



APJESS

Journal of Engineering
and **Smart Systems**

Volume : 11

Issue : 3

Year : 2023

Volume 11 / Issue 3

Academic Platform Journal of Engineering and Smart Systems

Editor in Chief (Owned By Academic Perspective)

Dr. Mehmet SARIBIYIK, Sakarya University of Applied Sciences, Türkiye

Editors

Dr. Caner ERDEN, Sakarya University of Applied Sciences, Türkiye

Dr. John YOO, Bradley University, USA

Editorial Board

Dr. Abdullah Hulusi KÖKÇAM, Sakarya University, Türkiye

Dr. Ali Tahir KARAŞAHİN, Karabuk University, Türkiye

Dr. Aydın MÜHÜRÇÜ, Kırklareli University, Türkiye

Dr. Ayşe Nur AY, Sakarya University of Applied Sciences, Türkiye

Dr. Cengiz KAHRAMAN, Istanbul Technical University, Türkiye

Dr. Elif Elçin GÜNAY, Sakarya University, Türkiye

Dr. Erkan ÇELİK, Istanbul University, Türkiye

Dr. Fatih VARÇIN, Sakarya University of Applied Sciences, Türkiye

Dr. Gürcan YILDIRIM, Abant İzzet Baysal University, Türkiye

Dr. Hacı Mehmet ALAKAŞ, Kirikkale University, Türkiye

Dr. Huseyin SEKER, Birmingham City University, Birmingham, United Kingdom

Dr. Kerem KÜÇÜK, Kocaeli University, Türkiye

Dr. Marco A. ACEVES-FERNÁNDEZ, Universidad Autónoma De Querétla, Mexico

Dr. Mazin MOHAMMED, University Of Anbar, Iraq

Dr. Mehmet Emin AYDIN, University of The West Of England, United Kingdom

Dr. Muhammet KURULAY, Yıldız Technical University, Türkiye

Dr. Muhammed Maruf ÖZTÜRK, Suleyman Demirel University, Türkiye

Dr. Ömer AYDIN, Celal Bayar University, Türkiye

Dr. Rakesh PHANDEN, Amity University Uttar Pradesh, India

Dr. Uğur Erkin KOCAMAZ, Bursa Uludağ University, Türkiye

Dr. Tuğba TUNACAN, Abant İzzet Baysal University, Türkiye

Dr. Turgay Tugay BİLGİN, Bursa Technical University, Türkiye

Dr. Tülay YILDIRIM, Yıldız Technical University, Türkiye

Dr. Valentina E. BALAS, Aurel Vlaicu University of Arad, Romania

Language Editor

Dr. Hakan ASLAN, Sakarya University, Türkiye

Editorial Assistants

Selim İLHAN, Sakarya University, Türkiye

İbrahim MUCUK, Sakarya University, Türkiye

Correspondence Address

Academic Platform Journal of Engineering and Smart Systems

Akademik Perspektif Derneği, Tığcılar Mahallesi Kadir Sokak No:12

Kat:1 Adapazarı SAKARYA

+90 551 628 9477 (WhatsApp only)

<https://dergipark.org.tr/tr/pub/apjess>

Issue Link: <https://dergipark.org.tr/tr/pub/apjess/issue/80086>

Aim and Scope

Academic Platform Journal of Engineering and Smart Systems (APJESS) is a peer reviewed open-access journal which focuses on the research and applications related to smart systems and artificial intelligence. APJESS accepts both **original research papers** and **review articles** written in **English**. It is essential that the information created in scientific study needs to be new, suggest new method or give a new dimension to an existing information. Articles submitted for publication are evaluated by at least two referees in case the editor finds potential scientific merit, and final acceptance and rejection decision are taken by editorial board. The authors are not informed about the name of referees who evaluate the papers. In similar way, the referees are not allowed to see the names of authors. The papers which do not satisfy the scientific level of the journal can be refused with unexplained reason.

There are two key principles that APJESS was founded on: Firstly, to publish the most exciting, novel, technically sound, and clearly presented researches with respect to the subjects of smart systems and artificial intelligence. Secondly, to provide a rapid turn-around time possible for reviewing and publishing, and to disseminate the articles freely for research, teaching and reference purposes.

Any information about a submitted manuscript cannot be disclosed by the editor and any other editorial staff to anyone other than the corresponding author, reviewers, potential reviewers, other editorial advisers, and the publisher. No confidential information or ideas obtained through peer review can be used for personal advantage.

Journal History

The journal was published between 2013-2021 with the title of "Academic Platform - Journal of Engineering and Science". It will be published under its new title "Academic Platform Journal of Engineering and Smart Systems" after 2022.

Former Title: Academic Platform - Journal of Engineering and Science

Years: 2013-2021

Scope

APJESS aims to publish research and review papers dealing with, but not limited to, the following research fields:

- Knowledge Representation and Reasoning,
- Data Mining & Data Science,
- Supervised, Semi-Supervised and Unsupervised Learning,
- Machine Learning (ML) and Neural Computing,
- Evolutionary Computation,
- Natural Language Processing, Internet of Things, Big Data
- Fuzzy Systems,
- Intelligent Information Processing,
- AI Powered Robotic Systems,
- Multi-agent Systems and Programming for Smart Systems

Author Guidelines

Article Types

Manuscripts submitted to APJESS should neither be published previously nor be under consideration for publication in another journal.

The main article types are as follows:

Research Articles: Original research manuscripts. The journal considers all original research manuscripts provided that the work reports scientifically sound experiments and provides a substantial amount of new information.

Review Articles: These provide concise and precise updates on the latest progress made in a given area of research.

Checklist for Submissions

Please,

- read the [Aims & Scope](#) to see if your manuscript is suitable for the journal,
- use the [Microsoft Word template](#) to prepare your manuscript;
- Download [Copyright Transfer Form](#) and signed by all authors.
- make sure that issues about [Ethical Principles and Publication Policy](#), [Copyright and Licensing](#), [Archiving Policy](#), [Repository Policy](#) have been appropriately considered;
- Ensure that all authors have approved the content of the submitted manuscript.

The main text should be formed in the following order:

Manuscript: The article should start with an introduction written in scientific language, putting thoughts together from diverse disciplines combining evidence-based knowledge and logical arguments, conveying views about the aim and purpose of the article. It must address all readers in general. The technical terms, symbols, abbreviations must be defined at the first time when they are used in the article. The manuscript should be formed in the following order:

Introduction,

Material and Method,

Findings,

Discussion and Conclusion.

References: At the end of the paper provide full details of all references cited in-text. The reference list should be arranged in the order of appearance of the in-text citations, not in an alphabetical order, beginning with [1], and continuing in an ascending numerical order, from the lowest number to the highest. In the reference list, only one resource per reference number is acceptable.

References must be numbered in order of appearance in the text (including citations in tables and legends) and listed individually at the end of the manuscript. We recommend preparing the references with a bibliography software package, such as EndNote, Reference Manager or Zotero to avoid typing mistakes and duplicated references. Include the digital object identifier (DOI) for all references where available. Please use IEEE style.

IEEE Sample Reference List

[1] R. E. Ziemer and W. H. Tranter, Principles of Communications: Systems, Modulation, and Noise, 7th ed. Hoboken, NJ: Wiley, 2015.

[2] J. D. Bellamy et al., Computer Telephony Integration, New York: Wiley, 2010.

- [3] C. Jacks, High Rupturing Capacity (HRC) Fuses, New York: Penguin Random House, 2013, pp. 175–225.
- [4] N. B. Vargafik, J. A. Wiebelt, and J. F. Malloy, "Radiative transfer," in Convective Heat. Melbourne: Engineering Education Australia, 2011, ch. 9, pp. 379–398.
- [5] H. C. Hottel and R. Siegel, "Film condensation," in Handbook of Heat Transfer, 2nd ed. W. C. McAdams, Ed. New York: McGraw-Hill, 2011, ch. 9, pp. 78–99.
- [6] H. H. Gaynor, Leading and Managing Engineering and Technology, Book 2: Developing Managers and Leaders. IEEE-USA, 2011. Accessed on: Oct. 15, 2016. [Online]. Available: <http://www.ieeeusa.org/communications/ebooks/files/sep14/n2n802/Leading-and-Managing-Engineering-and-Technology-Book-2.pdf>
- [7] G. H. Gaynor, "Dealing with the manager leader dichotomy," in Leading and Managing Engineering and Technology, Book 2, Developing Leaders and Mangers. IEEE-USA, 2011, pp. 27–28. Accessed on: Jan. 23, 2017. [Online]. Available: <http://www.ieeeusa.org/communications/ebooks/files/sep14/n2n802/Leading-and-Managing-Engineering-and-Technology-Book-2.pdf>
- [8] M. Cvijetic, "Optical transport system engineering," in Wiley Encyclopedia of Telecommunications, vol. 4, J. G. Proakis, Ed. New York: John Wiley & Sons, 2003, pp. 1840–1849. Accessed on: Feb. 5, 2017. [Online]. Available: <http://ebscohost.com>
- [9] T. Kaczorek, "Minimum energy control of fractional positive electrical circuits", Archives of Electrical Engineering, vol. 65, no. 2, pp.191–201, 2016.
- [10] P. Harsha and M. Dahleh, "Optimal management and sizing of energy storage under dynamic pricing for the efficient integration of renewable energy", IEEE Trans. Power Sys., vol. 30, no. 3, pp. 1164–1181, May 2015.
- [11] A. Vaskuri, H. Baumgartner, P. Kärhä, G. Andor, and E. Ikonen, "Modeling the spectral shape of InGaAlP-based red light-emitting diodes," Journal of Applied Physics, vol. 118, no. 20, pp. 203103–203103-7, Jul. 2015. Accessed on: Feb. 9, 2017. [Online]. Available: doi: 10.1063/1.4936322
- [12] K. J. Krishnan, "Implementation of renewable energy to reduce carbon consumption and fuel cell as a back-up power for national broadband network (NBN) in Australia," Ph.D dissertation, College of Eng. and Sc., Victoria Univ., Melbourne, 2013.
- [13] C. R. Ozansoy, "Design and implementation of a Universal Communications Processor for substation integration, automation and protection," Ph.D. dissertation, College of Eng. and Sc., Victoria Univ., Melbourne, 2006. [Online]. Accessed on: June 22, 2017. [Online]. Available: <http://vuir.vu.edu.au/527/>
- [14] M. T. Long, "On the statistical correlation between the heave, pitch and roll motion of road transport vehicles," Research Master thesis, College of Eng. and Sc., Victoria Univ., Melb., Vic., 2016.
- [15] Safe Working on or Near Low-voltage Electrical Installations and Equipment, AS/NZS 4836:2011, 2011.

Ethical Principles and Publication Policy

Peer Review Policy

Academic Platform Journal of Engineering and Smart Systems (APJESS) applies double blind peer-review process in which both the reviewer and the author are anonymous. Reviewer selection for each submitted article is up to area editors, and reviewers are selected based on the reviewer's expertise, competence, and previous experience in reviewing papers for APJES.

Every submitted article is evaluated by area editor, at least, for an initial review. If the paper reaches minimum quality criteria, fulfills the aims, scope and policies of APJES, it is sent to at least two reviewers for evaluation.

The reviewers evaluate the paper according to the Review guidelines set by editorial board members and return it to the area editor, who conveys the reviewers' anonymous comments back to the author. Anonymity is strictly maintained.

The double-blind peer-review process is managed using “ULAKBİM Dergi Sistemleri”, namely Dergipark platform.

Open Access Policy

APJESS provides immediate open access for all users to its content on the principle that making research freely available to the public, supporting a greater global exchange of knowledge.

Archiving Policy

APJESS is accessed by Dergipark platform which utilizes the LOCKSS system to create a distributed archiving system among participating libraries and permits those libraries to create permanent archives of the journal for purposes of preservation and restoration.

Originality and Plagiarism Policy

Authors by submitting their manuscript to APJESS declare that their work is original and authored by them; has not been previously published nor submitted for evaluation; original ideas, data, findings and materials taken from other sources (including their own) are properly documented and cited; their work does not violate any rights of others, including privacy rights and intellectual property rights; provided data is their own data, true and not manipulated. Plagiarism in whole or in part without proper citation is not tolerated by APJESS. Manuscripts submitted to the journal will be checked for originality using anti-plagiarism software.

Journal Ethics and Malpractice Statement

For all parties involved in the publishing process (the author(s), the journal editor(s), the peer reviewers, the society, and the publisher) it is necessary to agree upon standards of expected ethical behavior. The ethics statements for APJESS are based on the Committee on Publication Ethics (COPE) Code of Conduct guidelines available at www.publicationethics.org.

1. Editor Responsibilities

Publication Decisions & Accountability

The editor of APJESS is responsible for deciding which articles submitted to the journal should be published, and, moreover, is accountable for everything published in the journal. In making these decisions, the editor may be guided by the journal's editorial board and/or area editors, and considers the policies of the journal. The editor should maintain the integrity of the academic record, preclude business needs from compromising intellectual and ethical standards, and always be willing to publish corrections, clarifications, retractions, and apologies when needed.

Fair play

The editor should evaluate manuscripts for their intellectual content without regard to race, gender, sexual orientation, religious belief, ethnic origin, citizenship, or political philosophy of the author(s).

Confidentiality

The editor and any editorial staff must not disclose any information about a submitted manuscript to anyone other than the corresponding author, reviewers, potential reviewers, other editorial advisers, and the publisher, as appropriate.

Disclosure, conflicts of interest, and other issues

The editor will be guided by COPE's Guidelines for Retracting Articles when considering retracting, issuing expressions of concern about, and issuing corrections pertaining to articles that have been published in APJES.

Unpublished materials disclosed in a submitted manuscript must not be used in an editor's own research without the explicit written consent of the author(s). Privileged information or ideas obtained through peer review must be kept confidential and not used for personal advantage.

The editor should seek so ensure a fair and appropriate peer-review process. The editor should recuse himself/herself from handling manuscripts (i.e. should ask a co-editor, associate editor, or other member of the editorial board instead to review and consider) in which they have conflicts of interest resulting from competitive, collaborative, or other relationships or connections with any of the authors, companies, or (possibly) institutions connected to the papers. The editor should require all contributors to disclose relevant competing interests and publish corrections if competing interests are revealed after publication. If needed, other appropriate action should be taken, such as the publication of a retraction or expression of concern.

2. Reviewer Responsibilities

Contribution to editorial decisions

Peer review assists the editor in making editorial decisions and, through the editorial communication with the author, may also assist the author in improving the manuscript.

Promptness

Any invited referee who feels unqualified to review the research reported in a manuscript or knows that its timely review will be impossible should immediately notify the editor so that alternative reviewers can be contacted.

Confidentiality

Any manuscripts received for review must be treated as confidential documents. They must not be shown to or discussed with others except if authorized by the editor.

Standards of objectivity

Reviews should be conducted objectively. Personal criticism of the author(s) is unacceptable. Referees should express their views clearly with appropriate supporting arguments.

Acknowledgement of sources

Reviewers should identify relevant published work that has not been cited by the author(s). Any statement that an observation, derivation, or argument had been previously reported should be accompanied by the relevant citation. Reviewers should also call to the editor's attention any substantial similarity or overlap between the manuscript under consideration and any other published data of which they have personal knowledge.

Disclosure and conflict of interest

Privileged information or ideas obtained through peer review must be kept confidential and not used for personal advantage. Reviewers should not consider evaluating manuscripts in which they have

conflicts of interest resulting from competitive, collaborative, or other relationships or connections with any of the authors, companies, or institutions connected to the submission.

3. Author Responsibilities

Reporting standards

Authors reporting results of original research should present an accurate account of the work performed as well as an objective discussion of its significance. Underlying data should be represented accurately in the manuscript. A paper should contain sufficient detail and references to permit others to replicate the work. Fraudulent or knowingly inaccurate statements constitute unethical behavior and are unacceptable.

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.

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. Parallel submission of the same manuscript to more than one journal constitutes unethical publishing behavior and is unacceptable.

Acknowledgement of sources

Proper acknowledgment of the work of others must always be given. Authors should also cite publications that have been influential in determining the nature of the reported work.

Authorship of a manuscript

Authorship should be limited to those who have made a significant contribution to the conception, design, execution, or interpretation of the reported study. All those who have made significant contributions should be listed as co-authors. Where there are others who have participated in certain substantive aspects of the research project, they should be named in an Acknowledgement section. The corresponding author should ensure that all appropriate co-authors are included in the author list of the manuscript, and that all co-authors have seen and approved the final version of the paper and have agreed to its submission for publication. All co-authors must be clearly indicated at the time of manuscript submission. Request to add co-authors, after a manuscript has been accepted will require approval of the editor.

Hazards and human or animal subjects

If the work involves chemicals, procedures, or equipment that has any unusual hazards inherent in their use, the authors must clearly identify these in the manuscript. Additionally, manuscripts should adhere to the principles of the World Medical Association (WMA) Declaration of Helsinki regarding research study involving human or animal subjects.

Disclosure and conflicts of interest

All authors should disclose in their manuscript any financial or other substantive conflict of interest that might be construed to influence the results or their interpretation in the manuscript. All sources of financial support for the project should be disclosed.

Fundamental errors in published works

In case 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's editor to either retract the paper or to publish an appropriate correction statement or erratum.

4. Publisher Responsibilities

Editorial autonomy

Academic Perspective Foundation is committed to working with editors to define clearly the respective roles of publisher and of editors in order to ensure the autonomy of editorial decisions, without influence from advertisers or other commercial partners.

Intellectual property and copyright

We protect the intellectual property and copyright of Academic Perspective Foundation, its imprints, authors and publishing partners by promoting and maintaining each article's published version of record. Academic Perspective Foundation ensures the integrity and transparency of each published article with respect to: conflicts of interest, publication and research funding, publication and research ethics, cases of publication and research misconduct, confidentiality, authorship, article corrections, clarifications and retractions, and timely publication of content.

Scientific Misconduct


In cases of alleged or proven scientific misconduct, fraudulent publication, or plagiarism the publisher, in close collaboration with the editors, will take all appropriate measures to clarify the situation and to amend the article in question. This includes the prompt publication of a correction statement or erratum or, in the most severe cases, the retraction of the affected work.

