stone use in fabricated filigree products is more significant than in handcrafted filigree products. This is because it can be designed to place as many stones as desired in the mold environment. This stone placement process in handmade filigree products is a complex task and requires patience and labor.

Two hundred filigree products and filigree-looking mold products obtained from the market with the help of these distinguishing features were examined in shape and taking into account the silver adjustment measure. The resulting data is shown in Table 1.

TABLE I

TABLE OF DATA ON THE DISTINGUISHING PROPERTIES OF FILIGREE AND MOLD WORK FILIGREE PRODUCTS

product	silver jewelry color	welding condition in jewelry	Silver carat	use of brass material	smoothness of the inner filling motif	smoothness of the framework wire	inner filling placement	symmetrical status of motifs on jewelry	inner filling tightness ratio	precious stone	result
	silver colour : 1	available: 1	925	available: 1	mold product regular:1	mold product regular:1	mold product regural:1	symmetrical : 1	little: 1	available: 1	
	blackout:2	unavailable:2	950	unavailable:2	mold product little defective:2	mold product little defective:2	mold product little defective:2	asymmetric: 2	very: 2	unavailable: 2	
	rhodium - plated:3	Additional piece welding available:3			handmade regular:3	handmade regular:3	handmade regular:3				
	gold-plated :4				handmade little defective:4	handmade little defective:4	handmade little defective:4				
	silver color and gold-plated : 5										
1	1	1	950	2	3	3	4	2	2	2	1
2	1	1	950	2	4	4	4	2	2	2	1
3	1	2	925	2	1	1	1	1	1	2	1
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
198	1	1	950	2	4	4	3	2	2	2	1
199	1	1	925	2	3	3	3	1	2	2	0
200	2	2	925	2	1	1	1	1	2	1	0

In the light of these distinguishing features, the determined data of the compared jewels are listed with the help of Excel software, and the data set has been created. Results were observed with these used data. An educational model was created with the Keras and TensorFlow libraries, an open-source neural network using Python programming. Artificial neural networks are given to the system using the deep learning model.

The confusion matrix is a frequently used method to describe the generated performance of a classification model in which actual values are known on a set of test data. It is easy to understand, but its terminology is complicated. The confusion matrix we performed in this study is shown below:

Confusion Matrix

[[13 0]

[0 22]]

50/50 [==========]- 0s 100us/step

Test set "accuracy" value 1.00

Test set "loss" value 0.00

In this study, 75% of the data set is separated for education while 25% is for the test. 200 samples were used for jewelry: 150 for training and the rest 50 for testing.

TABLE II
DATA SET TABLE OF FILIGREE PRODUCTS AND MOLD WORK FILIGREE PRODUCTS

Out[1]:

	0	1	2	3	4	5	6	7	8	9	10	11
0	1	1	1	950	2	3	3	4	2	2	2	1
1	2	1	1	950	2	4	4	4	2	2	2	1
2	3	1	2	925	2	1	1	1	1	1	2	0
3	4	2	3	925	1	2	1	2	1	1	1	0
4	5	2	3	925	2	1	2	1	1	1	1	0

TABLE III
DATA SET ENTRY TABLE OF FILIGREE PRODUCTS AND MOLD WORK FILIGREE PRODUCTS

Out[2]:

	0	1	2	3	4	5	6	7	8	9	10
0	1	1	1	950	2	3	3	4	2	2	2
1	2	1	1	950	2	4	4	4	2	2	2
2	3	1	2	925	2	1	1	1	1	1	2
3	4	2	3	925	1	2	1	2	1	1	1
4	5	2	3	925	2	1	2	1	1	1	1

3. RESULTS

In this study, handcrafted filigree products were found in the market, and mold filigree-looking products were provided; and then these products were examined, and a data set was obtained. An educational model was created by using Python Software with Keras and TensorFlow libraries, an open-source neural network. Artificial neural networks were given to the system using the deep learning model. In this developed system, the classification of filigree jewelry was successfully carried out. In the following figures (Figure 3 and 4), the classification success chart and the loss value chart that took place during training were shown. As can be seen from the figures, the success rate was achieved at a high rate of approximately 100%.

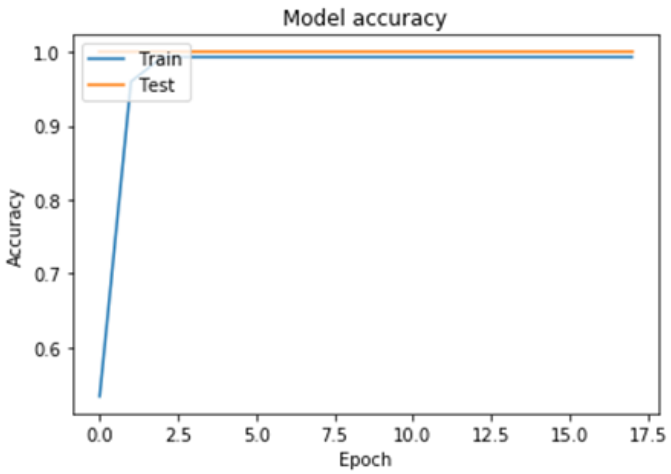


Figure 3. Classification success chart

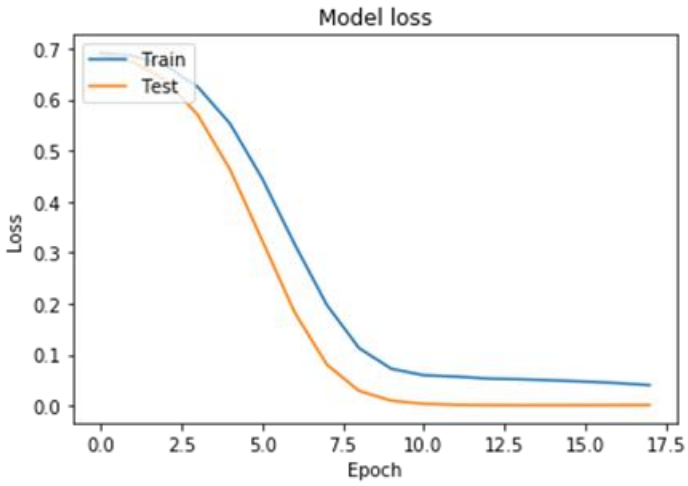


Figure 4. Loss value graph

As seen in Figure 3, with obtained data, the classification success of the computer is high.

In Figure 4, curves showing lost value data are given. The desired criterion in the lost value chart is that the angles decrease. As understood in the chart, the line of test and training data shows that success is in the desired results.

Creating a data set with more details made us achieve more realistic results. More distinguishing features can be added to improve the data set in this study. For more information, specific results can be obtained by adding data such as the determination of alloying rates of silver jewelry, the type of visual materials and precious stones used, the chemical structure of jewelry, and measurement values. Also, image processing can be carried out along with the application of deep learning in artificial neural networks. A data set can be created by processing many handcrafted filigree and mold work filigree images on the market. As a result, good results can be achieved.

REFERENCES

[1] Türe A, Savaşçın MY. Birth of jewelery Goldaş publications 2000.
[2] Öztemel E. Artificial neural networks Papatya publications April 2012.
[3] Deng L, Yu D. Deep Learning: Methods and Applications, vol. 7. 2013.
[4] LeCun Y, Bengio Y, Hinton G. Deep learning Nature İnternational journel of science pages 436–444 (28 May 2015)
[5] Goodfellow I. "Chapter06 Deep Feedforward Networks," Deep Learning Book, no. 1, pp. 169–229, 2015.
[6] Buduma N, Locascio N. Fundamentals of Deep Learning, vol. 521. 2015.
[7] Ahmetoğlu H, Daş R. Classification of Attack Types from Big Data Sets with Deep Learning 2019 International Artificial Intelligence and Data Processing Symposium (IDAP)

BIOGRAPHIES

Hamit ADİN born in 1972, received his PhD degree from the University of Fırat, Elazığ, Turkey in 2007 and has been Professor of Mechanical Engineering at the University of Batman, Turkey, since 2021. He has done research in the areas of mechanics, composite materials, adhesive, adhesion and finite element analysis. His research includes both theoretical and experimental studies.

European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.14, No.1, 2024

Research Article

Medical Insurance Cost Prediction MedCost: Machine Learning Ensemble Approaches

Murat EMEÇ[1*]

[1*]Istanbul University, Department of Information Technology, 34116, Fatih, Istanbul, Turkey. (e-mail: murat.emec@istanbul.edu.tr).

ARTICLE INFO

Received: Oct., 16. 2023
Revised: Dec., 31. 2024
Accepted: Jan, 14. 2024

Keywords:
Medical cost
Healthcare insurance
Machine Learning
Ensemble methods
Regression

Corresponding author: *Murat EMEÇ*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1375677

ABSTRACT

Healthcare insurance costs are a significant concern for individuals and providers. Accurately predicting these costs can assist in financial planning and risk assessment. This study explores machine learning ensemble methods to predict healthcare insurance costs based on various factors, including age, sex, body mass index (BMI), number of children, smoking status, and region. Additionally, new features were introduced by incorporating the mean and standard deviation of BMI and smoking habits, which are known to affect insurance costs substantially.

The study began with a comprehensive statistical analysis of the dataset, followed by feature engineering to enhance its predictive power. Categorical variables such as sex, smoking status, and region were appropriately encoded. Two datasets were constructed: one containing all the original features and the other having the engineered features. Ensemble learning methods, including Bagging, Stacking, and the proposed MedCost-AdaBoost model, were employed to predict the insurance costs for both datasets. The results revealed that the MedCost-AdaBoost model outperformed the other methods in terms of lower Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values, along with higher R-squared (R^2) scores.