Contents

Research Articles		
Title	Authors	Pages
Human Activity Recognition with Smartwatch Data by using Mahalanobis Distance-Based Outlier Detection and Ensemble Learning Methods	Ensar Arif SAĞBAŞ, Serkan BALLI	95 - 106
Optimizing PID Gains of a Vehicle using the state-of-the-art Metaheuristic Methods	Hilal ARSLAN, Mustafa Atakan AFŞAR	107 - 117
Is ChatGPT Leading Generative AI? What is Beyond Expectations?	Ömer AYDIN, Enis KARAARSLAN	118 - 134
Home Healthcare Routing Problem for Multiple Hospitals with Balanced Workload	Gülçin DİNÇ YALÇIN, Tuğçe YAVUZ, Şüheda ALTINTAŞ	135 - 150
A Novel Clustering-based Forecast Framework: The Clusters with Competing Configurations Approach	Gökhan DEMİRKIRAN, Miray ALP	151 - 162
Artificial Neural Networks-Based Route Selection Model for Multimodal Freight Transport Network During Global Pandemic	Elif CESUR, Yaşanur KAYIKCI	163 - 173
An Explorative Analysis of Tweets Sentiments for Investment Decision in Stock Markets	Aysun GÜRAN, Emine ATEŞ	174 - 183
Synchronization of Incommensurate Fractional-Order King Cobra Chaotic System	Abdullah GÖKYILDIRIM, Haris ÇALGAN	184-190

Human Activity Recognition with Smartwatch Data by using Mahalanobis Distance-Based Outlier Detection and Ensemble Learning Methods

*¹Ensar Arif SAĞBAŞ, ²Serkan BALLI

¹ Department of Information Systems Engineering, Muğla Sıtkı Kocman University, Muğla, Türkiye, arifsagbas@mu.edu.tr 

² Department of Software Engineering, Burdur Mehmet Akif Ersoy University, Burdur, Türkiye, serkanballi@mehmetakif.edu.tr 

Abstract

Recognition of human activities is part of smart healthcare applications. In this context, the detection of human activities is an area that has been studied for many years. In these studies, various motion sensors placed in the body are utilized. With the increase in the usage of smart devices, smartphones, and smartwatches have become the constant equipment of these studies thanks to their internal sensors. Sometimes abnormal data are included in data sets due to the way the data were collected and for reasons arising from the sensors. For this reason, it becomes important to detect outlier data. In this study, step counter and heart rate sensors were used in addition to an accelerometer and gyroscope in order to detect human activities. Afterward, the outliers were detected and cleared with a Mahalanobis distance-based approach. With the aim of achieving a better classification performance, machine learning methods were used by strengthening them with ensemble learning methods. The obtained results showed that step counter, heart rate sensors, and ensemble learning methods positively affect the success of the classification. In addition, it was found that the Mahalanobis distance-based outlier detection method increased the classification accuracy significantly.

Keywords: activity recognition; smartwatch; mahalanobis distance; ensemble learning; sensor data

1. INTRODUCTION

Multiple participants, such as doctors and patients, hospitals, and research organizations, are involved in smart healthcare. Disease prevention and monitoring, diagnosis and treatment, hospital management, health-related decision-making, and medical research are all part of this organic whole. Smart health services, for example, are built on the foundation of information technologies such as the Internet of Things, mobile internet, cloud computing, big data, 5G, microelectronics, and artificial intelligence, as well as current biotechnology. These technologies are frequently employed in smart healthcare in all sectors. Patients can utilize wearables to always keep track of their health, receive medical assistance through virtual assistants, and use remote homes to implement remote services; physicians can employ a range of sophisticated clinical decision support systems to aid and improve diagnosis. The adoption of mobile medical platforms can help patients have a better experience. For scientific research institutions, techniques like machine learning can be used instead of manual drug screening, and big data can be used to locate appropriate themes [1].

Recognition of human activities is a useful task in many subjects such as fall detection of elderly people, healthcare

applications, and tracking daily routines [2, 3], etc. With the expansion of the usage of wearable sensors, recognition of human activities is possible by using ubiquitous devices, such as smartphones and smartwatches. The increasing popularity of smartwatches facilitates personal health monitoring [4]. These devices have a lot of built-in sensor equipment for instance accelerometer, gyroscope, step counter, etc. With the aid of these sensors, the classification of human activities may be done by using machine learning methods. However, sometimes these sensor signals are exposed to effects such as noise. This situation causes a bad fit between the data and includes it in the general pattern calculation [5]. Clearing outlier data is an important preprocessing step to create more consistent models. Ensemble methods are also used in activity recognition. Ensembles classifiers train more than one base learner instead of the single base learner. Thus, these methods can contribute to increasing the accuracy rate.

In this study, step counter and heart rate sensors were used in addition to the accelerometer and gyroscope sensors to create a new dataset. Dataset consists of walking, jogging, writing on paper, writing on the blackboard, typing, stationary, vacuuming, and brushing teeth activities. Outlier data from the created data set was cleared with a structure

built based on Mahalanobis distance. This approach based on the detection and cleaning of outliers and the combination of multiple sensors used is the main novelty of this study. Then performances of sensor data combinations were evaluated by using machine learning methods and ensemble learning approaches. Machine learning methods and ensemble approaches which are frequently encountered in the literature and have high success rates were chosen [6]. The most accurate result was obtained from the Random Subspace ensemble of the kNN method with the accelerometer, gyroscope, heart rate, and step counter sensors combination. Step counter and heart rate sensors increased the success rate.

The paper is organized as follows: previous studies will be mentioned in section two. After that machine learning methods and ensemble learning approaches, the creation of the dataset, and Mahalanobis distance-based outlier detection will be described, used sensors clarified in Section three. Afterward, the experimental results of the study will be handled in Section four. In the continuation, the impact of Mahalanobis distance-based outlier detection on classification will be discussed in Section five. Eventually, section six will conclude the paper.

2. RELATED WORK

While examining the related works, various studies were found about activity recognition by using machine learning methods and wearable devices. Asarakaya and Ünsal [7] aimed to define human activities using machine learning methods on data obtained from smart sensors. Sağbaşı and Ballı [8] detected the transportation modes (traveling by bus or car, cycling, running, and walking) of the users by using smartphone sensors (accelerometer, gyroscope, and GPS). Six different machine learning techniques (Bayesian Network, Naive Bayes, kNN, Random Forest, J48, and Logistic Regression) were tested and their performances were compared. Erin et al. [9] performed the detection of human activities based on the internet of things by using the accelerometer sensor of the device with the android software developed for the mobile device. Voicu et al. [10] classified six human activities in eight different scenarios by using smartphone sensors. Ballı et al. [2] proposed a mobile solution for the detection of falls using together smartwatches and smartphone sensors. Peker et al. [11] predicted human activities with the data obtained from the smartwatch. They first applied the ReliefF attribute selection and then classified them with the Kernel-Based Extreme Learning Machine method. Ahmed et al. [12] proposed a hybrid feature selection method for human activity detection. Yahaya et al. [13] proposed an approach to identify sources of anomalies in human activities. Li et al. [14] proposed a new method of feature extraction based on linear predictive analysis (LPA) to reduce the computational complexity in activity classification using acceleration signals. Gani et al. [15] offered a computationally efficient, smartphone-based human activity recognition system based on chaos theory and dynamic systems. Elsts et al. [16] proposed an energy-efficient activity recognition framework with two key components by using a wearable accelerometer.

In addition, it is possible to come across various studies using deep learning methods. Challa et al. [17] used a hybrid of CNN and BiLSTM models to design a robust classification model for human activity recognition using wearable sensor data. Metin and Karasulu [18] compared the performance of deep learning techniques to classify daily human activities. Munez-Organero [19] proposed an outlier detection algorithm based on Deep recurrent neural networks for detecting human activities. Zhou et al. [20] proposed a Convolutional Neural Network-based structure to detect nine indoor human actions from smartphone sensor data. Wan et al. [21] designed a smartphone accelerometer and deep learning-based architecture for human activity detection. Zhou et al. [22] designed a semi-supervised deep learning framework that efficiently uses weakly labeled sensor data in activity detection. Altuve et al. [23] classified six different human activities using bidirectional LSTM. Mukherjee et al. [24] determined human body movements by using data obtained from smart device sensors and a collection of three classification models, namely CNN-Net, Encoded-Net, and CNN-LSTM, which are called EnsemConvNet.

Ensemble methods were also investigated in various studies. Catal et al. [25] investigated the power of the ensemble of classifiers approach for accelerometer-based activity recognition and built a novel activity estimation model grounded on machine learning classification methods. Elamvazuthi et al. [26] tested five different ensemble learning methods for classifying six daily activities. They gained inertial sensor data from smartphones. Ballı et al. [3] proposed a hybrid structure using principal component analysis and Random Forest methods for activity recognition with smartwatches. Herrera-Alcantara et al. [27] observed the activities of students with smartwatches. They obtained the most satisfactory result from the Random Forest method. Irvine et al. [28] proposed a new neural network ensemble method that is aiming to improve the human activity recognition dataset. The ensemble-based approach to detecting human activity was discussed in detail in the study conducted by Brajesh and Ray [29]. Sekiguchi et al. [30] increased the classification success of activity detection with an ensemble model that includes a CNN model and a gradient-boosting model. Subasi et al. [31] used the Adaboost ensemble to classify human activities. Dwivedi et al. [32] introduced a new skeleton-based feature for human activity recognition and used it to train the Random Forest classifier.

This study differs from the other study in the literature with the used sensor combinations. In addition, the effect of outlier data cleaning on classification success was investigated. High classification accuracies were achieved with classification ensemble learning-based approaches.

3. MATERIALS AND METHODS

In this section, the machine learning methods and ensemble learning approaches, the smart device sensors used in the study, the creation of the data set, and Mahalanobis distance-based outlier detection will be explained.

3.1. Machine Learning Methods

Machine learning methods were successfully used in human activity recognition in previous research. Machine learning methods construct an algorithm and make predictions from the dataset. In this section, the machine learning methods used in this study are summarized.

3.1.1. Naïve Bayes (NB)

The Naive Bayes Classifier is a simple probability-based algorithm with a strong assumption of attribute independence. The Naive Bayes Classifier performs learning through test data and incorporates the best proportion of the instance into the class [33, 34]. The Bayes theorem given in Equation 1 is used to estimate the class based on the test data.

$$p(A|B) = \frac{p(A)p(B|A)}{p(B)} \quad (1)$$

Here, $P(A|B)$ is the probability that event A will occur when event B is known. $P(B|A)$ is the probability of event B occurring when event A is known. $P(A)$ is the probability of occurrence of event A. $P(B)$ is the probability of occurrence of event B.

3.1.2. k Nearest Neighbor (kNN)

kNN is based on the similarity between sample data and dataset. The number of nearest neighbors to be considered for the classification in the kNN algorithm is expressed as a positive integer such as k. In determining the closest neighbors, the closeness between the samples in the training set and the selected sample is determined. The closeness between the samples is sorted in ascending. It presents the order from the nearest neighbor to the farthest neighbor of the selected sample [35, 36]. The euclidian distance formula in Equation 2 was used in this study to calculate distance.

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (2)$$

Here p and q represent points on the hyperplane of the data.

3.1.3. J48

J48 is the Java implementation of the C4.5 algorithm. The C4.5 algorithm consists of the classification trees of attributes between numerical and categorical values. It is important which attribute to start branching when creating classification trees. Using a training data set to reveal all possible tree structures and choose the most suitable one among these tree structures causes a lot of repetitions. Therefore, classification tree algorithms are going to create trees according to these values by calculating various values at the beginning of the process. For this purpose, the concept of entropy can be applied. The branching of the tree starts by considering the value of entropy [25, 37]

3.1.4. Random Forest (RF)

Breiman [38] presented the idea of using a huge number of tree structures instead of an individual tree, that is, the use of a forest for classification purposes. In the Random Forest

method, samples are selected from the dataset by the Bootstrap method. Then classification trees are generated based on these samples. Using these classification trees, the class of observation is estimated for each tree and the most repeated class value is selected among the classifications [39]. Due to the tree data structure, J48 and Random Forest methods are harder to implement and take more time than other methods for activity recognition applications.

3.2. Ensemble Methods

Four ensemble methods were compared in this study. These methods are AdaBoostM1, Bagging, Random Subspaces, and Voting. In this section, these methods are briefly described.

3.2.1. AdaBoostM1

AdaBoost (Adaptive Boosting) is a popular ensemble algorithm. It was presented in 1996. This algorithm focuses on patterns that are harder to classify. In each iteration of the algorithm, the weights of misclassified samples are increased, and the weights of correctly classified samples are reduced [40].

3.2.2. Bagging

The term Bagging (Bootstrap AGGREGatING) is introduced by Breiman [41]. Bagging is a simple and appealing idea. This idea builds a new classifier from the training set. In this case, some samples in the dataset are not included in the newly created training set, while others are found more than once. Each basic classifier in the ensemble structure is trained with these training sets that contain different examples. The output of the classifier is combined with plurality voting to conclude [42].

3.2.3. Random Subspaces

The Random Subspaces algorithm is an ensemble method that randomly selects several components of the given feature vector in constructing each classifier. The purpose of this algorithm is to avoid overfitting while boosting the predictive performance [6].

3.2.4. Voting

Voting is a simple method that combines basic learning algorithms. There are various methods of combining the outputs of basic classification algorithms. These hybrid methods include majority voting and weighted majority voting. In simple majority voting, the result decided to receive the most votes among the k basic classification algorithms is determined as the output of the ensemble classifier [6, 43].

3.3. Dataset and Feature Extraction

Step counter, heart rate monitor, three axes accelerometer, and gyroscope sensor data acquired from the smartwatch (Moto 360) in Figure 1 were used in this study. All data were collected while the smartwatch is attached to the wrist of the

user. This device has 512 MB RAM, a quad-core 1.2 GHz processor, and a built-in gyroscope, step counter, accelerometer, optical heart rate monitor, and ambient light sensors. It was set to collect 50 samples per second.



Figure 1. Smartwatch that used in this study

All sensor data were labeled during the data collection phase and split into two seconds size of windows. 500 patterns which include 100 sample sensor data were obtained for each activity: jogging, walking, brushing teeth, writing on paper, typing, writing on a blackboard, stationary, and vacuuming. Jogging means running at a slower velocity. Especially, heart rate value varies in jogging activity and running activity. Stationary includes sitting and standing activities. Using the keyboard means working at a computer and typing with the keyboard.

In the dataset, each activity has the same number of samples. The average value of heart rate, accelerometer, and gyroscope sensors, the standard deviation of accelerometer and gyroscope sensors, and the number of steps (total 14 features) form a pattern. The list of all features and their ranges is given in Table 1.

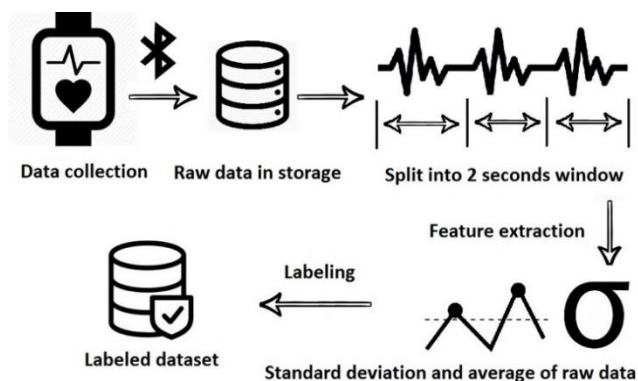


Figure 2. Dataset lifecycle

Table 1. The list of all features and their ranges

Feature name	Range
Average value of accelerometer X	0.0004-10.7696
Average value of accelerometer Y	0.0043-11.6919
Average value of accelerometer Z	0.0028-0.9369
Standard deviation value of accelerometer X	0.0055-12.2600
Standard deviation value of accelerometer Y	0.0098-16.4635
Standard deviation value of accelerometer Z	0.0053-6.4985
Average value of gyroscope X	0.0000-1.1898
Average value of gyroscope Y	0.0000-1.3245
Average value of gyroscope Z	0.0000-0.8611
Standard deviation value of gyroscope X	0.0034-4.9206
Standard deviation value of gyroscope Y	0.0012-2.3982
Standard deviation value of gyroscope Z	0.0008-2.6774
Number of steps	0-20
Average value of heart rate	0-180.83

Before extracting the features, raw data were not filtered, and all variables were calculated per two second window. Through two-second window, the developed activity recognition system continues to accurately identify activities if the beginning is missed. The dataset lifecycle is given in Figure 2.

3.4. Sensors

To detect human activities based on wrist motion; an accelerometer, gyroscope, step counter, and heart rate sensors were used. This section briefly describes these sensors.

3.4.1. Accelerometer

Acceleration force is applied to a device on the x, y, and z axes (Figure 3), including the force of gravity. The accelerometer measures the acceleration force in m/s² [44].



Figure 3. Smartwatch accelerometer axes

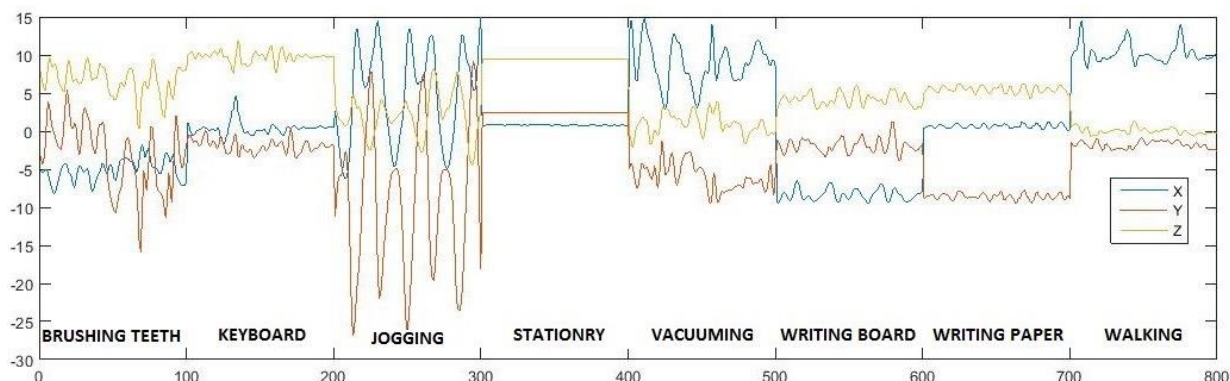


Figure 4. Amplitude changes of accelerometer axes

Figure 4 shows the amplitude change of accelerometer axes for eight different daily activities (brushing teeth, using the keyboard, jogging, stationary, vacuuming, writing on the blackboard, writing on paper, and walking).

In this study, the standard deviation and the average value of accelerometer data were selected as acceleration features. 3-D representations of standard deviations of accelerometer data for each activity are shown in Figure 5.

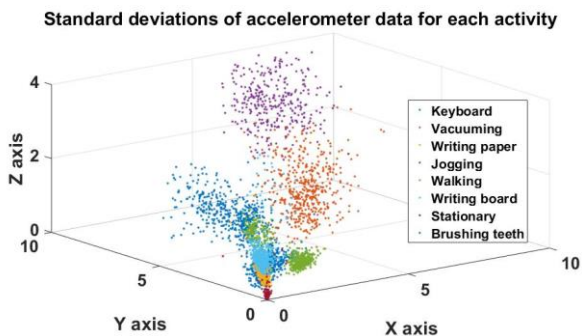


Figure 5. Standard deviations of accelerometer data for each activity

3.4.2. Gyroscope

The gyroscope detects the roll, pitch, and yaw motions of the devices along the x, y, and z-axes, respectively and it calculates the device's rotation rate. The axes' directions are shown in Figure 3. The raw data stream from a gyroscope sensor is the rate of the rotation around each of the three physical axes in rad/s (radian per second) [44]. In this study, the standard deviation and the average value of gyroscope

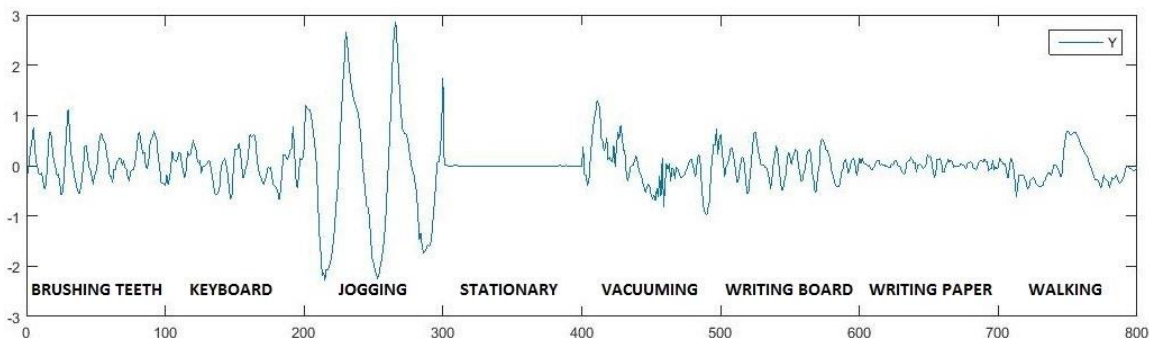


Figure 6. Amplitude changes of gyroscope y-axis

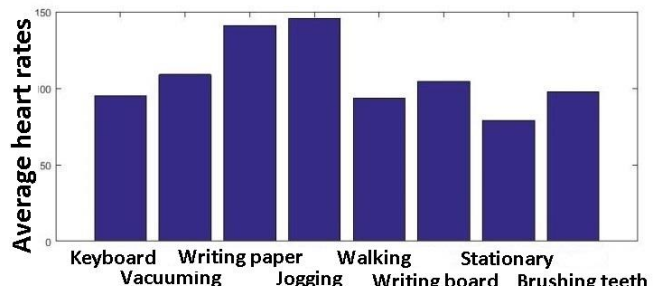


Figure 8. Average heart rates for each activity

3.5. Mahalanobis Distance-Based Outlier Detection

In an ellipse, some points are closer to the center than others (Figure 9), but it cannot be concluded that the more distant

data were used. Figure 6 shows the y-axis amplitude change of the gyroscope sensor for eight different daily activities.

3.4.3. Step Counter

The step counter sensor returns the number of users' steps since the last reboot while activated. The value is returned as a float (with the fractional part set to zero) and is reset to zero only on a system restart. The timestamp of the event is set to the time when the last step for that event was taken [3]. The average number of steps for each activity is given in Figure 7.

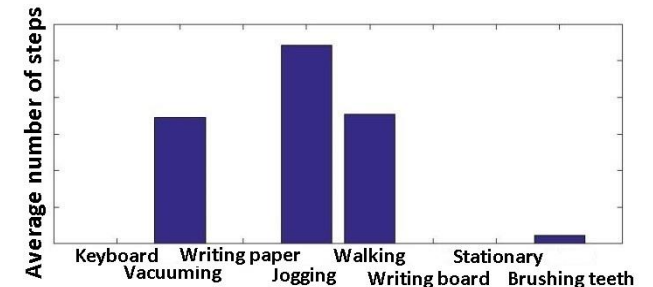


Figure 7. Average number of steps for each activity

3.4.4. Heart Rate Monitor

The reported value is the heart rate in beats per minute. The reported accuracy represents the status of the monitor during the reading [2]. The heart rate sensor rarely returns 0 when the signal is distorted. Average heart rates for each activity are given in Figure 8.

points belong less than the points closer to the sample, since this is part of the basic model of the normal distribution. Therefore, instead of the classical distance, it is recommended to use a distance that considers the shape of the observations under investigation, and such a distance is the Mahalanobis distance denoted by d [45]:

$$d = \sqrt{(x - m)^T C^{-1} (x - m)} \tag{3}$$

where x is a vector of variables $x = (x_1, x_2, \dots, x_k)$, $m = (m_1, m_2, \dots, m_k)$ is a k -dimensional vector and C is a $k \times k$ symmetric matrix. It measures the distance from a point x to the center of m in metric C , meaning that the distance depends on the shape. Naturally, the values of m and C are practically unknown and therefore need to be estimated.

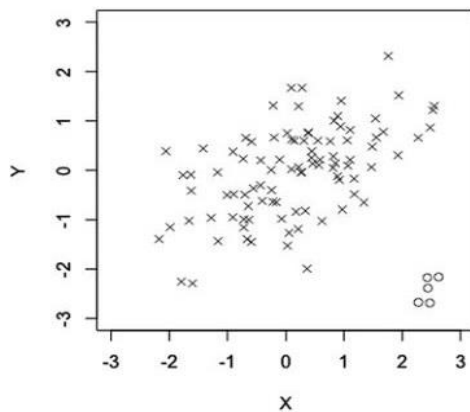


Figure 9. Scatter plot of two variables X, Y sampled from the normal distribution and 5 outliers (circles) [46]

In this study, the data were considered class by class in clearing out the outliers. Mahalanobis distances were calculated to detect outliers. Then, the probability values (1-ChiSquare) were calculated with the SPSS program. The calculated probability, those below the 0.001 threshold, were marked as outliers and deleted from the data set. After clearing the outliers, the number of patterns and cleared data ratios belonging to the classes are presented in Table 2.

It is seen that approximately 4% of the data for each class was selected as outlier data and cleared from the data set. It is striking that the rate of outlier data pertaining only to writing on the board is higher than in the other classes. The size of the dataset decreased by 4.4% after the cleaning process.

The outliers in the sensor data are caused by the participant's out-of-class movements and noisy data. Considering that the step counter sensor is also a multiple motion detector, it is likely to be affected by noisy data. When Table 1 in Section 3.3 is examined, it is seen that the lowest value of the heart rate is 0. This indicates that erroneous readings of the heart rate sensor are also present. In summary, it can be said that there is a certain amount of outlier data in all sensor data.

Table 2. Number of patterns belonging to classes after outliers are cleared

Class name	Number of patterns	Cleared data ratio
Brushing teeth	481	3.8%
Writing on the paper	479	4.2%
Writing on the board	452	9.6%
Walking	484	3.2%
Vacuuming	477	4.6%
Stationary	479	4.2%
Keyboard	486	2.8%
Jogging	485	3.0%

4. EXPERIMENTAL RESULTS

To compare the performance metrics, six different types of sensor combinations (Table 3) were tested with machine learning and ensemble methods. A Diagram of the system architecture is shown in Figure 10. k-fold cross-validation was employed to improve the performance of the used learning method. In cross-validation, the dataset is split into k groups, and the method is employed in each group. In each trial, one of the k-groups is selected as the test set and the other k-1 groups are used to create a training set. Then the mean error through all k trials is calculated [47]. With this approach, all samples in the dataset are used both in the testing phase and in the training phase.

Naive Bayes, kNN, J48, and Random Forest, which are frequently encountered in the literature, were used as base classification methods. In ensemble construction, four ensemble methods (AdaBoost M1, Bagging, Random Subspaces, and Vote) were analyzed. In the voting decision step, NB, kNN, J48, and Random Forest algorithms were chosen, and the average of probabilities combination rule was selected. The experiments were applied with WEKA [48] toolkit version 3.8. The results of the tests were compared with classification accuracy. Classification accuracy rates of methods for datasets obtained without cleaning outliers are given in Table 4 and Figure 11. According to tests carried out, obtaining the best results is displayed in bold.

Table 3. Model structure and input variables for the six models

Model No	Model Name	Contents
1	Accelerometer only	Standard deviation and average value of x, y and z axes accelerometer sensor data (6 features)
2	Accelerometer and Gyroscope	Standard deviation and average value of x, y and z axes accelerometer and gyroscope sensor data (12 features)
3	Accelerometer, Gyroscope and Step Counter	Standard deviation and average value of x, y and z axes accelerometer and gyroscope sensor data and number of steps (13 features)
4	Accelerometer, Gyroscope, Step Counter and Heart Rate	Standard deviation and average value of x, y and z axes accelerometer and gyroscope sensor data, number of steps and average heart rate (14 features)
5	Accelerometer and Step Counter	Standard deviation and average value of x, y and z axes accelerometer sensor data and number of steps (7 features)
6	Accelerometer, Step Counter and Heart Rate	Standard deviation and average value of x, y and z axes accelerometer sensor data, number of steps and average heart rate (8 features)

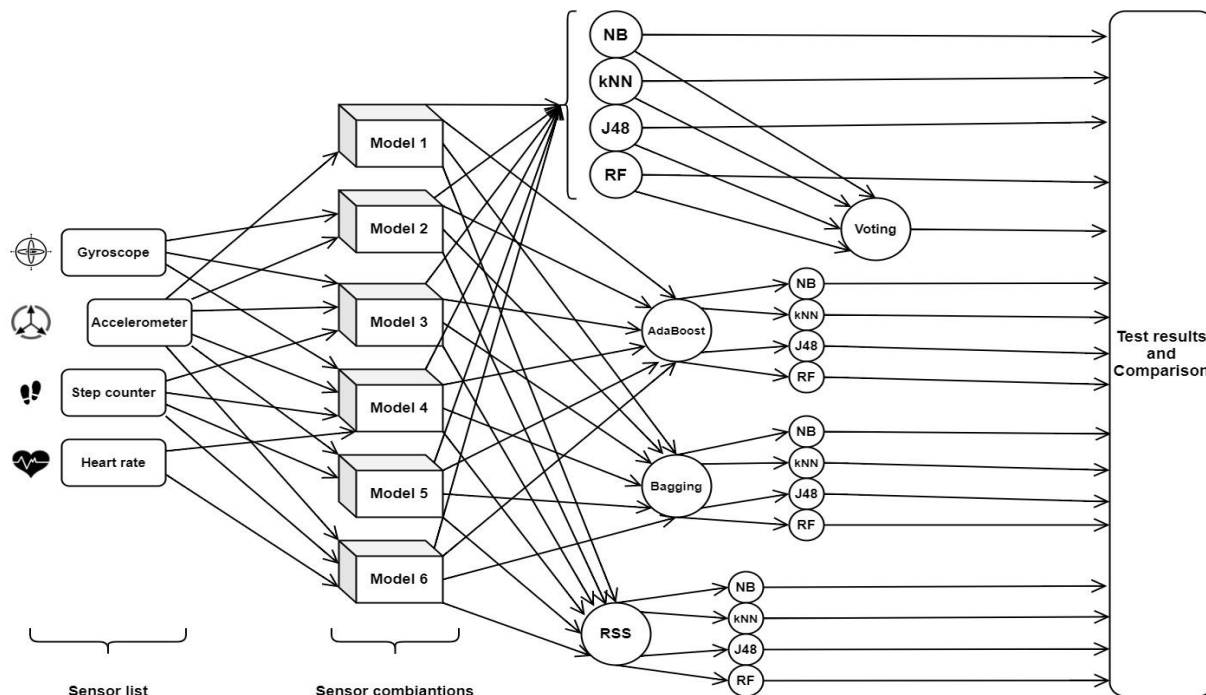


Figure 10. Diagram of the system architecture

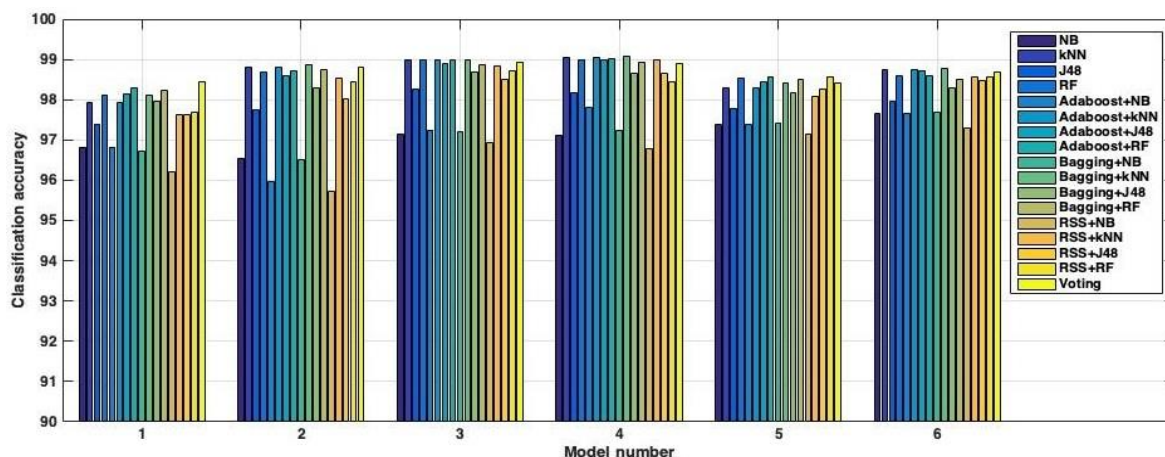


Figure 11. Bar chart of results from the dataset without cleaning outliers

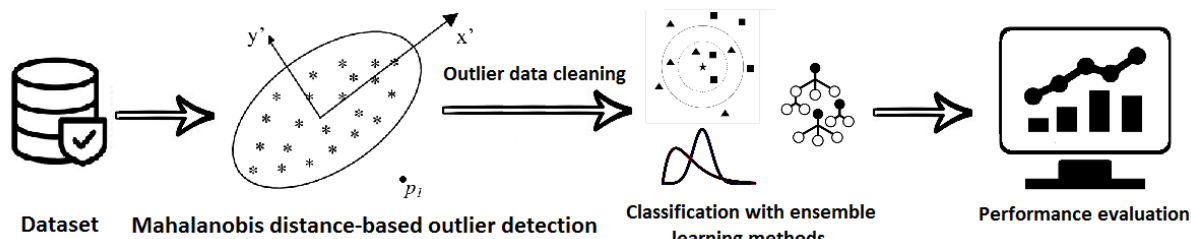


Figure 12. Flow chart of outlier data cleaning and performance evaluation stage

Six different types of sensor combinations were compared in this study, and it is seen that the best classifications based on methods are mostly obtained by using the Model 4 sensor combination (Accelerometer, gyroscope, step counter, and heart rate). It is noteworthy that the accelerometer and step counter have an important place in classification performance. The most successful method was found to be the Bagging ensemble of kNN with an accuracy rate of 99.075%. Thus, it can be said that the heart rate sensor also has a positive effect on classification success. The lowest classification accuracy (95.725%) was obtained from Model

2 (Accelerometer and Gyroscope) with RSS + NB. When comparing Model 2 with Model 1, the accuracy rates of all methods with NB combinations decreased. However, an increase was observed in other methods.

The tests performed with the ensemble learning methods were re-performed after clearing the outliers and their performance was compared. The flow chart of the outlier data cleaning and performance evaluation stage is shown in Figure 12. The accuracy rates for the dataset obtained with outlier data cleared are shown in Table 5 and Figure 13.

Table 4. Classification accuracy rates for dataset obtained without cleaning outliers

Method/Model	1	2	3	4	5	6
NB	96.825	96.550	97.150	97.125	97.375	97.650
kNN	97.925	98.800	98.975	99.050	98.300	98.750
J48	97.375	97.750	98.250	98.175	97.775	97.950
RF	98.125	98.700	99.000	99.000	98.525	98.600
AdaBoost+NB	96.825	95.975	97.250	97.800	97.375	97.650
AdaBoost+kNN	97.925	98.800	98.975	99.050	98.300	98.750
AdaBoost+J48	98.150	98.600	98.900	99.000	98.450	98.725
AdaBoost+RF	98.300	98.725	98.975	99.025	98.550	98.600
Bagging+NB	96.725	96.500	97.200	97.225	97.425	97.675
Bagging+kNN	98.100	98.875	98.975	99.075	98.425	98.775
Bagging+J48	97.950	98.300	98.700	98.650	98.175	98.300
Bagging+RF	98.225	98.750	98.875	98.925	98.500	98.500
RSS+NB	96.200	95.725	96.925	96.775	97.150	97.300
RSS+kNN	97.625	98.525	98.825	98.975	98.075	98.575
RSS+J48	97.625	98.025	98.500	98.650	98.275	98.475
RSS+RF	97.700	98.450	98.725	98.450	98.550	98.550
Voting	98.450	98.800	98.925	98.900	98.425	98.700

According to Table 5, the best performance (99.686%) was obtained by the Model 4 sensor combined with the Random Subspace ensemble of the kNN method. It was observed that only the accelerometer sensor provides a high success rate (minimum 97.489%). Step counter and heart rate sensors and ensemble methods contribute to increasing the accuracy. Model 6 (Accelerometer + Step counter + Heart rate) provided the best result in 9 of the 17 classification methods. When Model 1 and Model 2 were compared, there was no increase in the classification success of only 4 methods. Three classification methods provided the most successful

results with Model 3 (Accelerometer + Gyroscope + Step counter). There was not any method that achieved the most successful result with Model 5. These comparisons highlight the importance of the gyroscope sensor. However, the most successful classifications include step counter and heart rate sensors. The positive effects of the new sensors tested are supported by these experiments. In addition, obtaining the highest classification accuracy rate with the Random Subspace ensemble shows that not all the extracted features are efficient.