These findings underscore the effectiveness of ensemble learning techniques in predicting healthcare insurance costs, with feature engineering playing a crucial role in improving prediction accuracy. Despite certain limitations, such as the dataset size, this study provides valuable insights for researchers and professionals in the healthcare insurance industry. Future research could explore additional factors and larger datasets to enhance the predictive models in this domain further.

1. INTRODUCTION

Health insurance is a vital financial tool that facilitates access to healthcare services and offsets unexpected medical expenses. However, health insurance companies' risk assessment and premium determination processes are complex and involve several factors. These factors include the insured's age, sex, medical history, region of residence, insurance plan coverage, etc. Machine learning and data analytics play essential roles in health insurance cost estimation.

Health insurance companies use big data and machine-learning techniques to offer fairer premiums to customers and better manage their financial risks. Therefore, health insurance cost estimations are essential for insurance companies and PHs. This study focuses on health insurance cost estimation and examines methods for improving these estimates using machine learning ensemble approaches. Additionally, the importance of health insurance cost estimation and the advantages of accurate estimation are essential areas of discussion.

Health insurance cost estimation is essential for several reasons.

- Fair Premium Setting: Health insurance companies must offer fair and competitive premiums to their customers. Accurate cost estimates help set insurance premiums fairly. This implies more favorable and competitive insurance prices for customers.
- Risk Management: Health insurance companies use accurate cost estimates to forecast payments and manage financial risks. Inaccurate estimates can cause financial difficulties for insurance companies.
- Insured Making Informed Decisions: Individuals need accurate information to understand their health insurance options and costs. Precise cost estimates help insureds make informed decisions.

Access to Health Services: Accurate cost estimates can help insurance companies provide individuals access to affordable

healthcare services. This contributes to making healthcare services accessible to the broader population.

Therefore, accurate and precise health insurance cost estimations are essential. Machine-learning ensemble approaches offer a way to improve these estimates and make better decisions. This study examines how health insurance cost estimation can be improved and how ensemble methods can be used.

The primary motivation and contributions of this study are as follows:

- Accurately estimating health insurance costs is challenging with current diagnostic methods. In particular, individuals' health profiles are complex and include age, sex, smoking habits, BMI, and other essential factors.
- Traditional methods are often inadequate for estimating health insurance costs. For example, some models do not adequately consider certain factors, while others cannot accurately identify effects and relationships.
- MedCost-AdaBoost, the method presented in this study, can make more precise estimates, especially by considering certain health profile factors. In particular, with new features, including determinants such as BMI and smoking habit, it has a predictive capability beyond traditional methods.
- This study aims to contribute to related studies to estimate health insurance costs better. The highlights of MedCost-AdaBoost contribute to more accurate financial planning and risk assessment in industry and academia.

The remainder of this paper works as Section 2. In Section 3, the materials and methods used are described. Section 4 presents the results and evaluation. A discussion and comparison are presented in Section 5. Finally, conclusions are presented in Section 6.

2. RELATED WORKS

Emerging technologies and growing data sources have become important research topics in health insurance cost estimation, machine learning, and data analytics. This section reviews the literature on health insurance cost estimations and machine-learning ensemble methods. Health insurance cost estimation is essential for healthcare providers, insurance companies, and individuals, and accurate estimates play a critical role in financial planning and resource management. In this context, they understand the contributions of previous studies on health insurance cost estimation and how machine learning ensemble methods can be used to form the basis of this research.

Research using various machine learning algorithms and data mining techniques has enriched the literature on health insurance and cost estimations. Studies such as those in [1,2] have highlighted how traditional machine learning algorithms can predict insurance claims and medical insurance prices. To improve the interpretability of the models, explainable artificial intelligence (XAI) methods were investigated, and health insurance datasets were considered [3]. XGBoost and machine learning-based regression methods have been investigated for health insurance premium prediction [4,5] and focused on current methods of estimating health insurance costs [6-8]. Finally, studies have been conducted to assess health insurance claim costs using data mining techniques [9-10]. Related studies reflect various health insurance cost estimation approaches and shed light on future research. In this context, to update the literature, we have also included critical recent works [11-17]. These references reflect current applications of ensemble methods for estimating health insurance costs and new developments in the research field. Recently, many studies have focused on regression techniques based on machine learning techniques. [18-21].

A review of these studies showed that machine learning and data mining techniques have been widely used in health insurance cost estimations, and significant results have been achieved. However, each of these approaches has advantages and limitations. Therefore, this study adopted a strategy called ensemble modeling. Ensemble modeling allows us to reach more robust and stable predictions by combining different machine learning algorithms. This study uses this approach to evaluate stacking, bagging, and boosting ensemble methods and examines their impact on health insurance cost prediction. The results based on the ensemble model can significantly improve health insurance cost estimations by providing more precise and reliable estimates.

3. MATERIALS AND METHODS

This section presents detailed information on the dataset, methods, and analyses used for health insurance cost estimation. The main objective of this study was to estimate health insurance costs using different ensemble methods and evaluate the performance of these estimates. First, the dataset's sources, components, and importance are explained. The data preprocessing steps, feature engineering techniques, and ensemble methods are described in detail. This study's health insurance cost estimation methodology was carefully designed and implemented to obtain accurate and reliable estimates.

3.1. Materials

3.1.1. Data Set and Features

The dataset used to understand and estimate the key factors affecting health insurance costs and the characteristics of this dataset are critical. This section discusses the data sources that form the basis of our health insurance cost estimation analysis and the essential features of these sources. To better understand the various factors determining health insurance costs, we examined patient characteristics such as age, sex, body mass index (BMI), number of children, smoking habits, region of residence, and individual health costs. These characteristics provide the primary data for estimating and analyzing health insurance costs. Therefore, by exploring this dataset and its characteristics more closely, we aim to better understand these variables' impact on health insurance cost estimation. The features of the dataset are as follows.

- Age is an essential factor in health insurance costs. Usually, older individuals pay higher insurance fees because of the increased risk of health problems.
- Sex: Sex can affect health insurance fees. Women, especially those of childbearing age, may tend to use more health services, but the effect may vary depending on other factors.
- Body mass index (BMI): BMI assesses whether a person has a healthy body weight. A high BMI can indicate an increased risk of chronic disease, leading to higher healthcare costs.

- Child: The number of dependent children can affect insurance costs. More dependents usually pay higher insurance costs.
- Smokers: Smoking is a risk factor for many health problems, and smokers often have higher health insurance rates.
- Region: The region where you live can affect health insurance costs. The costs of living, healthcare costs, and health habits in different areas can vary.
- Charges: Individual medical costs incurred by health insurance. All other factors influence them, and the main objective of this analysis is to understand the relationship between independent variables (age, sex, BMI, children, smokers, and region) and costs.

Assessing the impact of these essential features and datasets on health insurance cost estimation is one of the central aims of this study. Using these characteristics, we attempt to estimate health insurance costs by evaluating different ensemble methods. A better understanding of how age, sex, BMI, number of children, smoking habits, and region of residence affect health insurance costs is needed to provide an essential roadmap for future health insurance planning and decisions.

3.1.2. Statistical analysis of the dataset

This section focuses on the statistical analysis of the dataset used to estimate health insurance costs. The statistical properties of the dataset, its distribution, and basic summary statistics formed the basis of this analysis. We will also use correlation analysis to examine the relationships between the features and visual studies to understand the overall structure of the dataset better. These statistical analyses will help us to understand the factors influencing health insurance costs and how these factors are interrelated. This approach helps us identify the variables of interest when building future forecasting models.

TABLE I
THE STATISTICAL DISTRIBUTION OF THE MEDICAL COST DATASET

Feature(s)	Count	Mean	STD	MIN	MAX
AGE	1338	39.207	14.049	18	64
SEX	1338	0.5	0.5	0.0	1
BMI	1338	30.664	6.097	1.596	5.313
CHILDREN	1338	1.094	1.205	0	5
SMOKER	1338	0.2	0.4	0	1
REGION	1338	1.5	1	0	3

The statistics in Table 1 describe the distribution and key features of each characteristic of the dataset. For example, age generally comprises young and middle-aged individuals, sex is evenly distributed, BMI is usually between 26 and 34, and the proportion of smokers is low. It also shows that wages are spread over a wide range and have large standard deviations. These statistics help us better understand the characteristics of the dataset and allow us to prepare for the analysis. In particular, the wide distribution of wages is essential for studying their impact on cost estimations, as this can help us understand how health insurance costs are associated with different factors. More advanced statistical methods, such as correlation and visual analyses, will help us examine the relationships between these characteristics in more detail.