Table 5. Classification accuracy rates for dataset obtained with outlier data cleaned

Method/Model	1	2	3	4	5	6
NB	98.012	98.117	98.484	98.640	98.718	98.771
kNN	99.111	99.425	99.503	99.582	99.215	99.608
J48	98.666	98.535	98.928	98.849	98.928	98.954
RF	99.346	99.503	98.247	99.608	99.608	99.634
AdaBoost+NB	98.666	99.058	99.163	99.163	99.032	98.849
AdaBoost+kNN	99.111	99.425	99.503	99.582	99.215	99.608
AdaBoost+J48	99.372	99.372	99.425	99.503	99.398	99.320
AdaBoost+RF	99.398	99.582	99.582	99.660	99.608	99.608
Bagging+NB	97.934	98.064	98.509	98.666	98.666	98.771
Bagging+kNN	99.163	99.320	99.477	99.529	99.268	99.555
Bagging+J48	98.980	99.006	99.032	99.006	98.928	99.111
Bagging+RF	99.425	99.555	99.608	99.555	99.477	99.529
RSS+NB	97.489	96.940	98.718	98.666	98.404	98.666
RSS+kNN	98.980	99.582	99.582	99.686	99.241	99.425
RSS+J48	99.032	99.268	99.425	99.555	99.425	99.425
RSS+RF	99.215	99.503	99.634	99.660	99.477	99.503
Voting	99.451	99.451	99.451	99.425	99.294	99.503

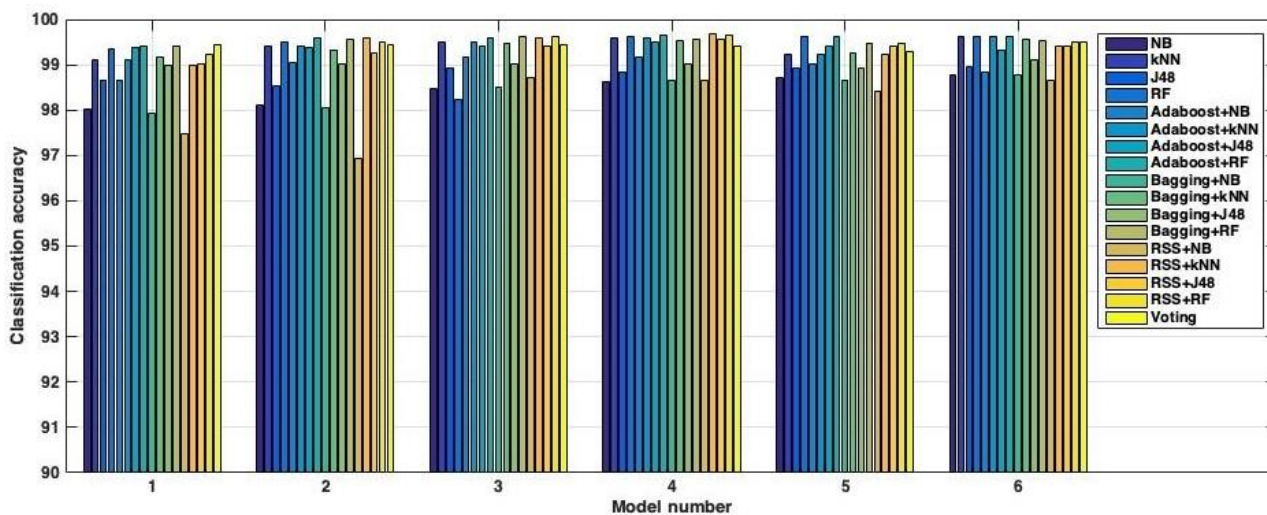


Figure 13. Bar chart of results from the dataset with outlier data cleaned

5. DISCUSSION

In this section, the effect of Mahalanobis distance-based outlier detection and ensemble learning approaches on classification performance will be discussed. When Table 4 and Table 5 in Section 4 are examined, the effect of sensor data on classification success can be seen clearly. The best results were obtained with Model 3, Model 4, and Model 6. All these Models include a step counter. The Model 4 and Model 5 include a heart rate sensor in addition to the step

counter. The improvement rates are presented in Table 6 and Figure 14 to better understand the extent to which the outlier data cleaning affects the classification success.

When the differences between classification accuracies are examined, it is seen that only 1 of the 102 experiments performed decreased. An increase was observed in all other experiments. The highest increase was provided by AdaBoost + NB method and Model 2. Mahalanobis distance-based outlier detection and cleaning increased the

classification accuracy by an average of 1 point. An N-way analysis of variance test was used to determine statistical significance. The attained p-value given across the multiple classification methods was p 0.01 at a significance threshold of 0.05, indicating substantial differences in the accuracy rate achieved by the different classification methods. The confusion matrix for Model 4 sensor combination with Random Subspace ensemble of kNN is given in Table 7.

According to Table 7, recognition accuracies for walking, writing on paper, and jogging were 100%. Writing on the board activities was confused with brushing teeth and brushing teeth activities were confused with writing on board and vacuuming. One of the vacuuming and stationary activities was misclassified as walking. Because the user was performing these activities in a standing position like walking.

It is not possible to directly compare this study, which was carried out by creating a new data set, with other studies in the literature. In addition to motion sensors, heart rate sensors were used in the study. Considering the total number of samples in the dataset, it was thought that it would be appropriate to use classical machine learning methods instead of deep learning methods. The ensemble learning

approach applied after outlier data cleaning provided high-accuracy classification success. But the main drawback of the proposed method is that ensemble learning methods (they have a sequential or parallel operation) need a long computation time and relatively much memory.

Table 6. Improvement rates after outlier data cleaning

Method/Model	1	2	3	4	5	6
NB	1.187	1.567	1.334	1.515	1.343	1.121
kNN	1.186	0.625	0.528	0.532	0.915	0.858
J48	1.291	0.785	0.678	0.674	1.153	1.004
RF	1.221	0.803	-0.753	0.608	1.083	1.034
AdaBoost+NB	1.841	3.083	1.913	1.363	1.657	1.199
AdaBoost+kNN	1.186	0.625	0.528	0.532	0.915	0.858
AdaBoost+J48	1.222	0.772	0.525	0.503	0.948	0.595
AdaBoost+RF	1.098	0.857	0.607	0.635	1.058	1.008
Bagging+NB	1.209	1.564	1.309	1.441	1.241	1.096
Bagging+kNN	1.063	0.445	0.502	0.454	0.843	0.78
Bagging+J48	1.03	0.706	0.332	0.356	0.753	0.811
Bagging+RF	1.2	0.805	0.733	0.63	0.977	1.029
RSS+NB	1.289	1.215	1.793	1.891	1.254	1.366
RSS+kNN	1.355	1.057	0.757	0.711	1.166	0.85
RSS+J48	1.407	1.243	0.925	0.905	1.15	0.95
RSS+RF	1.515	1.053	0.909	1.21	0.927	0.953
Voting	1.001	0.651	0.526	0.525	0.869	0.803

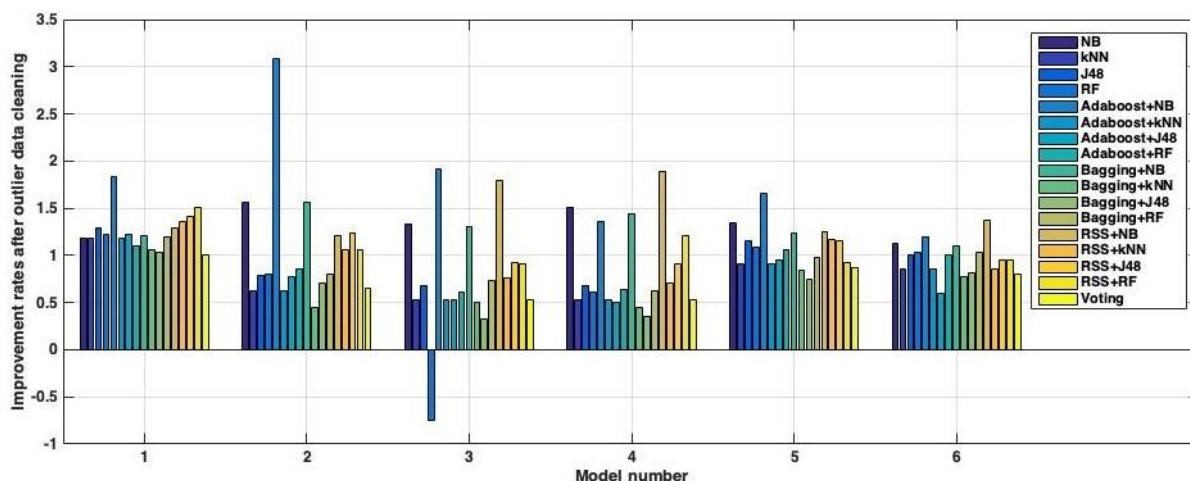


Figure 14. Bar chart of the improvement rates after outlier data cleaning

Table 7. Confusion matrix of Model 4 sensor combination with Random Subspace ensemble of kNN

Classified as	a	b	c	d	e	f	g	h	%
a=Brushing teeth	477	0	2	0	2	0	0	0	99.17
b=Writing on the paper	0	479	0	0	0	0	0	0	100
c=Writing on the board	3	0	448	0	1	0	0	0	99.12
d=Walking	0	0	0	484	0	0	0	0	100
e=Vacuuming	1	0	0	1	475	0	0	0	99.58
f=Stationary	0	0	0	1	0	478	0	0	99.79
g=Keyboard	0	1	0	0	0	0	485	0	99.79
h=Jogging	0	0	0	0	0	0	0	485	100

6. CONCLUSION

This study presents an effective outlier data clearing and ensemble learning-based approach to the recognition of human activities using smartwatch sensors. With the

Mahalanobis distance-based outlier detection, approximately 4% of the total data was detected as an outlier and cleared from the data set. Various ensemble approaches were tested in past studies. In this study, four different ensemble approaches were tested and the Random Subspace

ensemble of the kNN method achieved the most successful result for activity recognition by using the smartwatch. In addition, step counter and heart rate sensor data performances are investigated in this paper. These sensors also increase accuracy. According to the dataset used in the study, each ensemble method increases the success rate of different sensor combination models. The best result is obtained from the Model 4 sensor combined with the Random Subspace ensemble of the kNN method between all test options. The highest accuracy rate of 99.686% among all test options was obtained from this method. This demonstrates that all the sensors used in the study contribute to the classification and the RSS approach increases the classification success of the kNN method. In future works, this activity recognition study can be improved in various ways such as by increasing the number of classes (e.g.: handshake, smoking, cooking, and drinking, etc.) and expanding the dataset by collecting sensor data from different users since especially heart rate value may vary from person to person.

Author contributions: All authors contributed equally to the creation of the idea and the design. The mobile application used in the study was developed by the first author. The second author contributed to checking the article in terms of spelling and content.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: This study is supported by Mugla Sitki Kocman University Scientific Research Projects under the grant number 016-061.

REFERENCES

- [1] S. Tian, W. Yang, J.M. Le Grange, P. Wang, W. Huang, and Z. Ye, "Smart healthcare: making medical care more intelligent", *Global Health Journal*, vol. 3, no. 3, pp. 62-65, 2019. <https://doi.org/10.1016/j.glohj.2019.07.001>
- [2] S. Balli, E.A. Sağbaş, and M. Peker, "A mobile solution based on soft computing for fall detection". In *Mobile Solutions and Their Usefulness in Everyday Life*, Ed. Sara Paiva, 2019, ch. 14, pp. 275-294.
- [3] S. Balli, E.A. Sağbaş, and M. Peker, "Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm", *Measurement and Control*, vol. 52, no. 1-2, pp. 37-45, 2019. <https://doi.org/10.1177/0020294018813692>
- [4] B. Mortazavi, S. Nyamathi, S.I. Lee, T. Wilkerson, H. Ghasemzadeh, and M. Sarrafzadeh, "Near-realistic mobile exergames with wireless wearable sensors", *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 2, pp. 449-456, 2013. <https://doi.org/10.1109/JBHI.2013.2293674>
- [5] M.C. Limas, J.B.O. Meré, F.J.M. de Pisón Ascacibar, and E.P.V. González, "Outlier detection and data cleaning in multivariate non-normal samples: the PAELLA algorithm", *Data Mining and Knowledge Discovery*, vol. 9, no. 2, pp. 171-187, 2004. <https://doi.org/10.1023/B:DAMI.0000031630.50685.7c>
- [6] A. Onan, S. Korukoğlu, and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification", *Expert Systems with Applications*, vol. 57, pp. 232-247, 2016. <https://doi.org/10.1016/j.eswa.2016.03.045>
- [7] S. Asarkaya, and E. Ünsal, "Recognition of Human Activities Using Machine Learning Methods on Smart Sensor Data", *Journal of Science and Technology*, vol. 1, no. 1, pp. 9-14, 2022.
- [8] E.A. Sağbaş, and S. Balli, "Transportation mode detection by using smartphone sensors and machine learning", *Pamukkale University Journal of Engineering Sciences*, vol. 22, no. 5, pp. 376-383, 2016. <https://doi.org/10.5505/pajes.2015.63308>
- [9] K. Erin., C. Bayılmış, and B. Boru, "Real-Time Human Activity Detection Based on Accelerometers and Internet of Things", *Academic Platform-Journal of Engineering and Science*, vol. 9, no. 1, pp. 194-198, 2021. <https://doi.org/10.21541/apjes.809777>
- [10] R.A. Voicu, C. Dobre, L. Bajenaru, and R.I. Ciobanu, "Human physical activity recognition using smartphone sensors", *Sensors*, vol. 19, no. 3, pp. 458, 2019. <https://doi.org/10.3390/s19030458>
- [11] M. Peker, S. Balli, and E.A. Sağbaş, "Predicting human actions using a hybrid of ReliefF feature selection and kernel-based extreme learning machine", In *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications*, Ed. Mehdi Khosrow-Pour, 2020, ch. 17, pp. 307-325.
- [12] N. Ahmed, J.I. Rafiq, and M.R. Islam, "Enhanced human activity recognition based on smartphone sensor data using hybrid feature selection model", *Sensors*, vol. 20, no. 1, pp. 317, 2020. <https://doi.org/10.3390/s20010317>
- [13] S.W. Yahaya, A. Lotfi, and M. Mahmud, "Detecting anomaly and its sources in activities of daily living", *SN Computer Science*, vol. 2, no. 1, pp. 1-18, 2021. <https://doi.org/10.1007/s42979-020-00418-2>
- [14] R. Li, H. Li, and W. Shi, "Human activity recognition based on LPA", *Multimedia Tools and Applications*, vol. 79, no. 41, pp. 31069-31086, 2020. <https://doi.org/10.1007/s11042-020-09150-8>
- [15] M.O. Gani, T. Fayezeen, R.J. Povinelli, R.O. Smith, M. Arif, A.J. Kattan, and S.I. Ahamed, "A light weight smartphone based human activity recognition system with high accuracy", *Journal of Network and Computer Applications*, vol. 141, pp. 59-72, 2019. <https://doi.org/10.1016/j.jnca.2019.05.001>

- [16] A. Elsts, N. Twomey, R. McConville, and I. Craddock, "Energy-efficient activity recognition framework using wearable accelerometers", *Journal of Network and Computer Applications*, vol. 168, pp. 102770, 2020. <https://doi.org/10.1016/j.jnca.2020.102770>
- [17] S. K. Challa, A. Kumar, and V. B. Semwal, "A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data", *The Visual Computer*, vol. 38, no. 12, pp. 4095-4109, 2022. <https://doi.org/10.1007/s00371-021-02283-3>
- [18] İ. A. Metin, and B. Karasulu, "A novel dataset of human daily activities: Its benchmarking results for classification performance via using deep learning techniques", *Journal of the Faculty of Engineering and Architecture of Gazi University*, vol. 36, no. 2, pp. 759-777, 2021. <https://doi.org/10.17341/gazimmfd.772849>
- [19] M. Munoz-Organero, "Outlier detection in wearable sensor data for human activity recognition (HAR) based on DRNNs", *IEEE Access*, vol. 7, pp. 74422-74436, 2019. <https://doi.org/10.1109/ACCESS.2019.2921096>
- [20] B. Zhou, J. Yang, and Q. Li, "Smartphone-based activity recognition for indoor localization using a convolutional neural network", *Sensors*, vol. 19, no. 3, pp. 621, 2019. <https://doi.org/10.3390/s19030621>
- [21] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, "Deep learning models for real-time human activity recognition with smartphones", *Mobile Networks and Applications*, vol. 25, no. 2, pp. 743-755, 2020. <https://doi.org/10.1007/s11036-019-01445-x>
- [22] X. Zhou, W. Liang, I. Kevin, K. Wang, H. Wang, L.T. Yang, and Q. Jin, "Deep-learning-enhanced human activity recognition for Internet of healthcare things", *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6429-6438, 2020. <https://doi.org/10.1109/JIOT.2020.2985082>
- [23] M. Altuve, P. Lizarazo, and J. Villamizar, "Human activity recognition using improved complete ensemble EMD with adaptive noise and long short-term memory neural networks", *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 901-909, 2020. <https://doi.org/10.1016/j.bbe.2020.04.007>
- [24] D. Mukherjee, R. Mondal, P.K. Singh, R. Sarkar, and D. Bhattacharjee, "EnsemConvNet: a deep learning approach for human activity recognition using smartphone sensors for healthcare applications", *Multimedia Tools and Applications*, vol. 79, no. 41, pp. 31663-31690, 2020. <https://doi.org/10.1007/s11042-020-09537-7>
- [25] C. Catal, S. Tufekci, E. Pirmit, and G. Kocabag, "On the use of ensemble of classifiers for accelerometer-based activity recognition", *Applied Soft Computing*, vol. 37, pp. 1018-1022, 2015. <https://doi.org/10.1016/j.asoc.2015.01.025>
- [26] I. Elamvazuthi, L.I. Izhar, and G. Capi, "Classification of human daily activities using ensemble methods based on smartphone inertial sensors", *Sensors*, vol. 18, no. 12, pp. 4132, 2018. <https://doi.org/10.3390/s18124132>
- [27] O. Herrera-Alcántara, A.Y. Barrera-Animas, M. González-Mendoza, and F. Castro-Espinoza, "Monitoring student activities with smartwatches: On the academic performance enhancement", *Sensors*, vol. 19, no. 7, pp. 1605, 2019. <https://doi.org/10.3390/s19071605>
- [28] N. Irvine, C. Nugent, S. Zhang, H. Wang, and W.W. Ng, "Neural network ensembles for sensor-based human activity recognition within smart environments", *Sensors*, vol. 20, no. 1, pp. 216, 2020. <https://doi.org/10.3390/s20010216>
- [29] S. Brajesh, and I. Ray, "Ensemble approach for sensor-based human activity recognition", In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, pp. 296-300, 2020. <https://doi.org/10.1145/3410530.3414352>
- [30] R. Sekiguchi, K. Abe, T. Yokoyama, M. Kumano, and M. Kawakatsu, "Ensemble learning for human activity recognition", In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, pp. 335-339, 2020. <https://doi.org/10.1145/3410530.3414346>
- [31] A. Subasi, D.H. Dammas, R.D. Alghamdi, R.A. Makawi, E.A., Albiety, T. Brahimi, and A. Sarirete, "Sensor based human activity recognition using adaboost ensemble classifier", *procedia computer science*, vol. 140, pp. 104-111, 2018. <https://doi.org/10.1016/j.procs.2018.10.298>
- [32] N. Dwivedi, D.K. Singh, and D.S. Kushwaha, "Orientation invariant skeleton feature (oisf): a new feature for human activity recognition", *Multimedia Tools and Applications*, vol. 79, no. 29, pp. 21037-21072, 2020. <https://doi.org/10.1007/s11042-020-08902-w>
- [33] S. Balli, and E.A. Sağbaş, "The usage of statistical learning methods on wearable devices and a case study: activity recognition on smartwatches" *Advances in statistical methodologies and their application to real problems*, Ed. Tsukasa Hakimoto, 2017, ch. 13, pp. 259-277.
- [34] G.H. John, P. Langley, "Estimating continuous distributions in Bayesian classifiers", In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pp. 338-345, August 1995.

- [35] B.K. Alsberg, R. Goodacre, J.J. Rowland, and D.B. Kell, "Classification of pyrolysis mass spectra by fuzzy multivariate rule induction-comparison with regression, K-nearest neighbour, neural and decision-tree methods", *Analytica Chimica Acta*, vol. 348, no. 1-3, pp. 389-407, 1997. [https://doi.org/10.1016/S0003-2670\(97\)00064-0](https://doi.org/10.1016/S0003-2670(97)00064-0)
- [36] E.A. Sağbaşı, and S. Ballı, "Estimation of Human Activities by Using Wrist Movements", 3. International Management Information Systems Conference, pp. 52-58, October 2016.
- [37] S. Ballı, E.A. Sağbaşı, "Diagnosis of transportation modes on mobile phone using logistic regression classification", *IET Software*, vol. 12, no. 2, pp.142-151, 2018. <https://doi.org/10.1049/iet-sen.2017.0035>
- [38] L. Breiman, "Random forests", *Machine learning*, vol. 45, no. 1, pp. 5-32, 2001. <https://doi.org/10.1023/a:1010933404324>
- [39] S. Ballı, and E.A. Sağbaşı, "Classification of Human Motions with Smartwatch Sensors", *Süleyman Demirel University Journal of Natural and Applied Sciences*, vol. 21, no. 3, pp. 980-990, 2017. <https://doi.org/10.19113/sdufbed.32689>
- [40] L. Rokach, "Ensemble-based classifiers", *Artificial intelligence review*, vol. 33, no. 1, pp. 1-39, 2010. <https://doi.org/10.1007/s10462-009-9124-7>
- [41] L. Breiman, "Bagging predictors", *Machine learning*, vol. 24, no. 2, pp. 123-140, 1996. <https://doi.org/10.1007/BF00058655>
- [42] L.I. Kuncheva, "Combining pattern classifiers: methods and algorithms", John Wiley & Sons, 2014.
- [43] E.A. Sağbaşı, and S. Ballı, "Activity recognition by voting method via motion sensors", In 2017 International Conference on Computer Science and Engineering (UBMK), 2017. <https://doi.org/10.1109/UBMK.2017.8093557>
- [44] E.A. Sağbaşı, and S. Ballı, "Usage of the smartphone sensors and accessing raw sensor data", *Academic Computing Conferences*, pp. 180-186, 2015.
- [45] P.C. Mahalanobis, "On tests and measures of groups divergence", *IJ Asiatic Soc. Bengal*, vol. 26, pp. 541, 1930.
- [46] C. Leys, O. Klein, Y. Dominicy, and C. Ley, "Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance", *Journal of Experimental Social Psychology*, vol. 74, pp. 150-156, 2018. <https://doi.org/10.1016/j.jesp.2017.09.011>
- [47] K. Polat, and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform", *Applied Mathematics and Computation*, vol. 187, no. 2, pp. 1017-1026, 2007. <https://doi.org/10.1016/j.amc.2006.09.022>
- [48] I.H. Witten, and E. Frank, "Data mining: practical machine learning tools and techniques with Java implementations", *Acm Sigmod Record*, vol. 31, no.1, pp. 76-77, 2002.

Optimizing PID Gains of a Vehicle using the State-of-the-art Metaheuristic Methods

¹Mustafa Atakan AFŞAR, ^{*2}Hilal ARSLAN

¹ Department of Aerospace Engineering, Ankara Yıldırım Beyazıt University, Ankara, Türkiye, atakanafsar@gmail.com 

² Department of Software Engineering, Ankara Yıldırım Beyazıt University, Ankara, Türkiye, hilalarlan@aybu.edu.tr 

Abstract

PID controllers are important control methods that are widely used in industrial processes. Proper tuning of PID gains is critical for achieving state-of-the-art system performance. Therefore, the optimization of PID gains is an important research topic in control engineering. In this study, PID controller gains are automatically tuned using metaheuristic optimization methods. These methods use an iterative approach to calculate optimal values of PID controller gains based on different optimization techniques. The interaction between artificial intelligence and control systems requires a multidimensional approach across different disciplines. In the study, we perform Particle Swarm Optimization, Gray Wolf Optimization, Whale Optimization Algorithm, Firefly Algorithm, Harris Hawks Optimization, Artificial Hummingbird Algorithm, and African Vulture Optimization Algorithm to determine PID gains. In the simulation, step input is applied to the dynamic equation of the unmanned free-swimming submersible vehicle. The fitness function is determined for the controller integral square error, settling time value, and maximum percent overshoot value. We also evaluate the optimization time of the selected algorithms based on the fitness function. Experimental results show that the Artificial Hummingbird Algorithm, Gray Wolf Optimization, and Particle Swarm Optimization achieve significant performance. This underlines that using metaheuristic methods in PID gain optimization increases overall system performance.

Keywords: PID gain optimization; metaheuristic methods; artificial intelligence; PID controller design

1. INTRODUCTION

PID (Proportional-Integral-Derivative) controllers in control systems have great acceptance and reputation. Despite the emergence of numerous complex control theories and techniques, control loops generally continue to rely on proportional-integral-derivative (PID) controllers since they have a simple structure, provide easy design as well as implementation, and have a wide range of implementation. A PID controller has mainly a proportional gain (KP), an integral gain (KI), as well as a derivative gain (KD). The main goal of the PID controller is to minimize over a short time, which is computed by taking the difference between a set point and the measured process variable. Hence, determining the optimum value of the three gains (KP, KI, KD) is challenging, especially when the process to control does not have a linear behavior.

Tuning of the first PID controller parameters was achieved by Ziegler and Nichols (ZN) [1]. The main disadvantage of this tuning method is to have a large settling time as well as overshoot. Hence, each gain has to be refined later. Cohen and Coon (C-C) [2] proposed another method that optimizes three gains. The major limitation of this method is that all

computed gains are constant as well as they are not able to be on-line reconfigured.

Various state-of-the-art methods have been proposed to develop a better adjustment mechanism for each optimal gain (KP, KI, KD) of the PID controller. These methods may be grouped as fuzzy control [3], robust control [4], adaptive control [5], and neural network control. With the advent of computer technology, several control design methods based on artificial intelligence have been proposed. The usage of artificial intelligence methods provides a quick and accurate method for tuning parameters to reach precise control.

Recently, various types of metaheuristic methods have been proposed to parameter tuning of PID controllers. In this study, we perform Particle Swarm Optimization (PSO), Grey Wolf Optimization (GWO), Whale Optimization Algorithm (WOA), Firefly Algorithm (FA), Harris Hawks Optimization (HHO), Artificial Hummingbird Algorithm (AHA) and African Vultures Optimization Algorithm (AVOA). The other sections are organized as follows: In section 2, related works about metaheuristic methods are given. In section 3, details of the proposed methods are given. Section 4 includes results and discussion. Section 5 concludes the study.

2. RELATED WORKS

Kim and Cho [6] used genetic algorithm (GA), PSO, and BFO methods for tuning the PID gains in the automatic voltage regulator system; and also implemented hybrid GA-PSO and BF-GA methods. The BF-GA hybrid algorithms gave the best results in their system. According to the simulation results, hybrid applications have a better response than GA and PSO applications. Srinivas et al. [7] compared traditional PID tuning methods with the GA method. According to the results of their study, GA was faster and had low overshoot compared to traditional methods such as the Z-N method, C-C method, and minimum error integral criteria method. El-Deen et al. [8] and Bassi and Dada [9] tuned the PID gains for DC motor speed control using GA and controlled the system. In their studies, GA was compared with the classical method Z-N and Active Set Optimization Algorithm (ASOA). In both studies, the GA method produced a better solution to control system dynamics. Emmanuel and Inyama [10] tried to apply Z-N, Fuzzy Logic (FL), GA and Artificial Neural Network (ANN) methods in controller design for a robot manipulator. According to their studies, FL, GA and ANN methods are more suitable to obtain optimum system performance. Although these methods are more complex and slower to model than classical models, modern control methods have high accuracy.

Sandoval et al. [11] tuned PID gains with ACO to control the robotic arm and studied as experimentally. In their study, the Z-N method was compared with manual adjustment and ACO. The ACO was better than other methods. In addition, the system response improved as the number of ant was increased.

El-Telbany [12] compared the ABC, PSO, GA and ZN methods for DC motor controller optimization, and the best rise time and settling time for steady-state response were obtained with the ABC method. Liao et al. [13] used ABC to optimize DC motor PID gains. It has been stated that this method is effective for many optimization problems because it is easy to implement. Senberber and Bagis [14] stated that the application of the Fractional Order PID and ABC method greatly increased the performance of the system. Annisa et al. [15] tried to reduce the vibration at the endpoint by using a PID controller for a flexible robotic arm as a hybrid with PSO and ABC methods. According to the simulation results, it has been revealed that the PSO method was better than the ABC method. Zhi [16] applied the ABC, PSO, GA and ZN methods to determine the PID gains of the single-phase inverter system. In their study, the ABC method performed significantly better than other applications.

Kotteeswaran and Sivakumar [17] applied the BA method for tuning gains of the PI controller in the coal gasifier system. The controller controlled a complex system that was a non-linear multidimensional process. With the current controller, the system did not meet the performance criteria when sinusoidal pressure disturbance was applied while the system was at 0% load. With the optimum gain values determined by the BA, the system exhibited the desired performance under 0%, 50% and 100% load. Katal et al. [18] optimized PID gains with BA for the liquid-level control

system. The BA presented better results than classical methods. Singh et al. [19] compared BA and PSO algorithms for servo motor PID gain optimization. The integral of the time square error was used as a stopping criterion in the study. According to the simulation results, the BA presented better rise time and settling time than the PSO application. Premkumar and Manikandan [20] used PSO, CS, and BA algorithms to determine gains of Fuzzy PD and Fuzzy PID controllers for brushless DC motor speed control applications. The BA method achieved faster results in Fuzzy PD optimization compared to other methods.

Bayoumi and Salem [21] applied the BFO method to design a robust PID controller. As a fitness function, they optimized percent overshoot, settling time, rise time, and steady-state error. The BFO-PID application showed better voltage regulation than the classical Z-N application. Benbouabdallah and Zhu [22] automatically tuned the PID gains of the mobile robot moving in an unknown route with BFO. The BFO and PSO applications were used as a hybrid to obtain optimum gains. According to the simulation results, the hybrid BFO-PSO application had better results than BFO, PSO and GA applications. Sivakumar et al. [23] used BFO to determine the PID gains in a MIMO system. Three different cost functions were compared. According to the simulation results, these functions presented very similar performances. Agarwal et al. [24] tried to minimize the maximum overshoot, peak time, settling time, and rise time with an integral absolute error of the system by applying BFO in their study. Compared with the classical Z-N method, the BFO method determined more optimum gain values. Jasim [25] tuned DC servo motor PID gains with BFO. Similar to other studies, fitness function was tried to be minimized. According to the simulation results, the high performance of the BFO method was emphasized in the study.

The PSO method is used quite frequently in control applications. Chang and Chen [26] optimized the PID gains for a MIMO system using PSO. An improved PSO method compared with different machine learning algorithms performed better than other algorithms. Rajesh and Ananda [27] optimized PID gains with the PSO method for the control of the camera moving in 2 axes and located on a UAV. The PSO algorithm was compared with classical Z-N and C-C applications. The PSO performed better than other applications. Lodhi and Saraf [28] tuned PID gains using the Real-valued Genetic Algorithm (RGA), GA, and PSO methods. According to the simulation results, it was stated in the study that PSO was better than other algorithms. Fister et al. [29] used two different evolutionary algorithms which are Differential Evolution (DE) and GA for determining PID gains on a SCARA-type robot with 2 degrees of freedom; and four different swarm intelligence-based algorithms, namely BA, Hybrid Bat Algorithm (HBA), PSO and CS, was applied experimentally. In their study, algorithms were compared according to many parameters. The HBA, BA, DE, PSO, CS, and GA operating under the same conditions produced results in 1.87, 1.93, 2.30, 2.64, 3.03, and 21.13 milliseconds from fast to slow, respectively. However, when the results in the robot axes were examined, the method with the least standard deviation compared to the average was obtained as PSO. According to these average results, HBA

presented the worst result. Joseph and Dada [30] automatically tuned PID gains using PSO for the control of the inverted pendulum system. The PSO showed better system performance than conventional methods.

Kumar et al. [31] used the CS method in the PID controller design of nonlinear systems such as ship roll dynamics, oscillators and inverted pendulums. Integral square error, integral time square error and integral absolute error criteria were used as fitness functions. It was compared with the system designed with PSO by another researcher. According to the simulation results, the CS performed better than the PSO application. Gholap et al. [32] applied PSO, GA and SA to obtain optimum PID gains. They used two fitness functions, integral time absolute error and time domain specifications in their work. Different types of algorithms were compared in the study, taking into account parameters such as the number of iterations and cost. According to the results, although all three applications can control the system, the PSO and SA applications showed better results than GA applications. Bingul and Karahan [33] tuned the gains of the PID controller in the automatic voltage regulator system with CS. Performance criteria were determined as in the study of Bayoumi and Salem. The CS method proposed in the study performed better than the systems designed with PSO, ABC and BF-GA methods.

Bansal et al. [34] used the Multi-Objectives SA application for tuning PID gains. Single-Objective SA and classical Z-N methods were also applied in the study. The multi-objective SA application performed better than other methods. Vijay and Banu [35] optimized gains in attitude control with PID by the SA method. The fuel consumption of the thruster actuator in the attitude control system has been optimized. Lahcene et al. [36] used SA to optimally select the gains of the fractional order PID controller that controls the automatic voltage regulator system. Similar to other studies, the cost function was tried to be minimized. SA application performed better than Multi-Objective External Optimization and PSO methods. Shatnawi and Bayoumi [37] determined gains of PI and PID controllers designed for permanent magnet BLDC motors by SA methods. Maximum overshoot, rise time and settling time were optimized simultaneously. In their study, the SA applications performed better than PSO and classical Z-N applications.

Şen and Kalyoncu [38] tuned the gains of the PID controller for a quadruped robot system using GWO. For the same system, the GWO method performed faster and better than PSO and GA methods. Agarwal et al. [39] optimized gains of the fractional order PID controller for DC motor speed control with the GWO method. When the system response was compared, the GWO method was good and robust compared to Invasive Weed Optimization, Stochastic Fractal Search, and PSO methods. Yadav et al. [40] optimized PID control gains in the nonlinear inverted pendulum system with GWO. The integral square error was used as a cost function in the study. Sule et al. [41] determined the gains of the PI controller controlling the wind turbine system with the GWO method and compared them with the PSO and GA methods. For their system, the GWO method worked better and achieved faster results. Verma and Devarapalli [42] tuned gains of the fractional order PID controller that controls the

automatic voltage regulator system using different types of GWO. The modified GWO method suggested by the researchers presented better performance than standard GWO methods.

Hekimoglu et al. [43] optimized PID gains in the DC-DC buck converter system using WOA and compared them with the GA method. According to the simulation results, the WOA method achieved better results. Kumar and Kumar [44] optimized PID gains controlling the drilling machine system with WOA and compared them with PSO and classical Z-N methods. The WOA method was better than other methods when compared to the settling time and rise time values. Mosaad et al. [45] optimized PID gains in the automatic voltage regulator system with the WOA method. In their study, PID gains were optimized with many evolutionary algorithms (8 different metaheuristic methods) and compared with WOA. The WOA was designed as a better controller than other applications. Loucif et al. [46] optimized gains in the PID controller of the robot arm with 2 degrees of freedom with WOA and compared them with the PSO and GWO methods. According to the performance criteria in the study, the WOA was the most effective method.

Bendjeghaba et al. [47] proposed an algorithm for tuning PID controller parameters using the firefly algorithm. They compared their results with Ziegler-Nichols method and achieved efficient results. Coelho and Mariani [48] applied the firefly algorithm for tuning the PID controller. They introduced a chaotic firefly algorithm based on the Tinkerbell map. Their experimental results presented that their method tuned multi-loop PID controllers. Finally, Ekinci et al. [49] presented an application of the Harris Hawks optimization algorithm to find the optimal parameters of the PID controller.

3. METHODS

3.1. Dynamics of the System

An Unmanned Free-Swimming Submersible (UFSS) vehicle changes its depth with the help of an elevator which is driven by a motor while moving forward. The moment occurs about the pitch axis of UFSS with the change of the elevator angle. The depth of the UFSS can be changed concerning the desired reference angle by controlling the moment of this axis. It is possible to develop an autonomous system by controlling the moment of the pitch axis with the help of a controller.

In this study, the UFSS is selected as a vehicle since its transfer function is well-known in literature and easy-to-design a controller relative to the multi degree of freedom systems. The transfer function can be found in the Control System Engineering book released by Norman S. Nise [50]. The block diagram of the system is shown in Figure 1.

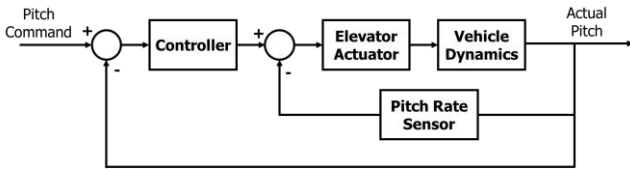


Figure 1. Closed-loop diagram of UFSS vehicle

The overall Transfer Function (OTF) in complex frequency domain, which expresses the actuator, vehicle dynamics and pitch rate sensor in the system, is shown in Equation 1.

$$OTF = \frac{-0.25s - 0.1087}{s^4 + 3.456s^3 + 3.457s^2 + 0.7193s + 0.04157} \quad (1)$$

The coefficients describing the system dynamics in OTF are derived by considering all system elements. System elements and detailed block diagram of the system can be found in [50]. The transfer function of a real system is obtained through its mathematical model that takes into account the dynamics of the system.

3.2. Controller Design

In control applications, the main purpose is to reduce the error to zero. In this study, the PID controller is used to make the error closer to zero. The PID controller is frequently used in many simulation and experimental studies such as DC motor, Quadrotors and Inverted Pendulum. It is possible to apply a PID controller for UFSS vehicle. The equation of the PID controller is shown in Equation 2 [51].

$$u(t) = K_p e(t) + K_I \int_0^t e(t) dt + K_D \frac{de(t)}{dt} \quad (2)$$

Table 1. Effect of increasing PID gains on system behavior

Close Loop Response	Rise Time	Overshoot	Settling Time	Steady-state Error
K_p	Decrease	Increase	Small Change	Decrease
K_I	Decrease	Increase	Increase	Decrease
K_D	Small Change	Decrease	Decrease	No Change

The proportional term adjusts control output proportionally depending on the size of the error. The integral term calculates the time integral of the error and adjusts the control output accordingly to prevent the control output from increasing with the accumulation of error over time. The derivative term allows control output to be changed quickly by calculating the rate of error. The effect of K_p , K_I and K_D gains on system behavior is shown in Table 1 [51].

3.3. Metaheuristic Algorithms

In this study, we apply the state-of-the-art metaheuristic methods, PSO, GWO, WOA, Firefly Algorithm (FA), Harris Hawks Optimization (HHO), Artificial Hummingbird Algorithm (AHA) and African Vultures Optimization Algorithm (AVOA) for optimizing PID gains. In the following, we summarize these methods.

In the PSO, new positions of particles are calculated at each iteration. This calculation is performed using the current

position and best position of each particle. Each particle is repositioned at a new position determined by its velocity and the difference of its best position. This process continues until the particles reach a position that minimizes or maximizes their fitness function. Detailed steps of the PSO algorithm are given in Algorithm 1 [52].