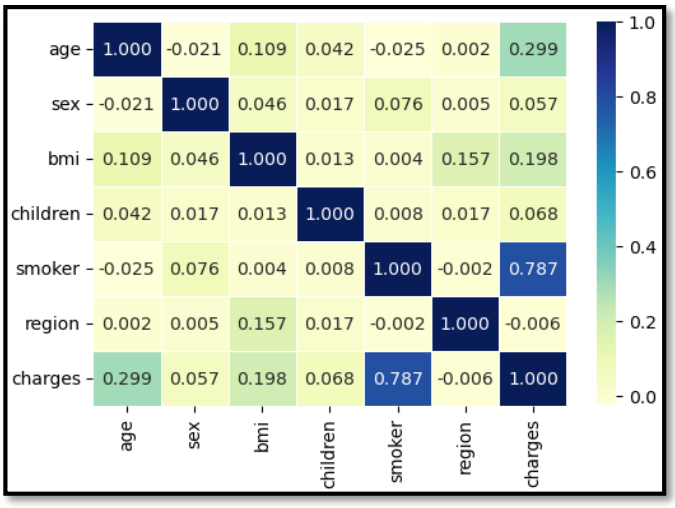


Figure 1. The correlation matrix of the medical cost dataset

The correlation matrix shown in Figure 1 measures the relationships between the variables in the dataset. This correlation determines the strength and direction of the relationship between two variables. The following is an interpretation of the correlations in the matrix:

- Age–charge correlation (0.299): A positive correlation of 0.299 was shown between age and charge. This finding indicates that insurance charges generally increase with age.
- Sex-charge correlation (0.057): A weak positive correlation exists between sex and charges. However, this correlation indicates that sex is not among the most critical factors in determining charges.
- Body mass index (BMI) – was correlated (0.198): A positive correlation of 0.198 existed between BMI and medical status. This finding indicates that higher BMI is generally associated with higher health insurance charges.
- Child-charge correlation (0.068): A weak positive correlation exists between the number of children and charges. However, this correlation was also lower than that of the other factors.
- Smokers-charge correlation (0.787): A strong positive correlation exists between smokers and charges. Smokers generally pay higher health insurance costs.
- Region charge correlation (-0.006): A very low negative correlation exists between the region and charges. This indicates that the region's influence on determining the charges is almost negligible.

These correlations show the relationships between factors that affect health insurance costs. Smoking and BMI appear to have a more significant impact on wages. Predictions. The grid search method was preferred for hyperparameter tuning. The MedCost-AdaBoost method combines weak learners with weights to develop a robust forecasting model. This model positively impacts the ability of the model to predict health insurance costs more accurately. The results and evaluation section presents the results and performance of the proposed method in more detail.

3.1.3. Feature engineering

Feature engineering, an essential step in the data analysis and modeling process, is being applied to estimate health insurance costs more accurately and better understand the factors behind these costs. In this phase, we plan to add new features and transform the existing features to enrich and better understand our current dataset. This approach allows us to understand better the critical factors that influence health insurance costs, thus enabling us to construct more powerful and precise predictive models. We will also code categorical variables appropriately, making them understandable in our model. We are fortunate in handling incomplete data because our dataset is complete, which allows us to obtain more reliable results. At this stage, we aim to step forward in health insurance cost estimation by making our dataset more robust and meaningful. The feature engineering stages are listed below:

- Adding New Features: To examine the effects of BMI and smoking on wages, new features were created from these two features. This is logical because these two features can significantly affect health insurance costs.

- Adding the mean and standard deviation: Adding the mean and standard deviation for BMI and smoking status is essential for measuring the distribution and variability of these characteristics in the dataset. This makes it easier to understand how common these characteristics are and how much they can vary.

Feature engineering is an essential step toward making the analysis and modeling of the dataset more robust [22]. In particular, adding new features and a better understanding existing features can improve the model's ability to predict health insurance costs. In addition, coding categorical variables can help the model better use the tables.

3.1.4. Data preprocessing

Data preprocessing, one of the critical steps in this study, forms the basis of our model for estimating health insurance costs. At this stage, we analyzed, cleaned, and prepared our dataset. First, we code our categorical variables appropriately and put them into a format machine-learning algorithm. In terms of missing data processing, there were no missing data, which increased the reliability of the dataset. New features are added using feature engineering methods so that our model can be fed with more information. Finally, we split our dataset appropriately for model training and evaluation. Data preprocessing is a fundamental step in more accurately predicting health insurance costs and is essential in helping us increase the model's power.

The steps listed below describe the key stages of the data preprocessing process and indicate that you are preparing to migrate to the model.

The following are descriptions of these steps:

- Encoding Categorical Variables: Convert categorical variables such as "sex," "smoker," and "region" into numeric values via the encoder method and introduce these variables into a format that machine learning algorithms can understand.
- Missing value: No missing data in the dataset means incomplete data processing methods are unnecessary. This means the dataset is clean and complete and helps the machine learning model produce reliable results.
- Creating datasets: Creating a dataset with new features added using feature engineering methods helps the model obtain more information. By creating two datasets, one has two separate datasets containing the original and new features, allowing us to examine how the model performs in different scenarios.
- Splitting the dataset: Separating the training and test datasets was essential for training and evaluating the model. Thus, we can determine the generalization ability of the model.

The data are divided as follows:

  - Train set (80%): 1.070
  - Test set (20%): 268

These preprocessing steps form the basic infrastructure of the health insurance cost estimation model. Using these datasets, one can train the machine learning algorithms and proceed to the evaluation phase.

3.2. Methods

3.2.1. Ensemble methods

This study uses various ensemble methods to estimate health insurance costs. These methods allow for more robust forecasting by combining multiple underlying models. The following is a brief introduction to these methods:

- Bagging (Bootstrap Aggregating): Bagging is an ensemble method in which several base models are trained independently on the same dataset, and their predictions are subsequently aggregated. Each base model was trained on a random dataset sample; thus, different models were generated with different samples. This method can increase the stability of the model while reducing the variance.
- Stacking: Stacking is an ensemble method in which the predictions of multiple base models are combined using a higher-level model (meta-model). Base models can be built on different features or algorithms, and metamodels use the predictions of these base models to generate the final prediction. This method is effective when different models complement each other and achieve better forecasting performance.
- Boosting: Boosting is an ensemble method that combines weak forecasting models to construct a robust model. The base models were trained sequentially, and each model was introduced by focusing on the errors of the previous model. In this way, the number of mistakes is reduced, and more robust predictions are obtained. Popular boosting methods include AdaBoost, gradient boosting, and XGBoost.

This study performed experiments to predict health insurance costs using ensemble methods. The performance of each method was evaluated separately, and the results are presented comparatively.

3.2.2. Proposed MedCost-AdaBoost Ensemble Methods

This section details the proposed AdaBoost ensemble method for forecasting health insurance costs. AdaBoost is an efficient ensemble method that aims to construct a robust prediction model by combining weak learners. The proposed architecture utilizes vulnerable learners, which are decision trees. Decision trees are suitable options for capturing the complexity of a dataset and understanding the effects of critical features.

The proposed model architecture includes a customized structure, as shown in Figure 2, to predict health insurance costs. This model architecture consists of three primary layers: an input layer, a MedCost-AdaBoost ensemble layer, and an output layer.

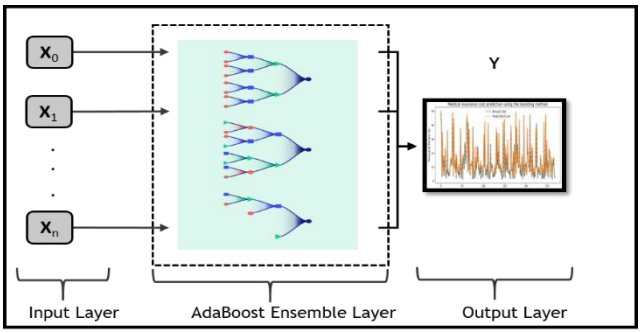


Figure 2. Proposed MedCost-AdaBoost Ensemble Architecture

- Input Layer: The first layer of the model, the input layer, contains the dataset's characteristics. These features include factors that affect health insurance costs, such as age, sex, body mass index (BMI), number of children, smoking status, and region (X0,…, Xn). The input layer uses these features and starts processing them.
- MedCost-AdaBoost Ensemble Layer: The layer that forms the core of the model contains the AdaBoost ensemble method. AdaBoost is an ensemble method that combines weak learners with a robust prediction model. This layer connects vulnerable learners like decision trees to obtain more robust predictions.
- OutPut Layer: The final layer of the model, the output layer, is where the predictions are generated and the results are obtained. This layer makes predictions from the AdaBoost Ensemble Layer and predicts medical insurance costs (Y).

The model architecture in Figure 2 was customized and optimized to predict health insurance costs by applying the AdaBoost ensemble method using the features in the dataset. Setting specific parameters during the training of the AdaBoost model is important. These parameters affected the performance and stability of the model. In particular, parameters such as the depth of the trees, learning rate, and number of estimators (n_estimators) must be chosen carefully [23]. These parameters help the model avoid overfitting and produce accurate predictions. The grid search method was preferred for hyperparameter tuning. The optimal parameters after tuning are listed below:

- Max_depth=5
- Max_features='auto'
- Max_leaf_nodes=70
- Min_samples_leaf=10
- Min_weight_fraction_leaf=0.1
- Splitter='best'
- N_estimators=10
- Learning_rate= 0.01

The MedCost-AdaBoost method combines weak learners with weights to develop a robust forecasting model. This model positively impacts the ability of the model to predict health insurance costs more accurately. The results and evaluation section presents the results and performance of the proposed method in more detail.

4. RESULTS AND EVALUATION

4.1. Evaluation Metrics

In this study, which used the AdaBoost ensemble method to predict health insurance costs, different model parameters and feature engineering approaches were examined. We are now exploring the results of this study to assess the effectiveness of this method and its impact on the ability to predict health insurance costs. This study highlights the importance of using various evaluation metrics to measure a model's performance and understand how it can be integrated into real-world applications [24].

1. Mean absolute error (MAE): MAE is a metric that measures how far the predictions are from the actual values. A lower MAE indicates a better prediction model.

$$MAE = \frac{\Sigma\,(p - a)^2}{n} \quad (1)$$

In Formula (1):

- p: predicted value
- a: actual value
- n: represents the number of observations.