Algorithm 1. Particle Swarm Optimization

```

1: Initialize Population: Positions and velocities of particles initialized as randomly
2: Define fitness function to evaluate each particle's fitness
3: Set initial personal best (pbest) for each particle to its current position
4: Set initial global best (gbest) to the position of the particle with the best fitness value
5: while termination criterion is not met do
6:   for each particle i do
7:     Update particle velocity using the following formula:
8:      $v_i = w * v_i + c1 * rand() * (pbest_i - x_i) + c2 * rand() * (gbest - x_i)$ 
9:     Update particle position using the following formula:
10:     $x_i = x_i + v_i$ 
11:    Evaluate fitness value of the new position
12:    if the new position is better than its personal best then
13:      Update personal best:  $pbest_i = x_i$ 
14:    end if
15:    if the new position is better than the global best then
16:      Update global best:  $gbest = x_i$ 
17:    end if
18:  end for
19: end while
20: return gbest

```

Algorithm 2. Gray Wolf Optimization

```

1: Initialize a population of gray wolves randomly
2: Evaluate fitness function to determine the fitness value of each wolf
3: Set alpha, beta, and delta as the three best wolves in the population
4: while termination criterion is not met do
5:   for each wolf i in the population do
6:     Calculate the distance between the current wolf i and alpha, beta, and delta
7:     Update the position of wolf i using the following formula:
8:      $x_i = \frac{\alpha + 2 * \beta + 2 * \delta - 5 * x_i}{4}$ 
9:     Apply a randomization operator to the new position of wolf i
10:    Evaluate the fitness value of the new position of wolf i
11:    if the new position is better than the position of alpha, beta, or delta
12:      then
13:        if the new position is better than alpha then
14:          Set the new position as alpha
15:        else if the new position is better than beta then
16:          Set the new position as beta
17:        else
18:          Set the new position as delta
19:        end if
20:      end if
21:    end for
22:  end while
23: return alpha

```

Algorithm 3. Whale Optimization Algorithm

```

1: Initialize a population of whales randomly
2: Evaluate fitness function to determine the fitness value of each whale
3: Set the current best position as the position of the whale with the best fitness value
4: while termination criterion is not met do
5:   for each whale  $i$  in the population do
6:     Generate a random vector  $A$  and  $C$ 
7:     if  $|A| < 0.5$  then
8:       if  $|C| < 1$  then
9:         Update the position of whale  $i$  towards the current best position using the following formula:
10:         $x_i = x_{best} - A \times |C \times x_{best} - x_i|$ 
11:       else
12:         Update the position of whale  $i$  randomly within the search space
13:       end if
14:     else
15:       Update the position of whale  $i$  towards a randomly selected whale  $j$  using the following formula:
16:        $x_i = x_j - A \times |C \times x_j - x_i|$ 
17:     end if
18:     Apply a randomization operator to the new position of whale  $i$ 
19:     Evaluate the fitness value of the new position of whale  $i$ 
20:     if the new position is better than the current best position then
21:       Set the new position as the current best position
22:     end if
23:   end for
24: end while
25: return current best position

```

Algorithm 4. Firefly Algorithm

```

1: Initialize a population of fireflies randomly
2: Evaluate the fitness of each firefly
3: Set the best solution as the current global best
4: while stopping criterion not met do
5:   for each firefly  $i$  do
6:     for each firefly  $j$  do
7:       if firefly  $j$  is brighter than firefly  $i$  then
8:         Compute the distance  $r_{ij}$  between fireflies  $i$  and  $j$ 
9:         Compute the attractiveness  $\beta(r_{ij})$  of firefly  $j$  towards firefly  $i$ 
10:        Move firefly  $i$  towards firefly  $j$  with a step size  $\alpha\beta(r_{ij})$ 
11:       end if
12:     end for
13:     Evaluate the fitness of the new position of firefly  $i$ 
14:     if the new position of firefly  $i$  is better than its current position then
15:       Set the new position as the current position of firefly  $i$ 
16:     end if
17:   end for
18:   Update the global best solution
19: end while
20: return the best solution found

```

Algorithm 5. Harris Hawks Optimization

```

1: Initialize a population of hawks randomly
2: Evaluate the fitness of each hawk
3: Set the best solution as the current global best
4: while stopping criterion not met do
5:   Sort the hawks in descending order of fitness
6:   Compute the position of the top predator hawk
7:   Compute the position of the other predator hawks
8:   for each prey hawk do
9:     Compute the position of the new prey hawk
10:    Evaluate the fitness of the new prey hawk
11:    if the new prey hawk is better than the current prey hawk then
12:      Replace the current prey hawk with the new prey hawk
13:    end if
14:  end for
15:  Update the global best solution
16: end while
17: return the best solution found

```

Algorithm 6. Artificial Hummingbird Algorithm

```

1: Initialize the population of hummingbirds with random positions and velocities
2: Evaluate the fitness of each hummingbird
3: while stopping criterion not met do
4:   for each hummingbird do
5:     Calculate the attraction towards the best hummingbird in the neighborhood
6:     Update the velocity of the hummingbird
7:     Update the position of the hummingbird
8:     Evaluate the fitness of the new position
9:     if new position is better than the current position then
10:      Replace the current position with the new position
11:     end if
12:   end for
13: end while
14: return the best hummingbird solution found

```

Algorithm 7. African Vultures Optimization Algorithm

```

1: initialize vulture population and their positions
2: initialize search radius and iteration counter
3: while not converged do
4:   for each vulture do
5:     update vulture's position based on its personal best and global best positions
6:     if vulture is outside the search radius then
7:       move it back inside
8:     end if
9:     if vulture is better than the current global best then
10:      update the global best position
11:     end if
12:   end for
13:   update the search radius based on the number of iterations
14:   update the iteration counter
15: end while

```

The GWO is based on three basic behaviors: hunting, sorting and mating. The wolf pack interacts with each other to optimize these behaviors. Hunting behavior involves exploring the search area to reach the best location. Sorting behavior allows wolves to adjust their position and strength against each other. Mating behavior, on the other hand, allows the creation of new wolves and the enlargement of the solution area. The Pseudocode of GWO is presented in Algorithm 2 [53].

The WOA works by maintaining a population of candidate solutions (whales) and improving them iteratively over several generations. At each iteration, each whale's position is updated using two random vectors, A and C , and a specific formula that simulates the hunting behavior of humpback whales. Some randomness is then introduced into the search process by applying the random join operator to each whale's new position. The fitness value of each whale's new position is calculated and compared against the best available position. If the new position is better than the current best position, it is assigned as the new current best position. This process continues until a certain termination criterion is met. The final solution is the best available position. Detailed steps of WOA are presented in Algorithm 3 [54].

In the FA, the difference in light intensity between fireflies is related to distance and the intensity of the environment. This difference determines each firefly's movement. The main purpose of the algorithm is to find the position where the difference in light intensity between many fireflies is the least. This location is a near-optimal solution. Using light intensities as motion vectors, the algorithm creates a swarm of fireflies that move through the solution range. Each firefly moves towards the light intensity value of the closest firefly, thus bringing the flock closer to the optimum solution. The Pseudocode of the FA is shown in Algorithm 4 [55].

The HHO is led by the strongest hawk chosen as the pack leader. Other hawks follow the leader's position and speed and move towards the positions the leader finds. Hawks also cooperate with each other to hunt non-leader hawks. This algorithm is particularly effective for multidimensional and complex optimization problems. An overview of the HHO is shown in Algorithm 5 [56].

The AHA is an optimization algorithm inspired by the interactions of flowers and hummingbirds in nature. The algorithm works by selecting the most suitable flowers using information such as the location of the flowers, the amount of nectar, and the taste value. The AHA can show high performance, especially in complex problems. The pseudocode of AHA is shown in Algorithm 6 [57].

The AVOA mimics the social interactions of vultures within the population and uses the leader-follower hierarchy. Initially, the fitness of each vulture is calculated. The population with the best fitness value is selected as the leader. Follower vultures keep track of the leader's position and update their position according to the leader's movement. Detailed steps of the AVOA are shown in Algorithm 7 [58].

Metaheuristic algorithms offer a more flexible and robust approach by exploring a larger solution space and considering multiple objective criteria than deterministic methods such as Ziegler-Nichols and Cohen-Cohen. They can handle nonlinear systems, take into account constraints, and find near-optimal solutions that outperform deterministic methods in terms of stability, performance, and robustness. Furthermore, they are particularly useful when dealing with complex control systems where the relationship between gains and system behavior is not well defined or where multiple conflicting objectives must be considered.

3.3.1. Input Parameters

The input parameters required to optimize the PID gains in the control problem are shown in Table 2. The number of search agents and the maximum number of iterations are set to 20 and 30, respectively. Solution quality may be enhanced by increasing these values. However, this increases the optimization time, and it has been observed that the effect on the result is quite low.

Lower boundary and upper boundary values are set to 0 and 100, respectively. It should be considered that these values may need to be adjusted for different control problems.

Table 2. Input parameters for optimization

Parameter	Value
Number of search agents	20
Maximum number of iterations	30
Lower boundary	[0 0 0]
Upper boundary	[100 100 100]
Problem Dimension	3

3.3.2. Fitness Function

The fitness function is used to evaluate the performance in the control problem, specifically for optimizing PID gains. In this study, the fitness function is calculated based on three metrics that are integral square error (ISE), overshoot (OS) and settling time (T_s). The formulation of the fitness function used in this study is given in Equation 3.

$$FitnessFunc = (J_1 \cdot ISE) + (J_2 \cdot \%OS) + (J_3 \cdot T_s) \quad (3)$$

where $J_1 = 5 \text{ rad}^{-2} \cdot \text{sec}^{-1}$, $J_2 = 10$ and $J_3 = 5 \text{ sec}^{-1}$. These values make the FitnessFunc variable unitless. Weights of ISE, OS and T_s are set by trial and error to get the step response more realistic. The reason why the OS value is larger than the ISE and T_s values is to prevent the system from large control signals. ISE and T_s weights are of equal importance for this problem. If the weights of these values are increased, oscillation occurs initially. This oscillation can cause instability in real system applications.

Input signals such as step, ramped, impulse, sinusoidal, parabolic etc. are applied to measure the performance of the controller. In this study, step input is applied to the system. ISE integrates the square of the error amount between the reference value (unit step = 1 rad) and the actual response numerically over the time interval of 0.1 sec. Thus, the overall error of the system is measured.

Overshoot is calculated by the difference between the reference value and the maximum response value. An overshoot penalty is applied if the maximum overshoot exceeds 1% threshold.

Settling time is measured according to 2% tolerance band. The response of the system is assigned to time vector by using loop which checks the specified tolerance band. If the settling time value is greater than 10 second threshold, the settling time penalty is applied.

4. RESULTS AND DISCUSSION

In this study, the depth of the UFSS vehicle is controlled by changing the elevator angle. The gain values of the PID controller used in the system are optimized with the PSO, GWO, WOA, FA, HHA, AHA, and AVOA metaheuristic methods. These algorithms are compared with respect to optimization time, error value, settling time value and % overshoot metrics.

The total optimization time of the algorithms performed in this study is shown in Figure 2. The time is determined based on the value of the fitness function. The fitness values achieved by the PSO, GWO, WOA, FA, HHO, AHA, and AVOA are 1.918, 1.912, 2.328, 1.911, 2.388, 1.924 and 2.012, respectively. As you can see in Figure 2, the FA requires much time when compared to the other methods and has a better cost value. The PSO, GWO, AHA, and AVOA converge in a short time with acceptable fitness values.

Error values in the proposed system where step input is applied are shown in Figure 3. All algorithms almost reach

the desired reference value. Error values of the AHA, PSO and FA are relatively low compared to other algorithms. Absolute error values of the PSO, GWO, WOA, FA, HHO, AHA, and AVOA are 0.00299 rad, 0.00381 rad, 0.05141 rad, 0.00306 rad, 0.00622 rad, 0.00237 rad, and 0.00535 rad, respectively. The error value of WOA is higher than other algorithms. However, it should be noted that all algorithms have fairly good error values for this control problem. For example, HHO has an error of 0.00622 radians ($\approx 0.36^\circ$) for 1 radian ($\approx 57.3^\circ$) input.

The settling time values, which are an important criterion for the evaluation of control problems, are compared in Figure 4. Settling time values of PSO, GWO, WOA, FA, HHO, AHA, and AVOA are obtained as 7.433 sec, 8.142 sec, 3.731 sec, 7.536 sec, 22.204 sec, 7.126 sec, and 25.924 sec, respectively. The WOA achieves the best performance when we focus on the settling time. The performances of the PSO, GWO, FA, and AHA are close to each other. On the other hand, HHO and AVOA methods require a larger time.

Another important benchmark for control problems is the percentage overshoot value that is presented in Figure 5. Percentage overshoot values of PSO, GWO, WOA, FA, HHO, AHA, and AVOA are calculated as 1.121, 1.022, 6.520, 1.029, 6.978, 0.747, and 2.414, respectively. While high overshoot may be appropriate in some systems, low overshoot may be appropriate in some systems. Overshoot evaluation is problem specific. In this study, a low overshoot value would be more appropriate. The performances of the AHA, GWO, and FA are close to each other and achieve the best results. Percentage overshoot values of WOA and HHO are greater than other algorithms.

Integral square error values of PSO, GWO, WOA, FA, HHO, AHA, and AVOA are calculated as 0.453 $\text{rad}^2 \cdot \text{sec}$, 0.382 $\text{rad}^2 \cdot \text{sec}$, 0.376 $\text{rad}^2 \cdot \text{sec}$, 0.382 $\text{rad}^2 \cdot \text{sec}$, 0.392 $\text{rad}^2 \cdot \text{sec}$, 0.374 $\text{rad}^2 \cdot \text{sec}$, and 0.384 $\text{rad}^2 \cdot \text{sec}$. AHA, WOA and GWO methods have better ISE values than other algorithms. ISE values are shown in Figure 6.

AHA, GWO and PSO have better performance than other algorithms according to simulation time, best cost value, percentage error, settling time and percentage overshoot. AHA, GWO and PSO have almost same performance.

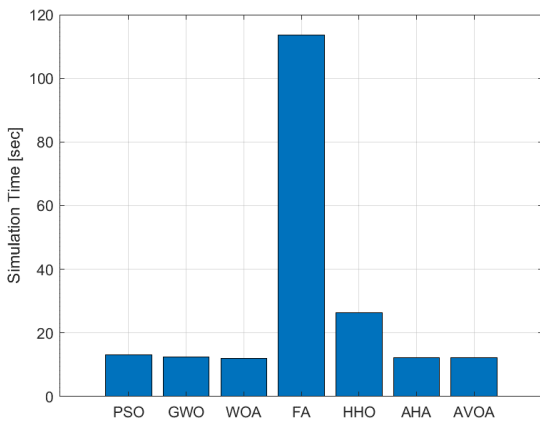


Figure 2. Total optimization time

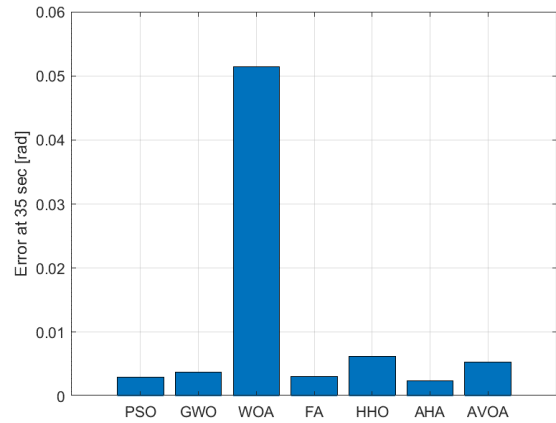


Figure 3. Error values at the end of simulation

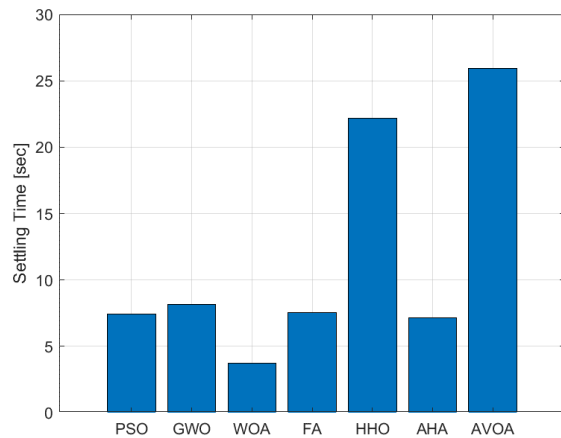


Figure 4. Settling time of step input

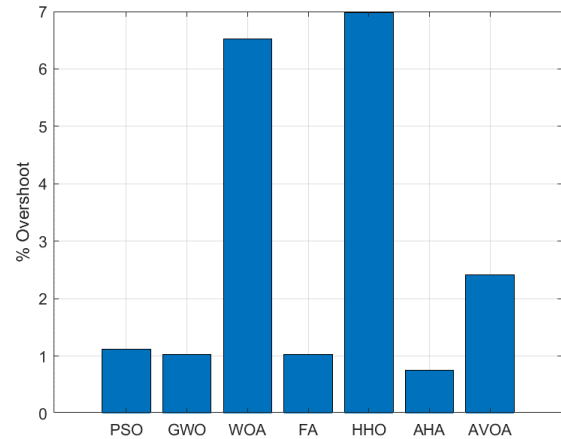


Figure 5. Percentage overshoot of step input

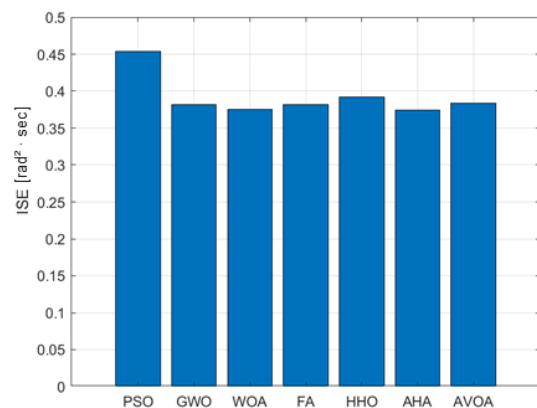


Figure 6. ISE values of all algorithms

Table 3. Optimized PID gains

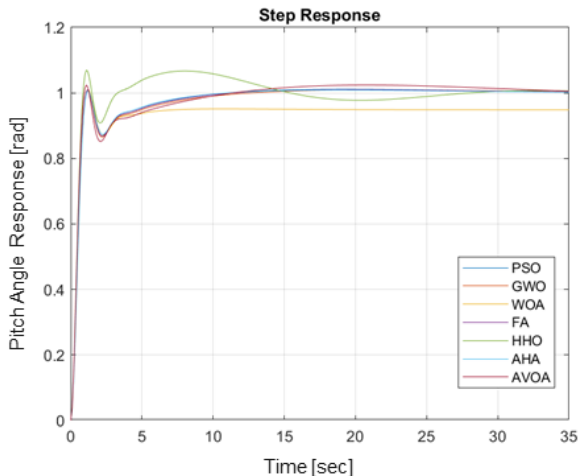
Algorithm	K _P	K _I	K _D
PSO	6.026	0.551	31.180
GWO	5.862	0.505	31.551
WOA	7.051	0.000	29.674
FA	6.058	0.535	31.130
HHO	7.960	2.746	35.660
AHA	6.311	0.532	30.176
AVOA	4.763	0.634	35.348

Table 4. Results of metaheuristic algorithms

MN	RT (sec)	ST (sec)	OS (rad)	PE (rad)	PT (sec)	BCV	ISE (rad ² ·sec)	SIT (sec)
PSO	0.66	7.43	1.12	1.01	19.29	1.92	0.45	13.20
GWO	0.66	8.14	1.02	1.01	1.21	1.91	0.38	12.39
WOA	0.63	3.73	6.52	1.01	1.23	2.33	0.38	12.05
FA	0.66	7.54	1.03	1.01	1.21	1.91	0.38	113.59
HHO	0.57	22.20	6.98	1.07	1.09	2.39	0.39	26.42
AHA	0.68	7.13	0.75	1.01	1.21	1.92	0.37	12.34
AVOA	0.60	25.92	2.41	1.02	20.76	2.01	0.38	12.17

MN: Method Name, RT: Rise Time, ST: Settling Time, OS: Overshoot, PE: Peak, PT: Peak Time, BCV: Best Cost Value, ISE: Integral Square Error, SIT: Simulation Time

PID gains obtained by algorithms are shown in Table 3. Detailed results of the metaheuristic methods are also shown in Table 4. Furthermore, step responses are shown in Figure 7 based on optimized gains.