2. Root mean squared error (RMSE): The RMSE is a metric that measures how much error a prediction makes relative to the actual values. A lower RMSE indicates a better prediction model.

$$RMSE = \sqrt{\left(\frac{\Sigma\,(p - a)^2}{n}\right)} \quad (2)$$

The values of p, a, and ne in formula (2) were used as defined in formula (1).

3. R^2 (R-squared): R^2 is a metric that measures how much of the variance in the dataset is explained by the prediction model. It has a value between 0 and 1, with higher R^2 values indicating better model performance:

The formula R^2:

$$R^2 = 1 - \frac{SSR}{SST} \quad (3)$$

- The sum of squares of residuals (SSR) represents the sum of squares of the actual value and forecast differences.
- Total Sum of Squares (SST): represents the sum of squares of the deviations from the mean of the actual values.

These are the standard metrics used to evaluate model performance and measure the accuracy of predictions. The performance of the proposed model was assessed in detail using these metrics.

4.2. Evaluation of model performance

The performance and model results of different ensemble methods for estimating health insurance costs were evaluated. The work is based on the original dataset containing all features and the MedCost dataset, which results from feature engineering. Bagging, stacking, and the proposed MedCost-AdaBoost model were applied to both datasets. For each ensemble method, mean absolute error (MAE), root mean squared error (RMSE), and R2 were used to evaluate the model's performance. These results will help us better understand which ensemble method is more effective for both datasets and how dataset characteristics affect the prediction performance. The results of the ensemble methods for both datasets are presented in Tables 2 and 3, respectively.

TABLE II
ALL-FEATURES DATASET

Ensemble Method(s)	MAE	RMSE	R^2
Bagging	2.4850	4.6641	84.46
Stacking	2.7359	4.5078	85.48
MedCost-AdaBoost	2.3052	4.2500	87.10

The evaluation metrics in Table 2 show the different results for each ensemble method. In this study, we evaluated these results. MedCost-AdaBoost has lower MAE and RMSE values than the other two methods. This indicated that the model predictions were closer to the actual values. In addition, the R2 value of MedCost-AdaBoost was greater than that of the others. This suggests that the model can explain a more significant proportion of the variance of the dataset, signaling a better prediction performance. Stacking performed moderately in terms of MAE and RMSE values, with a slightly higher R2 value. This offered an advantage over the other two models. Bagging was higher than the others in terms of MAE and RMSE values and had a lower R2 value, indicating that the predictions were less accurate. In conclusion, the MedCost-AdaBoost model shows better forecasting performance than the results obtained using all features. This model appears to be the preferable ensemble method for estimating health insurance costs.

TABLE III
MOST-ADABOOST FEATURE LEARNING DATASET

Ensemble Method(s)	MAE	RMSE	R^2
Bagging	3.9538	4.9418	89.75
Stacking	2.4656	4.7428	90.56
MedCost-AdaBoost	0.6749	0.9144	95.75

The evaluation metrics in Table 3 show the different results for each ensemble method. These results were evaluated as follows:

- MedCost-AdaBoost: This model performed exceptionally well on the MedCost dataset using all features. Both the MAE and RMSE values were shallow, and the R^2 value was high. This indicates that the model predictions are very close to the actual values and can explain a large part of the variance of the dataset.
- Stacking: The Stacking model showed moderate performance regarding MAE and RMSE, but the R^2 value was relatively high. This model offers more precise predictions than other methods.
- Bagging: The Bagging model is higher than MedCost-AdaBoost and Stacking regarding MAE and RMSE values and has a lower R^2 value. This indicates that the model's predictions are less precise and can explain less variance than those of the other two models.

In conclusion, the MedCost-AdaBoost model performed best on the MedCost dataset. In particular, the MAE and RMSE values were shallow, and the R2 value was extremely high. This model appears to be an effective ensemble method for estimating health insurance costs. In Tables 2 and 3, the MedCost-AdaBoost model after feature engineering on the MedCost dataset performs best when all features are used and after feature engineering. This model can better predict health insurance costs by improving the accuracy and precision of the predictions. Although the Bagging and Stacking models performed well, MedCost-AdaBoost was more effective.

5. DISCUSSION AND COMPARISON

This study aims to evaluate different ensemble methods for estimating health insurance costs using two other datasets. First, the results were obtained using Bagging, Stacking, and the proposed MedCost-AdaBoost model on the original dataset containing all the features. The same methods were applied to a new dataset created after feature engineering. We can evaluate which ensemble method and dataset performs better by comparing the results obtained for the two datasets.

Although the first dataset (All Features) performed well for the Bagging and Stacking models, the prediction performance of these models was lower than that of MedCost-AdaBoost. The MedCost-AdaBoost model obtained this dataset's lowest MAE and RMSE values and stood out as the model with the highest R2 value. These results show that the proposed ensemble method, MedCost-AdaBoost, can predict health insurance costs better when all features are used.

The new dataset created after feature engineering (Without-Feature Engineering) performed better, with higher R2 values for the Bagging and Stacking models. However, the MedCost-AdaBoost model still achieved the lowest MAE and RMSE values for this dataset, and the R2 value was high. This shows that the MedCost-AdaBoost model exhibited the best prediction performance on the without-feature engineering dataset.

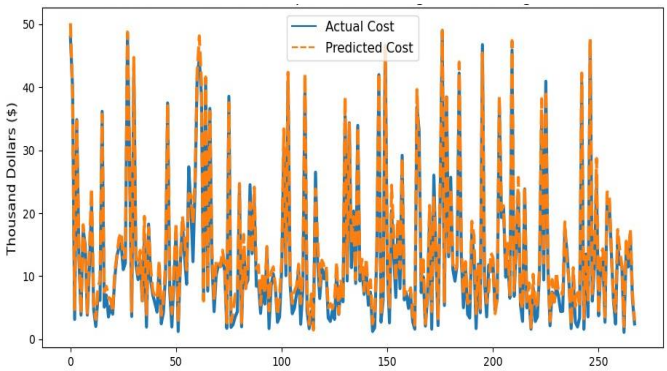


Figure 3. Forecast graph of Proposed MedCost-AdaBoost

Figure 3 shows the relationship between estimated and actual health insurance costs. The x-axis of the graph represents the "Actual Cost" values, and the y-axis represents the "Estimated Cost" values. An overview of the graph shows that, in general, the costs are in reasonably good agreement with the estimates. Instead of forming a line between the actual and estimated expenses, the dots appear densely packed. This shows that the forecasts are generally close to the actual costs and that the model performs well. However, there was a more significant deviation between the actual and estimated costs at some points. In particular, some estimates are lower than the actual costs of high-attraction insurance policies. These points may need to be considered, and the model may need to be improved to predict such cases better. Overall, this graph shows that the model successfully predicts health insurance costs and generally provides good agreement with actual expenses. However, some aspects can be further improved.

In conclusion, this study emphasizes that the MedCost-AdaBoost model is an effective ensemble method for predicting health insurance costs and performs well on a without-feature engineering dataset. This model can be a valuable tool for the cost estimation of health insurance companies and healthcare providers. Although the Bagging and Stacking models also performed well, the MedCost-AdaBoost model was more efficient and could make more precise predictions.

This study has some limitations. First, the dataset size is limited, which may limit its generalizability and require validation on a larger dataset. Moreover, the choice of characteristics used in the study may have excluded some other essential factors. For example, not including genetic factors, chronic diseases, or socioeconomic status may affect the model's predictive power. Therefore, future studies with more extensive and diverse datasets and a more comprehensive selection of characteristics may help make these models more robust and generally valid.

6. CONCLUSION

This study examined the usability and performance of machine learning methods in predicting health insurance costs. The dataset included factors such as age, sex, body mass index (BMI), number of children, smoking habit, and region and aimed to understand and predict the impact of these factors on health insurance costs. The statistical analysis of the dataset was performed at the beginning of the study. This analysis allowed us to examine the basic statistical properties and the distribution of the dataset. Subsequently, new features are added to the dataset through a feature engineering step. In particular, we have added the mean and standard deviation values of BMI and smoking status. Categorical variables were coded appropriately.

The main focus of this study was to evaluate the performance of ensemble learning methods. Health insurance costs were estimated on both datasets using Bagging, Stacking, and the proposed MedCost-AdaBoost model. The results show that the MedCost-AdaBoost model outperformed the other two methods. The MedCost-AdaBoost model achieved lower MAE and RMSE values and higher R2 values on the dataset with all features and the dataset generated after feature engineering. These results emphasize that ensemble learning methods are practical for predicting health insurance costs. This also shows that feature engineering can improve the model's performance and is vital in obtaining accurate predictions. However, this study has some limitations. The limited size of the dataset and exploration of potential possibilities for adding more features provide an opportunity for future work. Moreover, considering other factors affecting insurance costs and using a larger dataset may improve forecasting performance.

In conclusion, this study shows that ensemble learning methods are practical tools for predicting health insurance costs, and feature engineering plays a vital role in obtaining accurate predictions. This study can be a valuable reference for researchers and professionals seeking to improve cost estimations in the health insurance industry. Future studies can advance research by considering more factors and working on larger datasets.

ACKNOWLEDGEMENT

I want to thank my dear family, wife, and kids for their patience with my academic studies.

REFERENCES

[1] Saraswat, B. K., Singhal, A., Agarwal, S., & Singh, A. (2023, May). Insurance Claim Analysis Using Traditional Machine Learning Algorithms. In 2023 International Conference on Disruptive Technologies (ICDT) (pp. 623-628). IEEE.

[2] Vijayalakshmi, V., Selvakumar, A., & Panimalar, K. (2023, January). Implementation of Medical Insurance Price Prediction System using Regression Algorithms. In 2023, the 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 1529-1534). IEEE.