**Figure 7.** Step responses of all algorithms

Since this study is the application of metaheuristic methods in the field of control, control gains have been optimized by considering only the system dynamics. However, in real systems, the control effort should also be included into the fitness function. This ensures that the control signals sent to the actuator stay within acceptable limits. The control signal changes the angle of the elevator. This change is controlled by the torque produced by a motor. Both upper and lower limits need to be specified by the user, similar to setting the values for number of search agents, maximum number of iterations, and problem dimension. When selecting these limits, considering the control effort could be beneficial.

5. CONCLUSION

In this study, the gain values of the controller (PID) used in a control problem are optimized by new generation metaheuristics that are the PSO, GWO, WOA, FA, HHA, AHA and AVOA methods. To achieve a fair comparison, the algorithms are compared under the same conditions. A special fitness function is defined considering the optimization time, controller error, step input settling time and step input overshoot value in this study. Experimental results present that the AHA, GWO and PSO perform better than other algorithms. On the other hand, FA, HHO, and AVOA present lower performance.

The methods compared in the study are acceptable for PID gain optimization. Researchers can optimize PID gain values based on the fitness function they will determine in different control problems. The PSO, which is frequently used in other control problem studies in the literature, performed quite well in this study as well. As a result, the AHA, GWO and PSO can be preferred for PID gain optimization.

In future studies, new generation metaheuristics algorithms will be compared by optimizing PID gain values according to different fitness functions. Thus, the impact of the metrics can be compared as well. In addition, the optimized gain values can be used with the Fuzzy Logic method to design a better controller. Thus, the controller can produce a better response according to different reference signals. Furthermore, we will plan to implement the state-of-the-art metaheuristics by combining with traditional methods in parallel to save CPU time.

REFERENCES

- [1] Ziegler JG, Nichols NB. Optimum Settings for Automatic Controllers. *J Dyn Syst Meas Control* 1993; 115:220–2. <https://doi.org/10.1115/1.2899060>.
- [2] Cohen GH, Coon GA. Theoretical Consideration of Retarded Control. *Trans Am Soc Mech Eng* 1953; 75:827–34. <https://doi.org/10.1115/1.4015451>.
- [3] Fereidouni, A, Masoum, M. A. S., & Moghbel, M. (2015). A new adaptive configuration of PID type fuzzy logic controller. In *ISA Transactions* (Vol. 56, pp. 222–240). Elsevier BV. <https://doi.org/10.1016/j.isatra.2014.11.010>
- [4] Ali, R., Mohamed, T. H., Qudaih, Y. S., & Mitani, Y. (2014). A new load frequency control approach in an isolated small power systems using coefficient diagram method. In *International Journal of Electrical Power & Energy Systems* (Vol. 56, pp. 110–116). Elsevier BV. <https://doi.org/10.1016/j.ijepes.2013.11.002>
- [5] Mittal, A., Kapoor, A., & Saxena, T. K. (2013). Adaptive tuning of PID controller for a nonlinear constant temperature water bath under set-point disturbances using GANFC. *J. Auto Syst. Eng*, 7, 143–63.

- [6] Kim DH, Cho JH. A Biologically Inspired Intelligent PID Controller Tuning for AVR Systems. *Int J Control Autom Syst* 2006; 4:624–36.
- [7] Srinivas P, Vijaya Lakshmi K, Kumar VN. A Comparison of PID Controller Tuning Methods for Three Tank Level Process. *Int J Adv Res Electr Electron Instrum Eng* 2007;3.
- [8] El-Deen AT, Mahmoud AAH, El-Sawi AR, El-Deen AT, Mahmoud AAH, El-Sawi AR. Optimal PID Tuning for DC Motor Speed Controller Based on Genetic Algorithm. *Int Rev Autom Control* 2015; 8:80–5. <https://doi.org/10.15866/IREACO.V8I1.4839>.
- [9] Bassi JS, Dada EG. Automatic Tuning of Proportional–Integral–Derivative Controller using Genetic Algorithm. *Pacific J Sci Technol* 2018; 19:51–7.
- [10] Emmanuel AC, Inyiyama H. A SURVEY OF CONTROLLER DESIGN METHODS FOR A ROBOT MANIPULATOR IN HARSH ENVIRONMENTS. *Eur J Eng Technol* 2015; 3:64–73.
- [11] Sandoval D, Soto I, Adasme P. Control of direct current motor using Ant Colony optimization. *CHILECON 2015 - 2015 IEEE Chil. Conf. Electr. Electron. Eng. Inf. Commun. Technol. Proc. IEEE Chilecon 2015*, Institute of Electrical and Electronics Engineers Inc.; 2016, p. 79–82. <https://doi.org/10.1109/CHILECON.2015.7400356>
- [12] E.El-Telbany M. Tuning PID Controller for DC Motor: An Artificial Bees Optimization Approach. *Int J Comput Appl* 2013; 77:18–21. <https://doi.org/10.5120/13559-1341>.
- [13] Liao W, Hu Y, Wang H. Optimization of PID control for DC motor based on artificial bee colony algorithm. *Int. Conf. Adv. Mechatron. Syst. ICAMechS, IEEE Computer Society*; 2014, p. 23–7. <https://doi.org/10.1109/ICAMECHS.2014.6911617>.
- [14] Senberber H, Bagis A. Fractional PID controller design for fractional order systems using ABC algorithm. *Proc. 21st Int. Conf. Electron., Institute of Electrical and Electronics Engineers Inc.*; 2017. <https://doi.org/10.1109/ELECTRONICS.2017.7995218>.
- [15] Annisa J, Mat Darus IZ, Tokhi MO, Mohamaddan S. Implementation of PID Based Controller Tuned by Evolutionary Algorithm for Double Link Flexible Robotic Manipulator. *2018 Int. Conf. Comput. Approach Smart Syst. Des. Appl. ICASSDA 2018*, Institute of Electrical and Electronics Engineers Inc.; 2018. <https://doi.org/10.1109/ICASSDA.2018.8477615>.
- [16] Zhi D. Optimization of PID Controller for Single Phase Inverter Based on ABC. *2019 5th Int. Conf. Control. Autom. Robot. ICCAR 2019*, Institute of Electrical and Electronics Engineers Inc.; 2019, p. 501–4. <https://doi.org/10.1109/ICCAR.2019.8813361>.
- [17] Kotteeswaran R, Sivakumar L. Optimal partial-retuning of decentralised PI controller of coal gasifier using bat algorithm. *Int. Conf. Swarm, Evol. Memetic Comput., Springer*; 2013, p. 750–61.
- [18] Katal N, Kumar P, Narayan S. Optimal PID controller for coupled-tank liquid-level control system using bat algorithm. *Int. Conf. Power, Control Embed. Syst. ICPCES 2014*, Institute of Electrical and Electronics Engineers Inc.; 2014. <https://doi.org/10.1109/ICPCES.2014.7062818>.
- [19] Singh K, Vasant P, Elamvazuthi I, Kannan R. PID Tuning of Servo Motor Using Bat Algorithm. *Procedia Comput Sci* 2015; 60:1798–808. <https://doi.org/10.1016/J.PROCS.2015.08.290>.
- [20] Premkumar K, Manikandan B V. Bat algorithm optimized fuzzy PD based speed controller for brushless direct current motor. *Eng Sci Technol an Int J* 2016; 19:818–40. <https://doi.org/10.1016/J.JESTCH.2015.11.004>.
- [21] Bayoumi EH., Salem F. PID controller for series-parallel resonant converters using bacterial foraging optimization. *Electromotion Sci J* 2012; 19:64–79.
- [22] Benbouabdallah K, Zhu QD. Bacterial foraging oriented by particle swarm optimization of a Lyapunov-based controller for mobile robot target tracking. *Proc. - Int. Conf. Nat. Comput., IEEE Computer Society*; 2013, p. 506–11. <https://doi.org/10.1109/ICNC.2013.6818029>.
- [23] Sivakumar R, Deepa P, Sankaran D. A Study on BFO Algorithm based PID Controller Design for MIMO Process using Various Cost Functions. *Indian J Sci Technol* 2016; 9:1–6. <https://doi.org/10.17485/IJST/2016/V9I12/89942>.
- [24] Agarwal S, Yadav D, Verma A. Speed control of PMSM drive using bacterial foraging optimization. *4th IEEE Uttar Pradesh Sect. Int. Conf. Electr. Comput. Electron. UPCON 2017*, vol. 2018- January, Institute of Electrical and Electronics Engineers Inc.; 2017, p. 84–90. <https://doi.org/10.1109/UPCON.2017.8251027>.
- [25] Jasim MH. Tuning of a PID Controller by Bacterial Foraging Algorithm for Position Control of DC Servo Motor. *Eng Technol J* 2018; 36:287–94. <https://doi.org/10.30684/etj.36.3A.7>.
- [26] Chang W Der, Chen CY. PID controller design for MIMO processes using improved particle swarm optimization. *Circuits, Syst Signal Process* 2014;33:1473–90. <https://doi.org/10.1007/S00034-013-9710-4/FIGURES/8>.
- [27] Rajesh RJ, Ananda CM. PSO tuned PID controller for controlling camera position in UAV using 2-axis gimbal. *Proc. 2015 IEEE Int. Conf. Power Adv. Control Eng. ICPACE 2015*, Institute of Electrical and Electronics Engineers Inc.; 2015, p. 128–33. <https://doi.org/10.1109/ICPACE.2015.7274930>.
- [28] Lodhi RS, Saraf A. Survey on PID Controller Based Automatic Voltage Regulator. *Int J Adv Res Electr Electron Instrum Eng* 2016; 5:7424–9.


- [29] Fister D, Fister I, Fister I, Šafarič R. Parameter tuning of PID controller with reactive nature-inspired algorithms. *Rob Auton Syst* 2016; 84:64–75. <https://doi.org/10.1016/J.ROBOT.2016.07.005>.
- [30] Joseph SB, Dada EG. Proportional-integral-derivative (PID) controller tuning for an inverted pendulum using particle swarm optimisation (PSO) algorithm. *FUDMA J Sci* 2018; 2:73–9.
- [31] Kumar P, Nema S, Padhy PK. PID controller for nonlinear system using cuckoo optimization. *Int. Conf. Control. Instrumentation, Commun. Comput. Technol. ICCICT 2014, Institute of Electrical and Electronics Engineers Inc.; 2014, p. 711–6.* <https://doi.org/10.1109/ICCICT.2014.6993052>.
- [32] Gholap V, Dessai CN, Bagyaveereswaran V. PID controller tuning using metaheuristic optimization algorithms for benchmark problems. *IOP Conf Ser Mater Sci Eng* 2017; 263:052021. <https://doi.org/10.1088/1757-899X/263/5/052021>.
- [33] Bingul Z, Karahan O. A novel performance criterion approach to optimum design of PID controller using cuckoo search algorithm for AVR system. *J Franklin Inst* 2018; 355:5534–59. <https://doi.org/10.1016/J.JFRANKLIN.2018.05.056>.
- [34] Bansal R, Jain M, Bhushan B. Designing of Multi-objective Simulated Annealing Algorithm tuned PID controller for a temperature control system. *6th IEEE Power India Int. Conf., Institute of Electrical and Electronics Engineers (IEEE); 2015, p. 1–6.* <https://doi.org/10.1109/POWERI.2014.7117716>.
- [35] Vijay D, Banu US. PID controller tuned using Simulated Annealing for assured crew re-entry vehicle with PWPf thruster firing. *Proc. 2016 Online Int. Conf. Green Eng. Technol. IC-GET 2016, Institute of Electrical and Electronics Engineers Inc.; 2017.* <https://doi.org/10.1109/GET.2016.7916804>.
- [36] Lahcene R, Abdeldjalil S, Aissa K. Optimal tuning of fractional order PID controller for AVR system using simulated annealing optimization algorithm. *5th Int. Conf. Electr. Eng. - Boumerdes, ICEE-B 2017, vol. 2017- January, Institute of Electrical and Electronics Engineers Inc.; 2017, p. 1–6.* <https://doi.org/10.1109/ICEE-B.2017.8192194>.
- [37] Shatnawi M, Bayoumi E. Brushless DC motor controller optimization using simulated annealing. *Int. Conf. Electrical Drives Power Electron., vol. 2019-September, KoREMA; 2019, p. 292–7.* <https://doi.org/10.1109/EDPE.2019.8883924>.
- [38] ŞEN MA, KALYONCU M. Optimal Tuning of PID Controller Using Grey Wolf Optimizer Algorithm for Quadruped Robot. *Balk J Electr Comput Eng* 2018; 6:29–35. <https://doi.org/10.17694/BAJECE.401992>.
- [39] Agarwal J, Parmar G, Gupta R, Sikander A. Analysis of grey wolf optimizer based fractional order PID controller in speed control of DC motor. *Microsyst Technol* 2018; 24:4997–5006. <https://doi.org/10.1007/S00542-018-3920-4/FIGURES/13>.
- [40] Yadav S, Nagar SK, Mishra A. Tuning of parameters of PID controller using Grey Wolf Optimizer. *Proc. Int. Conf. Adv. Electron. Electr. Comput. Intell. 2019, Elsevier BV; 2019.* <https://doi.org/10.2139/SSRN.3575432>.
- [41] Sule AH, Mokhtar AS, Jamian JJ Bin, Khidrani A, Larik RM. Optimal tuning of proportional integral controller for fixed-speed wind turbine using grey wolf optimizer. *Int J Electr Comput Eng* 2020; 10:5251–61. <https://doi.org/10.11591/IJECE.V10I5.PP5251-5261>.
- [42] Verma SK, Devarapalli R. Fractional order PID controller with optimal parameters using Modified Grey Wolf Optimizer for AVR system. *Arch Control Sci* 2022; 32:429–50. <https://doi.org/10.24425/acs.2022.141719>.
- [43] Hekimoğlu B, Ekinci S, Kaya S. Optimal PID Controller Design of DC-DC Buck Converter using Whale Optimization Algorithm. *Int. Conf. Artif. Intell. Data Process. IDAP 2018, Institute of Electrical and Electronics Engineers Inc.; 2018.* <https://doi.org/10.1109/IDAP.2018.8620833>.
- [44] Kumar AA, Kumar G, Anil A, Dr K, Giriraj Kumar SM. Application of Whale Optimization Algorithm for tuning of a PID controller for a drilling machine Atal Anil Kumar, S M Giriraj Kumar. *Application of Whale Optimization Algorithm for tuning of a PID controller for a drilling machine Application of Whale Optimization Algorithm for tuning of a PID controller for a drilling machine. ICAARS 2018.*
- [45] Mosaad AM, Attia MA, Abdelaziz AY. Whale optimization algorithm to tune PID and PIDA controllers on AVR system. *Ain Shams Eng J* 2019; 10:755–67. <https://doi.org/10.1016/J.ASEJ.2019.07.004>.
- [46] Loucif F, Kechida S, Sebbagh A. Whale optimizer algorithm to tune PID controller for the trajectory tracking control of robot manipulator. *J Brazilian Soc Mech Sci Eng* 2020; 42:1–11. <https://doi.org/10.1007/S40430-019-2074-3/FIGURES/11>.
- [47] Bendjeghaba, O., Boushaki, S. I., & Zemmour, N. (2013, May). Firefly algorithm for optimal tuning of PID controller parameters. *4th International Conference on Power Engineering, Energy and Electrical Drives*. Presented at the 2013 IV International Conference on Power Engineering, Energy and Electrical Drives (POWERENG), Istanbul, Turkey. doi:10.1109/powereng.2013.663579
- [48] Coelho, L. dos S., & Mariani, V. C. (2012). Firefly algorithm approach based on chaotic Tinkerbell map applied to multivariable PID controller tuning. *Computers & Mathematics with Applications* (Oxford, England: 1987), 64(8), 2371–2382. doi: 10.1016/j.camwa.2012.05.007

- [49] Ekinci, S., Izci, D., & Hekimoglu, B. (2020, June). PID speed control of DC motor using Harris hawks optimization algorithm. 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE). Presented at the 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Istanbul, Turkey. doi:10.1109/icecce49384.2020.9179308
- [50] Nise NS. Control Systems Engineering. 7th Edition. John Wiley & Sons; 2013.
- [51] Introduction: PID Controller Design n.d. <https://ctms.engin.umich.edu/CTMS/index.php?example=Introduction§ion=ControlPID> (accessed November 13, 2022).
- [52] Kennedy J, Eberhart R. Particle swarm optimization. Proc. ICNN'95 - Int. Conf. Neural Networks, vol. 4, 1995, p. 1942–8 vol.4. <https://doi.org/10.1109/ICNN.1995.488968>.
- [53] Mirjalili S, Mirjalili SM, Lewis A. Grey Wolf Optimizer. Adv Eng Softw 2014; 69:46–61. <https://doi.org/https://doi.org/10.1016/j.advengsoft.2013.12.007>.
- [54] Mirjalili S, Lewis A. The Whale Optimization Algorithm. Adv Eng Softw 2016; 95:51–67. <https://doi.org/https://doi.org/10.1016/j.advengsoft.2016.01.008>.
- [55] Yang X-S. Firefly Algorithms for Multimodal Optimization. In: Watanabe O, Zeugmann T, editors. Stoch. Algorithms Found. Appl., Berlin, Heidelberg: Springer Berlin Heidelberg; 2009, p. 169–78.
- [56] Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H. Harris hawks optimization: Algorithm and applications. Futur Gener Comput Syst 2019; 97:849–72. <https://doi.org/https://doi.org/10.1016/j.future.2019.02.028>.
- [57] Zhao W, Wang L, Mirjalili S. Artificial hummingbird algorithm: A new bio-inspired optimizer with its engineering applications. Comput Methods Appl Mech Eng 2022; 388:114194. <https://doi.org/https://doi.org/10.1016/j.cma.2021.114194>.
- [58] Abdollahzadeh B, Gharehchopogh FS, Mirjalili S. African vultures optimization algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems. Comput Ind Eng 2021; 158:107408. <https://doi.org/https://doi.org/10.1016/j.cie.2021.107408>.