[3] Bora, A., Sah, R., Singh, A., Sharma, D., & Ranjan, R. K. (2022, October). Interpretation of machine learning models using xai-a study on health insurance dataset. In 2022, the 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) (pp. 1-6). IEEE.

[4] Jyothsna, C., Srinivas, K., Bhargavi, B., Sravanth, A. E., Kumar, A. T., & Kumar, J. S. (2022, May). Health Insurance Premium Prediction using XGboost Regressor. In 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (pp. 1645-1652). IEEE.

[5] Kaushik, K., Bhardwaj, A., Dwivedi, A. D., & Singh, R. (2022). Machine learning-based regression framework to predict health insurance premiums. International Journal of Environmental Research and Public Health, 19(13), 7898.

[6] Chittilappilly, R. M., Suresh, S., & Shanmugam, S. (2023, May). A Comparative Analysis of Optimizing Medical Insurance Prediction Using Genetic Algorithm and Other Machine Learning Algorithms. In 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (pp. 1-6). IEEE.

[7] Albalawi, S., Alshahrani, L., Albalawi, N., & Alharbi, R. (2023). Prediction of healthcare insurance costs. Computers and Informatics, 3(1), 9-18.

[8] Praveen, M., Manikanta, G. S., Gayathri, G., & Mehrotra, S. (2023, February). Comparative Analysis of Machine Learning Algorithms for Medical Insurance Cost Prediction. In International Conference On Innovative Computing and Communication (pp. 885-892). Singapore: Springer Nature Singapore.

[9] Sahare, A. N. (2023). Forecasting Medical Insurance Claim Cost with Data Mining Techniques (Doctoral dissertation, Dublin, National College of Ireland).

[10] Hassan, C. A., Iqbal, J., Hussain, S., AlSalman, H., Mosleh, M. A., & Sajid Ullah, S. (2021). A computational intelligence approach for predicting medical insurance cost. Mathematical Problems in Engineering, 2021, 1-13.

[11] Demirci, F., Emec, M., Gursoy Doruk, O., Ormen, M., Akan, P., & Hilal Ozcanhan, M. (2023). Prediction of LDL in hypertriglyceridemic subjects using an innovative ensemble machine learning technique. *Turkish Journal of Biochemistry*, (0).

[12] Kaya, Y., Yiner, Z., Kaya, M., & Kuncan, F. (2022). A new approach to COVID-19 detection from X-ray images using angle transformation with GoogleNet and LSTM. *Measurement Science and Technology*, *33*(12), 124011.

[13] Hemdan, E. E. D., El-Shafai, W., & Sayed, A. (2023). CR19: A framework for preliminary detection of COVID-19 in cough audio signals using machine learning algorithms for automated medical diagnosis applications. *Journal of Ambient Intelligence and Humanized Computing*, *14*(9), 11715-11727.

[14] AKDAĞ, S., Kuncan, F., & Kaya, Y. (2022). A new approach for classification of congestive heart failure and arrhythmia by downsampling local binary patterns with LSTM. *Turkish Journal of Electrical Engineering and Computer Sciences*, *30*(6), 2145-2164.

[15] Kaya, Y., & Kuncan, F. (2022). A hybrid model for classification of medical data set based on factor analysis and extreme learning machine: FA+ ELM. *Biomedical Signal Processing and Control*, *78*, 104023.

[16] Wu, X., Tang, H., Zhu, Z., Liu, L., Chen, G., & Yang, M. S. (2023). Nonlinear strict distance and similarity measures for intuitionistic fuzzy sets with applications to pattern classification and medical diagnosis. *Scientific reports*, *13*(1), 13918.

[17] Ayvaz, E., Kaplan, K., Kuncan, F., Ayvaz, E., & Türkoğlu, H. (2022). Reducing Operation Costs of Thyroid Nodules Using Machine Learning Algorithms with Thyroid Nodules Scoring Systems. *Applied Sciences*, *12*(22), 11559.

[18] Yurtsever, M., & Emeç, M. (2023). Potable Water Quality Prediction Using Artificial Intelligence and Machine Learning Algorithms for Better Sustainability. *Ege Academic Review*, *23*(2), 265-278.

[19] Orenc, S., Acar, E., & Özerdem, M. S. (2022, October). The Electricity Price Prediction of Victoria City Based on Various Regression Algorithms. In *2022 Global Energy Conference (GEC)* (pp. 164-167). IEEE.

[20] Gönenç, A., Acar, E., Demir, İ., & Yılmaz, M. (2022, October). Artificial Intelligence Based Regression Models for Prediction of Smart Grid Stability. In *2022 Global Energy Conference (GEC)* (pp. 374-378). IEEE.

[21] Ruzgar, S., & Acar, E. (2022). The statistical neural network-based regression approach for prediction of the optical band gap of CuO. *Indian Journal of Physics*, *96*(12), 3547-3557.

[22] Emeç, M., & Özcanhan, M. H. (2023). Veri Ön İşleme ve Öznitelik Mühendisliğinin Yapay Zekâ Yöntemlerine Uygulanması. MÜHENDİSLİKTE ÖNCÜ VE ÇAĞDAŞ ÇALIŞMALAR, 33-54.

[23] Emeç, M., & Özcanhan, M. H. (2023). Makine Öğrenmesi Algoritmalarında Hiper Parametre Belirleme. MÜHENDİSLİKTE ÖNCÜ VE ÇAĞDAŞ ÇALIŞMALAR, 71-98.

[24] Alzoubi, H. M., Sahawneh, N., AlHamad, A. Q., Malik, U., Majid, A., & Atta, A. (2022, October). Analysis Of Cost Prediction In Medical Insurance Using Modern Regression Models. In 2022 International Conference on Cyber Resilience (ICCR) (pp. 1-10). IEEE.

BIOGRAPHIES

Murat Emeç completed his computer engineering undergraduate degree at Ege University in 2015. In 2017, he received his master's in management information systems from Dokuz Eylül University. In 2022, he received his PhD in computer engineering from Dokuz Eylül University. Between 2010 and 2020, he worked as a senior software specialist in the Dokuz Eylül information technology (IT) department. Between 2021 and 2022, he worked at Marmara University as a lecturer and then as a lecturer doctor. He has been working at Istanbul University since September 2022. His research interests are the Internet of Things, Data Science, Artificial Intelligence, and Machine Learning.

European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.14, No.1, 2024

Research Article

A Communication System for Dynamic Leader Selection in Distributed UAV Swarm Architecture

Abdulmelik Bekmez[1] **Kadir Aram**[2*],

[1]Fatih Sultan Mehmet Vakıf University, Computer Engineering Department, 34445, Istanbul, Turkey. . (e-mail: abdulmelik.bekmez@gmail.com).
[2 *]Fatih Sultan Mehmet Vakıf University, Computer Engineering Department, 34445, Istanbul, Turkey (e-mail: karam@fsm.edu.tr).

ARTICLE INFO

Received: Jan., 18. 2024
Revised: Feb., 13, 2024
Accepted: Feb, 15, 2024

Keywords:
Formation Control
Leader-Follower
Robotics
Distributed
Communication

Corresponding author: *Kadir Aram*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1434751

ABSTRACT

Distributed swarm robot systems are made up of several robots that communicate with one another and often work together to complete a task or reach a predetermined objective. These systems frequently consist of many platforms, like unmanned aerial aircraft, mobile robots, or other types of vehicles.

This paper offers a comprehensive exploration of the design, modeling, and real-world hardware and software implementation of a distributed swarm system. The decision was made to employ standard Pixhawk hardware for the swarm agents. Pixhawk, a freely available hardware and software platform for autonomous flight control, is commonly utilized in autonomous cars, multirotor vehicles, drones, and various robotic applications. Operating autonomously from the ground control station, swarm agents dynamically identify leaders during operation and execute leader tracking navigation to model swarm behavior. Ensuring generality and dynamism in all protocols and communication was a primary focus during the research phase. To maintain this dynamism, each protocol and communication process is implemented in distinct threads on the computer, and synchronization is achieved through synchronization primitives, shared memory, and interthread communication.

1. INTRODUCTION

Robot swarms are important for automation systems in many areas, such as search and rescue operations, environmental monitoring, environmental cleaning, area surveillance, agricultural activities, and transportation of heavy loads [1-5] . Unmanned aerial vehicle (UAV) swarms provide a collaborative structure to perform complex operations with a single independent UAV.

Swarm robot systems are generally controlled through two main methods: centralized and distributed. In centralized systems, control originates from a single central point. On the other hand, in decentralized systems, each agent operates autonomously, making independent decisions and taking actions without relying on a central control point.

Various methods have been employed in the design of distributed UAV swarm control. These include a hybrid-flocking control algorithm, amalgamating vector field guidance, augmented Cucker-Smale model, and potential field techniques to attain path following, collective flocking behavior, and collision avoidance [6]. Another approach involves a mixed game theory utilizing a hierarchical learning algorithm for large-scale multi-agent systems, employing cooperative game, Stackelberg game, and mean field game for efficient coupling between leaders and followers [7]. Additionally, a distributed method, relying on monocular vision information, integrates a control model, target detection through a modified YOLOV3-tiny method, and orientation and distance estimation using geometric approaches [8]. Zhu and Deng proposed a distributed swarm control framework with limited interaction. In this framework, UAVs select limited interactive neighbors, combining interaction force and obstacle avoidance to ensure safety and effective guidance [9].

Various strategies are employed in the formation control of robot swarms, including leader-follower [10], virtual structure [11], and behavioral-based approaches [12].

The leader-follower strategy involves designating one agent as the leader, with the remaining agents following its movements. This study adopts the leader-follower strategy due to its capacity to minimize the number of connections. Numerous studies have explored variations of this strategy, employing different methodologies.

Zhang et al. introduced three strategies for formation reconfiguration, focusing on leader disengagement, follower detachment, and adding new members to minimize the need for frequent connection changes [13]. Restrepo and Loria proposed

two controllers for formation-tracking control of velocity-controlled unicycles in a leader-follower configuration, addressing scenarios with known and unknown leader velocities [14]. Lee et al. conducted a study on swarm control algorithms for unmanned surface vehicles (USVs), validating the effectiveness of a leader-follower swarm control method through sea area tests [15]. Pauli and Fichter proposed a leader-follower formation control algorithm for UAV swarms, ensuring precise lateral and vertical separation during turns and adeptly handling communication limitations among agents [16].

To use a leader-follower strategy in a distributed system, the leader robot must also be dynamically selected by the swarm agents. One approach utilizes behavior-based control and a repulsive force method to navigate maze-like environments, automatically designating a leader when robots are stranded [17]. Another strategy concentrates on choosing a minimal set of leaders, employing the sub modularity ratio and defining metrics based on consensus tracking criteria [18].

The study encompasses several key aspects, including the dynamic determination of a single leader by swarm agents, the communication protocols and techniques employed in this process, the implementation of these protocols, and the subsequent testing of the architecture. Upon reviewing the literature, it became evident that existing research on distributed swarm systems is predominantly theoretical and algorithmic. This study aims to bridge this gap by adapting algorithms for the communication of a distributed swarm system and the navigation of agents to real hardware. The second section delves into the hardware structure, communication protocol, and autopilot software. Following this, the communication architecture, dynamic leader selection process, and leader tracking strategy are elucidated. The third section showcases the interface and simulation environment of the study. The final section engages in a comprehensive discussion of the results.

2. MATERIALS AND METHODS

This study involves multiple stages, particularly in implementing the leader-follower strategy. In this strategy, the robots are required to autonomously designate a leader and then organize themselves into a formation around this leader, following them to the specified location. The consecutive steps for these processes are detailed below.

2.1. Materials

The study is designed to be implemented on real hardware, and for this purpose, Pixhawk has been chosen as the hardware for autonomous flight control. Pixhawk, an open-source platform for autonomous flight control hardware and software, includes a control board that seamlessly integrates with diverse sensors and actuators, providing extensive capabilities for flight control and automation [19]. Widely employed in various robotic applications like multirotor vehicles, drones, and autonomous vehicles, Pixhawk serves as a standard control platform. Several autopilot software, including Ardupilot and PX4, have been developed based on this standard.

Ardupilot and PX4 are the most popular open-source autopilot systems. They are designed to operate a wide range of autonomous vehicles, including submarines, rotary-wing platforms, and fixed-wing aircraft [20].

PX4 is an embedded robotics middleware and programming environment with a multithreaded, publish-subscribe design pattern. It offers a software interface for microcontroller applications. The PX4 autopilot platform can operate independently on Pixhawk standard hardware or be coupled with a companion computer for tasks demanding additional processing power or an external GPU [21]. The PX4 autopilot software was used because it is an open-source platform, uses an open-source hardware standard, and is well integrated with widely used software such as ROS2.

ROS is an open-source software environment tailored for both commercial and research-based robotic applications [22]. It's preferred for its versatility, enabling development in various languages, fostering communication between processes through a publisher-subscriber architecture, supporting project-specific message types, allowing packaging and sharing of applications, and providing extensive package support for common robotic challenges. ROS features tools like RVIZ for data visualization [23]. In ROS, processes are termed nodes, and they communicate via messages, with publisher nodes sending messages to subscriber nodes through topics [24].

ROS2 was created to address the limitations of ROS, which include a centralized network structure, a lack of network security mechanisms, and dependence on third-party software for integration into microprocessors [25]. ROS2 adopts a distributed network system, employing the network layer's Data Distribution Service (DDS). Additionally, for communication between drones and between drones and ground stations, the Zenoh protocol was implemented.

Zenoh, a Pub/Sub-Query protocol, intricately unifies computations, data in motion, and data at rest. It was purposefully designed to cater to the demands of the shift from micro-controllers to the cloud, offering a seamless integration of diverse network topologies and technologies. This enables Zenoh to deliver messages with heightened speed and efficiency, fulfilling the need for high-density bandwidth while keeping latency to a minimum. As a result, it has evolved into an essential component of the Cloud-to-Edge Continuum [26].

Zenoh facilitates communication between two processing units on the same computer and extends its functionality to units on different computers within the same local network. This functionality resembles the DDS network software found in ROS2, sharing the dynamic discovery feature with DDS. Notably, Zenoh exhibits greater efficiency than DDS in wireless networks[27]. In the upcoming swarm system design, where swarm agents lack specific IP addresses and need to communicate via RF communication, the Zenoh protocol will be employed. This choice is driven by Zenoh's capability to operate in systems without IP addresses. Consequently, swarm agents will communicate through the serial interface provided by telemetry modules, utilizing the Zenoh protocol.

Gazebo is a 3D simulation platform that enables the development and testing of robots to be operated in indoor and outdoor environments [28]. It is frequently used in robotics research projects, competitions and commercial applications. PX4 autopilot software supports the simulation of rotary-wing platforms, fixed-wing, and VTOL robots with the Software In The Loop technique in development processes using the Gazebo environment. The Gazebo Garden version was used during the study.

The UAV used for this research is based on a Holybro X500 quadrotor equipped with a Pixhawk flight controller. The appearance of the robot in the simulation environment is given in the Figure 1.

Figure 1. X500 in Gazebo Simulation

2.2. Communication Architecture

A crucial aspect of a distributed swarm architecture is the communication mechanism among swarm agents and other system elements. To establish a distributed framework, each agent must support peer-to-peer communication. Defining the communication protocols and message formats for these interactions is imperative. For agent-to-agent communication, the "swarm/<agent_name>" topical format is utilized when one agent sends a message to another. For instance, if agent1 intends to convey a message to agent2, it sends a P2P message to the "swarm/agent2" topical. This P2P message comprises the sender's name, the message type, and the P2PType data, encapsulating message-specific information. A P2P message includes 4 types of messages. These are:

• Heartbeat
• HeartbeatAck
• Pooling
• Selection

HeartBeat Message: A dynamically generated swarm leader sends the Heartbeat message to the swarm members. This message indicates that the swarm is connected and allows the swarm leader to assign tasks to the swarm members.

HeartbeatAck Message: The HeartbeatAck message indicates the message the swarm member sends to the leader. The swarm member sends the swarm leader position and orientation data about itself in this message. This message also determines the status of the swarm member's connection to the leader.

Pooling Message: Pooling message, a type employed for dynamic swarm leader selection, comprises four subtypes. These are:

- StartPool
- StartPoolAck
- SyncPool
- SyncPoolAck

StartPool Message One agent sends a StartPool message to another agent to initiate the pooling protocol to create a swarm.

StartPoolAck Message The "StartPoolAck" message serves as the response to an incoming "StartPool" message. Depending on the agent's present state, it dispatches a "Success" message if it is available and an "AlreadyInPool" message if it is already part of a pool. The "AlreadyInPool" message contains details about the current pool state.

SyncPool Message The "SyncPool" message is sent to agents during the pool creation process to synchronize the instantaneous pool with other potential swarm agents at a specific frequency. This message contains the pool data maintained by the agent.

SyncPoolAck Message SyncPoolAck message is sent by the agent receiving the SyncPool message to the sending agent as a reply. If the pools of the agents are the same, the 'Same' message is sent; if they are different, a 'Different' message is sent. The 'Different' message also holds the names of the elements in the pool. It also contains the status of whether the agent's pool is locked or not.

Selection Message: The Selection message dynamically specifies the type of message that each agent shares its selection with the other agents in the pool with a specific frequency in the selection part, which is the last part of the swarm leader selection process.

2.3. Dynamic Leader Selection

In order to have a distributed architecture, a swarm system should not be utterly dependent on any central authority. However, swarm systems with a fully distributed architecture are subject to communication constraints. A simple calculation of the number of connections is shown in equation 1.

$$c = \frac{(n-1)\times n}{2} \qquad (1)$$

Here: c is connection count; n is agent count.

For example, if a swarm system with ten agents wants to create a communication topology between each agent, it must create 45 peer-to-peer channels. For these reasons, the swarm system is designed to determine the swarm's leader and maintain this connection dynamically.

The process of dynamically selecting a leader consists of the following sub-parts.

- Selection pool initialization phase
- Expanding the selection pool
- Phase to lock the selection pool
- Election phase
- Leader selection

Selection pool initialization: While in this state, the agent broadcasts a self-introduction message via the "swarm/advertize" topical. Through this topic, the agent signals its readiness for potential connections and swarming, accompanied by sending its name to facilitate establishing a connection. Simultaneously, all available agents monitor this topic, including the agent who transmitted this message. Another agent receiving this message sends a "StartPool"

message to the sending agent. Then, it waits for the "StartPoolAck" message to arrive. The agent receiving the "StartPool" message, if it is not already in a pool, sends a Success message to start the pool and gives feedback to the other agent. However, if it is already in a pool and this pool is still accepting members, it sends an "AlreadyInPool" message and adds the agent that sent the message to the group. If the recipient of the "StartPool" message is not in any of these situations, it ignores the incoming message and does not send any response.

Expanding the selection pool: After the pool is initialized for dynamic leader election, the agents perform two types of operations.

- Send a pooling request to agents sending 'Advertize' messages

- Ensuring pool synchronization between agents in the pool

Agents that are expanding the pool also keep the last time it was updated. This time, data is updated if a member is added or removed from the pool, preventing the pool from being created and closed immediately. If a new member is added to the pool, the time data is updated, and as a result of synchronization messages, other agents also update their pools and the time data they keep.

Lock Selection Pool: Once the selection pool stabilizes for two seconds, it becomes locked for all agents. Agents that have successfully locked the pool continue transmitting SyncPool messages to others in the pool at a specific frequency. If there are agents whose pools are still unlocked, the pool locking status is indicated with the SyncPoolAck message. This ensures that if, for any reason, there are agents whose pools remain unlocked, the synchronization of pool locking among all agents is guaranteed. In the case where received pools differ, both agents merge their pools. Following the pool merging process, other agents update their pools through synchronous messages sent to one another. Similar to expanding the selection pool, the last update time is retained in the case of locking. This time, data is updated during an asynchronous state, and if there is no update for two seconds, the pool selection process is initiated.

Election Phase: The voting process can be tailored to the specific swarm architecture or problem at hand. In the devised architecture, each agent autonomously casts a vote for itself as the potential swarm leader, with each vote assigned a weight and randomized during implementation. Agents within the pool exchange their votes at a designated frequency. Upon receiving a vote message, the recipient agent scrutinizes the incoming vote. If it is still locked within the pool, the agent promptly updates its status to "Selection" and initiates the voting process. If the incoming vote weight surpasses its own, the agent updates its vote with the incoming one and the corresponding time variable.

Conversely, if the incoming vote carries less weight, the agent maintains its current vote. This mechanism ensures unanimity and synchronization among the agents. If, after 2 seconds, the voting pool converges on a consistent vote, the leader selection process within the pool is triggered, similar to other scenarios.

Leader Selection: Once a unanimous vote is secured in the election pool, the leader initiates a heartbeat message transmission to its swarm at a specified frequency. This transmission serves the dual purpose of confirming the leader's continued activity and assigning tasks to the swarm members. Concurrently, the leader awaits the reception of "HeartBeatAck" messages, indicating the ongoing activity of swarm members and facilitating the exchange of necessary data, such as Pose messages. If a "HeartBeatAck" message isn't received from a swarm member within a defined timeframe, the leader expels that agent from the swarm and discontinues the transmission of heartbeat messages to them. Conversely, if a swarm member fails to receive a heartbeat message from the leader within a specified period, it refrains from sending a "HeartbeatAck" message. Even if a heartbeat message is received subsequently, the swarm member departs from the swarm, opting to join another or create a new one. The flow chart of the leader selection process is shown in Figure 2.

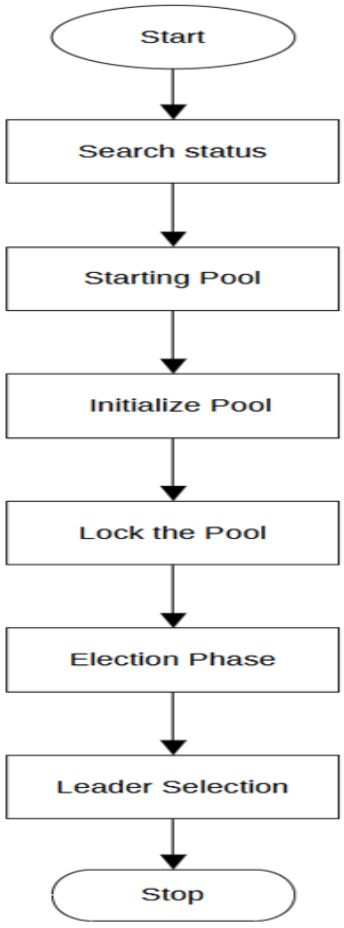


Figure 2. Leader Selection Flow Cart

2.4. *Formation Process*

The strategy involves the leader robot moving to a predefined target point while the other robots follow it. Continuous communication is essential between the leader and the followers. The leader robot continuously broadcasts its orientation and position, which is then transmitted to the follower robots. The followers, utilizing the leader's position, maintain a predetermined formation by following the leader's movements. The leader's current location serves as the goal for the follower robots, and they align their orientation with that of the leader. Throughout the tracking process, a constant distance between the leader and the followers is maintained.

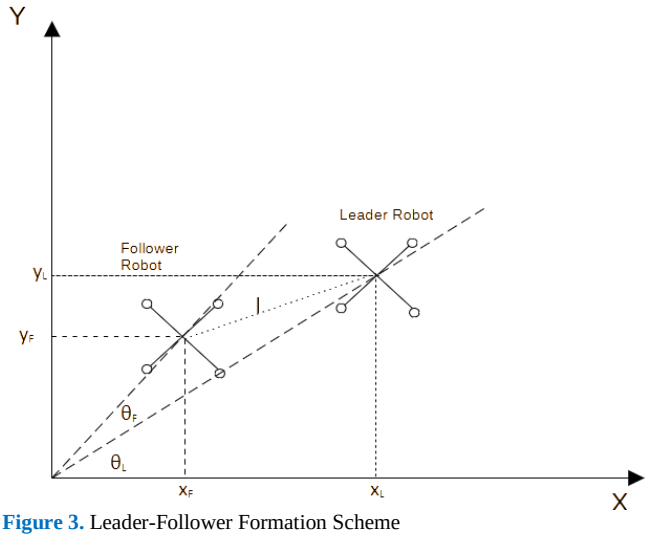


Figure 3. Leader-Follower Formation Scheme

In Figure:
X_L, Y_L : leader's position,
θ_L :is the leader's orientation
X_F, Y_F: follower robot's position
θ_F :follower robot's orientation.
l :the distance between the leader and the follower.
Equation 2 shows the distance between the leader robot and the follower robot.

$$dist\ (l) = \sqrt{(X_L - X_F)^2 + (Y_L - Y_F)^2} \qquad (2)$$

The PX4 autopilot software employs a position controller for guiding swarm agents from point A to point B. Using the Kalman Filter algorithm, a position estimation method, the position controller establishes a local coordinate plane, referencing the global position at the start time. Upon receiving a follow command from the leader through the interface, the leader communicates with each swarm agent, specifying the agent to be followed, the following direction, and the general heading angle for leader-follow navigation. Swarm members utilize peer-to-peer communication to subscribe to the Pose data of the designated agent. After receiving the agent's name, they follow the leader's command, creating a 10-meter vector in the opposite direction of the swarm's head rotation angle and a 5-meter direction vector based on the specified direction for the agent to stay. Position determination is accomplished by adding these vectors to the position of the agent being followed.

3. EXPERIMENTS

Swarm ground control station software has been developed for efficient swarm management, allowing dynamic connections and individual control of each agent. Agents introduce themselves to the ground control station via the swarm/GSAdvertize topical. The ground control station software consistently monitors this topic, sending connection requests to agents that identify themselves and initiate the connection. For the connection to be maintained, the agent must send a Heartbeat command with a specific frequency, indicating that the connection is active; otherwise, it disconnects from the ground control station software. To facilitate location tracking and control of swarm agents across the global map, a map interface was developed using OpenStreetMap data. This interface displays the locations of agents connected to the ground control station. The general map view of the ground control station is shown in Figure 4.

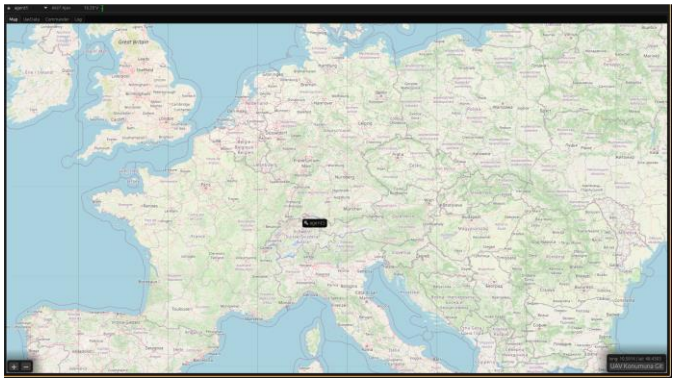


Figure 4. General Map View

A close-up map view of the ground control station is shown in Figure 5.

Figure 5. Close-up Map View

The interface enables the selection of the agent to be controlled, displaying instant telemetry data and the control panel of the selected agent. Additionally, individual swarm agents can be moved from one point to another using position controllers.
The user interface for position control of the swarm agents is shown in Figure 6.

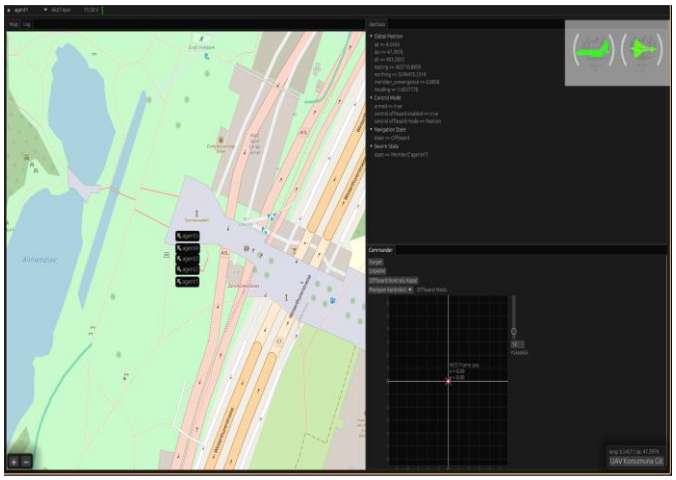


Figure 6. Position Controller User Interface

The position of the swarm robots in the Gazebo simulation environment is shown in Figure 7.

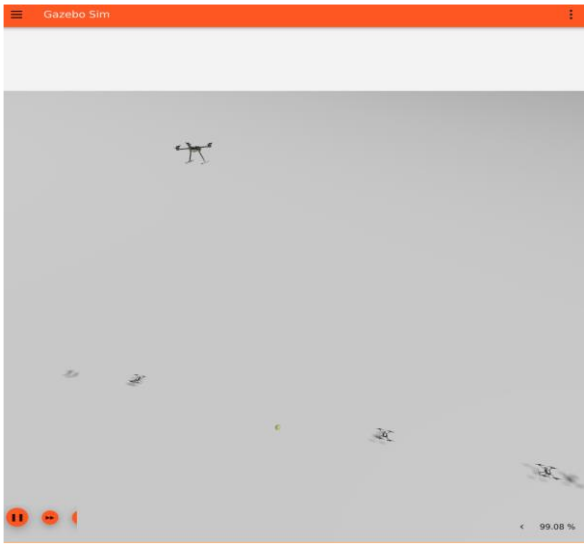


Figure 7. Swarm Robots in Gazebo Simulation

With the selection of the leader agent, the swarm system can be commanded to enter the leader-following state. Figure 8 shows the follower robots forming a v-formation behind the leader robot.

Figure 8. Swarm Robots in V-Formation

The view of the leader following the process in the Gazebo simulation environment is shown in Figure 9.

4. CONCLUSION

This study delves into the intricate realm of swarm systems with distributed architectures, acknowledging the inherent complexity compared to centralized architectures. The focal point is the analysis of formation control within a distributed architecture designed to operate on tangible hardware.

Utilizing PixHawk, an open-source autonomous flight control hardware and software platform, PX4 as open-source autopilot software, and employing ROS2 for implementation, the study takes a practical approach. Gazebo serves as the simulation environment for testing. Because swarm agents lack IP addresses in this system, communication relies on the Zenoh protocol. A standardized communication process is established for dynamic leader selection, and specific message formats are crafted for agents to execute these operations. Each robot is assigned a randomly determined vote weight during the leader election, and the leader is determined based on accumulated votes. The simulation demonstrates the leader's announcement and subsequent command for followership, adopting a V-formation for navigation. Simulation outputs affirm the successful realization of the study. Future endeavors involve extending the study to explore alternative leader selection methods and accommodating multiple leaders.

Figure 9. Leader Follower in Gazebo Simulation

REFERENCES

[1] W. Luo, S. S. Khatib, S. Nagavalli, N. Chakraborty, and K. Sycara, "Asynchronous distributed information leader selection in robotic swarms," in *2015 IEEE International Conference on Automation Science and Engineering (CASE)*, 2015, pp. 606–611. doi: 10.1109/CoASE.2015.7294145.

[2] M. Carpentiero, L. Gugliermetti, M. Sabatini, and G. B. Palmerini, "A swarm of wheeled and aerial robots for environmental monitoring," in *2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)*, 2017, pp. 90–95. doi: 10.1109/ICNSC.2017.8000073.

[3] M. Duarte *et al.*, "Application of swarm robotics systems to marine environmental monitoring," in *OCEANS 2016 - Shanghai*, 2016, pp. 1–8. doi: 10.1109/OCEANSAP.2016.7485429.

[4] J. Asbach, S. Chowdhury, and K. Lewis, "Using an Intelligent UAV Swarm in Natural Disaster Environments," in *International Design Engineering Technical Conferences and Computers and Information*

in Engineering Conference, 2018, p. V02AT03A013.

[5] J. Scherer *et al.*, "An autonomous multi-UAV system for search and rescue," in *Proceedings of the first workshop on micro aerial vehicle networks, systems, and applications for civilian use*, 2015, pp. 33–38.

[6] Y. Song *et al.*, "Distributed swarm system with hybrid-flocking control for small fixed-wing UAVs: Algorithms and flight experiments," *Expert Syst. Appl.*, vol. 229, p. 120457, 2023.

[7] S. Dey and H. Xu, "Intelligent Distributed Swarm Control for Large-Scale Multi-UAV Systems: A Hierarchical Learning Approach," *Electronics*, vol. 12, no. 1, p. 89, 2022.

[8] Y. Jia, M. Chen, Y. Gao, and H. Wang, "A Distributed Method to Form UAV Swarm based on Moncular Vision," in *2022 IEEE 28th International Conference on Parallel and Distributed Systems (ICPADS)*, 2023, pp. 41–48.

[9] B. Zhu and Y. Deng, "Distributed UAV swarm control framework with limited interaction for obstacle avoidance," *Aircr. Eng. Aerosp. Technol.*, no. ahead-of-print, 2022.

[10] R. Rafifandi, D. L. Asri, E. Ekawati, and E. M. Budi, "Leader--follower formation control of two quadrotor UAVs," *SN Appl. Sci.*, vol. 1, pp. 1–12, 2019.

[11] N. H. M. Li and H. H. T. Liu, "Formation UAV flight control using virtual structure and motion synchronization," in *2008 American Control Conference*, 2008, pp. 1782–1787.

[12] D. Xu, X. Zhang, Z. Zhu, C. Chen, P. Yang, and others, "Behavior-based formation control of swarm robots," *Math. Probl. Eng.*, vol. 2014, 2014.

[13] H. Zhang, G. Zhang, R. Yang, Z. Feng, and W. He, "Resilient Formation Reconfiguration for Leader--Follower Multi-UAVs," *Appl. Sci.*, vol. 13, no. 13, p. 7385, 2023.

[14] A. Lazri, E. Restrepo, and A. Lor\'\ia, "Robust leader-follower formation control of autonomous vehicles with unknown leader velocities," in *2023 European Control Conference (ECC)*, 2023, pp. 1–6.

[15] J.-H. Lee *et al.*, "Unmanned Surface Vehicle Using a Leader--Follower Swarm Control Algorithm," *Appl. Sci.*, vol. 13, no. 5, p. 3120, 2023.

[16] N. Pauli and W. Fichter, "Leader-Follower Formation Control with Longitudinal Separation along Lateral and Vertical Shifted Follower Paths," in *AIAA SCITECH 2023 Forum*, 2023, p. 484.

[17] L.-B. Wee and Y.-C. Paw, "Simultaneous Mapping Localization and Path Planning for UAV Swarm," in *2023 IEEE Aerospace Conference*, 2023, pp. 1–6.

[18] H. Y. Liu, J. Chen, K. H. Huang, G. Q. Cheng, and R. Wang, "UAV swarm collaborative coverage control using GV division and planning algorithm," *Aeronaut. J.*, vol. 127, no. 1309, pp. 446–465, 2023.

[19] L. Meier, P. Tanskanen, F. Fraundorfer, and M. Pollefeys, "Pixhawk: A system for autonomous flight using onboard computer vision," in *2011 ieee international conference on robotics and automation*, 2011, pp. 2992–2997.

[20] A. Allouch, O. Cheikhrouhou, A. Koubâa, M. Khalgui, and T. Abbes, "MAVSec: Securing the MAVLink Protocol for Ardupilot/PX4 Unmanned Aerial Systems," in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, 2019, pp. 621–628. doi: 10.1109/IWCMC.2019.8766667.

[21] L. Meier, D. Honegger, and M. Pollefeys, "PX4: A node-based multithreaded open source robotics framework for deeply embedded platforms," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 6235–6240. doi: 10.1109/ICRA.2015.7140074.

[22] A. BEKMEZ and A. Kadir, "Three Dimensional Formation Control of Unmanned Aerial Vehicles in Obstacle Environments," *Balk. J. Electr. Comput. Eng.*, vol. 11, no. 4, pp. 387–394, 2023.

[23] M. Quigley *et al.*, "ROS: an open-source Robot Operating System," in *ICRA workshop on open source software*, 2009, p. 5.

[24] L. Joseph and J. Cacace, *Mastering ROS for Robotics Programming: Design, build, and simulate complex robots using the Robot Operating System*. Packt Publishing Ltd, 2018.

[25] S. Macenski, T. Foote, B. Gerkey, C. Lalancette, and W. Woodall, "Robot Operating System 2: Design, architecture, and uses in the wild," *Sci. Robot.*, vol. 7, no. 66, p. eabm6074, 2022.

[26] J. J. Lopez Escobar, R. P. Diaz-Redondo, and F. Gil-Castineira, "Unleashing the power of decentralized serverless IoT dataflow architecture for the Cloud-to-Edge Continuum: a performance comparison," *Ann. Telecommun.*, pp. 1–14, 2024.

[27] A. Corsaro *et al.*, "Zenoh: Unifying Communication, Storage and Computation from the Cloud to the Microcontroller," vol. DSD 2023, 2023.

[28] Z. Tüfekçi and G. Erdemir, "Experimental Comparison of Global Planners for Trajectory Planning of Mobile Robots in an Unknown Environment with Dynamic Obstacles," in *2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2023, pp. 1–6.

BIOGRAPHIES

Abdulmelik Bekmez is currently studying in his bachelor's degree in the Department of Computer Engineering, Engineering Faculty, at the Fatih Sultan Mehmet Vakıf University. His current research areas include robotics and formation control.

Kadir Aram obtained his BSc and MSc degree in computer and control education Marmara University. He completed his PH.D. in Computer Science and Engineering at Istanbul Sabahattin Zaim University in 2023. He is research assistant in Computer Engineering department at Fatih Sultan Mehmet Vakıf University. His research interest are mobile robotics and natural language processing.