Is ChatGPT Leading Generative AI? What is Beyond Expectations?

*¹Ömer AYDIN, ²Enis KARAARSLAN

¹Department of Electrical and Electronics Engineering, Celal Bayar University, Manisa, Türkiye, omer.aydin@cbu.edu.tr 

²Department of Computer Engineering, Muğla Sıtkı Koçman University, Muğla, Türkiye, enis.karaarslan@mu.edu.tr 

Abstract

Generative AI has the potential to change the way we do things, and chatbot is one of the most popular implementation areas. Even though companies like Google and Meta had chatbots, ChatGPT became popular as it was publicly available. Although ChatGPT is still in the early stages of its development, it attracted the attention of people and capital groups. It has taken the public interest; people from different fields, ages, and education levels started using ChatGPT. There have been many trials with ChatGPT. It is possible to see wide range of news and social media posts about it on the Internet. The study aims to shed light on what is happening by the studies in the literature and give insight on the user expectations of ChatGPT and Generative AI. The technical and structural fundamentals of ChatGPT and its competitors (Google's Bard AI, Claude, Meta's Wit.ai and Tencent's HunyuanAide) are described. We try to shed light on who will win the race by sharing the early-stage due diligence and current situation analysis for all these points. Preprint papers and published articles are examined. We also used ChatGPT in editing the content of the available resources. Even with the presence of the wrong answers, Generative AI and its large language models are promising.

Keywords: ChatGPT, GPT4, Generative AI, Artificial Intelligence, LLM

1. INTRODUCTION

As humanity started to use Artificial intelligence (AI) in many areas, companies lead a new paradigm called Generative AI. Generative AI is focusing on creating new and unique content with the trained data. This content can be literature, images, music, video and etc. The written text can even be an academic paper or a film scenario. Possible usage areas are endless.

Chatbot is one of the implementation areas of generative AI. Chatbots have always been in focus as humanity had the question of "Can computers think?" [1]. Companies such as Google, Meta, DeepMind, and OpenAI are working on developing chatbots. ChatGPT gained popularity as the OpenAI company made its language models and user interfaces publicly available. Even though these environments are still evolving, their reliability is widely discussed in social networks and also gained the attention of academic studies.

This study aims to answer the following research questions; Is ChatGPT leading generative AI? What are the main competitors? Are these solutions satisfying people's expectations? In this paper, we present an overview of the technology, give a literature survey, an early-stage due

diligence and current situation analysis to address these research questions and discuss the challenges.

In the next section, fundamentals are given. In Section 3, the method is described. Then the literature review is given in Section 4. In section 5, discussion and evaluation are presented. The conclusion and future works are given in the last section.

2. FUNDAMENTALS

This section covers the main topics that form the basis of the used technologies; Natural Language Processing (NLP), tokens, Language Models (LMs), Large Language Models (LLMs), Generative Adversarial Network (GAN) and Generative AI.

2.1. NLP

Natural Language Processing is one of the important research areas of computer science. The ultimate goal to achieve with NLP is to create a computer system that can understand and respond to human speech or written text. NLP combines many technologies, algorithms, and models. In particular, it can use computational linguistics, machine learning, deep learning models, and artificial intelligence applications. While translating from one language to another,

it can generate answers or process commands according to the information supplied. Although what can be done with NLP is very diverse, we can list some of its basic applications as speech-to-text, part-of-speech tagging, word sense disambiguation, coreference resolution, sentiment analysis, and natural language generation [2].

Even though it takes many years for people to learn natural languages, and it is difficult to fully master the entire language; NLP can achieve this with significant success. There are still challenges with idioms, proverbs, spelling mistakes, and expression disorders.

2.2. Tokens, Language Models (LMs) & Large Language Models (LLMs)

We need to model languages to make the computer understand the human languages to use in speech recognition, machine translation, natural language generation, information retrieval and related applications. “A language model assigns a probability to a piece of unseen text, based on some training data” [3]. The studies on it had started with studies on probabilistic models of language generation [4].

A token is an instance of a string of characters that has been compiled into a meaningful semantic processing unit. Each document (text) is formed of several tokens. Unigram models use several one-state finite automata. In this approach, it is assumed that the probability of the tokens in a sequence are independent. n-gram (Q-gram) models are based on a continuous sequence of n pieces from a text.

Large-scale language models can be applied to machine translation as in [5]. Increasing the amount of training data to up to 2 trillion tokens was demonstrated experimentally; a 5-gram language model and 300 billion n-grams were produced. New architectures are being experimented that can scale up to 530 billion parameters [6] and more.

2.3. Generative Adversarial Network (GAN)

GANs are one of the most successful advancements in artificial neural network classification, prediction, and problem-solving. It is also one of the most influential developments in machine learning in recent years. Ian Goodfellow and colleagues introduced the Generative Adversarial Network (GAN) in 2014 [7, 8].

As shown in Figure 1, a GAN is constructed from two neural networks: a discriminator and a generator. These two structures work in the constant connection between them. While the generator creates new instances, the discriminator attempts to distinguish them from the genuine cases in the training dataset. The two networks are trained together, with the discriminator seeking to detect generated instances more correctly and the generator attempting to generate samples that the discriminator cannot distinguish as distinct from real ones.

GANs can create new images that are almost indistinguishable from reality. The generator produces an

image, and the discriminator checks whether the produced image is real or fake. A pre-trained model is not used, education takes place in the process. The discriminator has real pictures. The generator creates a new image and sends it to the discriminator. Discriminator compares the new image with the real images and returns a response to the generator. According to the answer, the generator constantly updates and improves itself. This process continues in a loop, and eventually, the generator starts producing more realistic pictures [9].

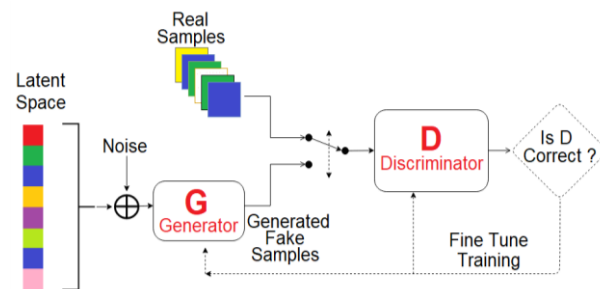


Figure 1. Generative adversarial network (GAN) model (adapted from [10])

2.4. Generative AI

Generative AI is an artificial intelligence field that concentrates on generating new and original information by machine learning on massive databases of experiences. There are several potential applications for generative AI, such as the generation of new pictures, text, and music. Also, it is used in computer vision, natural language processing, and speech recognition. As an example, generative models may be used to create realistic images for video games, simulations, and virtual reality, as well as novel chemical compounds for medicinal research.

This is performed by using a model that has been trained on a large dataset of examples and constructing new instances that are comparable to the training dataset. Natural language processing generative models may be used to produce new material that is equivalent to the style and substance of a specific author or subject. Among the most popular generative AI models are GANs, Variational Autoencoders (VAE), and Transformer-based models.

GAN is one of the most well-known models of generative AI. Another well-known generative AI technique is the VAE, a type of neural network that learns to create new instances by learning to compress and decompress input. After this model is trained on a collection of examples, a VAE learns to compress each example into a tiny, low-dimensional representation known as a latent code. The VAE can decompress a latent code to generate a new example that is analogous to those in the training dataset.

There are also other generative models, such as Transformer-based models (like GPT-2 and GPT-3), which use the attention mechanism to generate new text, pictures, and sounds that are remarkably similar to real-world instances. These models can generate high-quality, consistent text, images, and animation [11].

Table 1. Classification of the AI Tools [12, 13]

Classification	Tool Name and Web Page Address	Operation Mode
Text to Image (T2I)	DALL-E 2 (https://openai.com/product/dall-e-2)	Generation
	Stable Diffusion (https://stability.ai/)	Generation
	Craiyon (https://www.craiyon.com/)	Generation
	Jasper (https://www.jasper.ai)	Generation
	Imagen (https://imagen.research.google/)	Generation
	MidJourney (https://www.midjourney.com)	Generation
	NightCafe (https://nightcafe.studio/)	Generation
	GauGAN2 (https://gpt3demo.com/apps/gaugan2-by-nvidia)	Generation
	Wombo (https://www.w.ai/)	Generation
	Wonder (https://www.wonder-ai.com/)	Generation
	Pixray-test2image (https://pixray.gob.io/)	Generation
	Neural.love (https://neural.love/)	Classification, Generation, Transformation
	Text to Video (T2V)	Runway (https://runwayml.com/)
Fliki (https://fliki.ai/)		Generation
Synthesia (https://www.synthesia.io/)		Generation
MetaAI (https://ai.facebook.com/)		Generation
Google AI (https://ai.google/)		Generation
Text to Audio (T2A)	Phenaki (https://phenaki.video/)	Generation
	Play.ht (https://play.ht/)	Generation
	Murf.AI (https://murf.ai/)	Generation
	Resemble.AI (https://resemble.ai)	Generation
	WellSaid (https://wellsaidlabs.com/)	Generation
Text to Text (T2T)	Descript (https://www.descript.com/)	Classification, Generation, Transformation
	Simplified (https://simplified.com/)	Generation
	Jasper (https://www.jasper.ai)	Generation
	Frase (https://www.frase.io/)	Transformation
	EleutherAI (https://www.eleuther.ai/)	Generation
	Requstory (https://requstory.com/)	Generation
	Grammarly (https://www.grammarly.com)	Transformation
	Copy.ai (https://www.copy.ai/)	Generation
	MarketMuse (https://www.marketmuse.com/)	Generation
	AO21labs (https://www.ai21.com/)	Transformation
	HubSpot (https://www.hubspot.com/)	Classification
	InterKit (https://app.inferkit.com)	Generation
	GooseAI (https://goose.ai/)	Generation
	ResearchAI (https://research-ai.io/)	Generation
	Writesonic (https://writesonic.com/)	Generation
	Co:here (https://cohere.ai/)	Classification
	CHIBI (https://chibi.ai/)	Generation
	Ideas AI (https://ideasai.com/)	Generation
	Copysmith (https://copysmith.ai/)	Generation
	Flowrite (https://www.flowrite.com/)	Generation
	NICHESS\$ (https://nichess.com/)	Generation
	Sudowrite (https://www.sudowrite.com/)	Generation
	Ideasbyai (http://ideasbyai.com/)	Generation
	Text.cortex (https://textcortex.com/)	Classification, Transformation
	OpenAI GPT3 (https://openai.com/blog/chatgpt)	Generation
	AISEO Blog Idea Generator (https://aiseo.ai/templates/blog-idea-generator.html)	Generation
	Rytr AI https://rytr.me/	Generation
	PepperType AI https://www.peppertype.ai/	Generation
	Kafkai AI https://kafkai.com/en/	Generation
	Texta AI https://texta.ai/dark	Generation
	Anyword AI https://anyword.com/	Generation
	DeepL Write AI https://www.deepl.com/write	Generation
	Perplexity https://www.perplexity.ai/	Generation
Elicit https://elicit.org/	Generation	
Text to Motion (T2M)	TREELnd. (https://www.tree.industries/)	Generation
	MDM: Human Motion Diffusion Model (https://guytevet.github.io/mdm-page/)	Generation
Text to Code (T2C)	Replit Generate code(https://docs.replit.com/power-ups/ghostwriter/generate-code)	Generation
	Github Copilot (https://github.com/features/copilot)	Generation
Text to NFT (T2N)	LensAI (https://lens-ai.com/)	Generation
Text to 3D (T2D)	DreamFusion (https://dreamfusion3d.github.io/)	Generation
	Clip-Mesh (https://www.nasir.lol/clipmesh)	Generation
	GET3D (https://nv-tlabs.github.io/GET3D/)	Generation
Audio to Text (A2T)	Descript (https://www.descript.com/tools/audio-text)	Classification, Generation, Transformation
	AssemblyAI (https://www.assemblyai.com/)	Transformation
	Whisper (OpenAI) (https://github.com/openai/whisper)	Transformation
	Elsa Speaks https://elsaspeak.com/en/	Classification, Generation
Audio to Audio (A2A)	AudioLM (https://google-research.github.io/seanet/audiolm/examples/)	Transformation
	VOICEMOD (https://www.voicemod.net/)	Transformation
Brain to Text (B2T)	Speech from brain (Meta AI) (https://ai.facebook.com/blog/ai-speech-brain-activity/)	Generation
	Non-invasive brain recordings (https://hal.science/hal-03808317/document)	Generation
Image to Text (I2T)	Neural.love (https://neural.love/)	Classification, Generation, Transformation
	GPT-2 x Image Captions (https://huggingface.co/nlpcnnect/vit-gpt2-image-captioning)	Transformation

The classification of some selected Generative AI tools is given in Table 1. Generative AI tools are grouped as Text to Image (T2I), Text to Video (T2V), Text to Audio (T2A), Text to Text (T2T), Text to Motion (T2M), Text to Code(T2C), Text to NFT (T2N), Text to 3D (T2D), Audio to Text (A2T), Audio to Audio (A2A), Brain to Text (B2T), and Image to Text (I2T). These tools are categorized with their operation mode. These operations are listed as classification, generation, and transformation.

2.4.1. ChatGPT

ChatGPT is an advanced natural language processing model. The model was pre-trained on a large dataset, allowing it to create a language that is coherent, contextually appropriate, and often indistinguishable from human-written stuff. It makes use of a transformer architecture that merges the benefits of generative models with scalability. It creates natural language using a self-attention mechanism, allowing it to learn the context of the dialogue and offer more accurate replies.

ChatGPT enables text production in any terminology and topic from any specified prompt. Its modular architecture allows it to perform a variety of tasks, including question-answering, machine translation, language modeling, and more. It may also be used to develop virtual assistants and chatbots, as well as generate human-like discussions. ChatGPT's design and algorithms enable it to produce long-form and rich material.

ChatGPT has a big potential and is already being utilized in a variety of industries. ChatGPT may be used to develop complex applications in any sector, such as medicine, education, finance, and others. Its possible usage areas will be discussed in the literature survey section.

The exact implementation details and architecture of ChatGPT are proprietary to OpenAI. However, a simplified representation of the underlying algorithm and pseudocode used in the ChatGPT model can be given as follows. The ChatGPT model is built using a variant of the Transformer architecture, which is a popular deep-learning model for natural language processing tasks. The Transformer architecture consists of an encoder-decoder framework with self-attention mechanisms. The model is trained in a supervised manner using a large dataset of paired input-output examples. During training, the model learns to predict the most likely next token in a sequence given the previous context. This is done by minimizing a loss function, typically cross-entropy loss, between the predicted distribution and the actual target distribution.

The simplified pseudocode representation of the ChatGPT model can be given as follows [14]:

```
function ChatGPT(input_sequence){
    tokenized_input = tokenize(input_sequence)
    input_ids = convert_tokens_to_ids(tokenized_input)
    output_sequence = ""
    while not end_of_sequence(output_sequence):
        tokenized_output = tokenize(output_sequence)
        output_ids = convert_tokens_to_ids(tokenized_output)
        combined_ids = concatenate(input_ids, output_ids)
        encoded_inputs = encoder_model(combined_ids)
        decoded_output = decoder_model(encoded_inputs)
        next_token = sample_next_token(decoded_output)
        output_sequence += convert_id_to_token(next_token)
    return output_sequence
}
```

This pseudocode only shows the basic flow. It tokenizes the input sequence, converts the tokens to numeric IDs, and iteratively generates the output sequence by sampling the next token based on the binding encoded in the input sequence and the previously generated output.

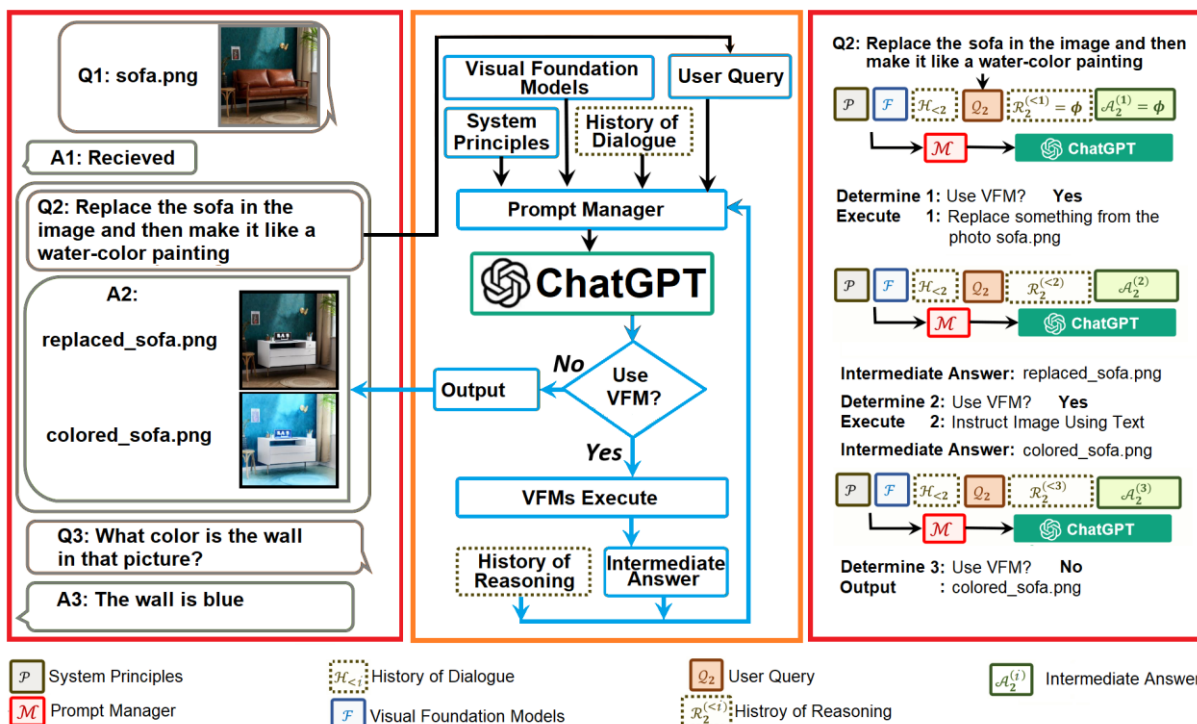


Figure 2. Visual ChatGPT System Architecture (adapted from [15])

2.4.1.1. Visual ChatGPT

Teams at Microsoft are engaged in the Visual ChatGPT research. It has been developed based on OpenAI's ChatGPT, as ChatGPT is currently unable to create images. Several enhancements are made to the core features of ChatGPT. Visual ChatGPT can interact with existing ChatGPT using both text and visual images. This research aims to reach the ability to add image processing to text-based communication. The product will allow dealing with complex visual questions that involve multiple steps and require multiple AI models to work together. Also, it will provide feedback and ask for corrected results.

The system structure of Visual GPT can be seen in Figure 2 [15]. A three-step dialogue is given on the left side of Figure 2. A flowchart is shared in the middle. This flowchart shows how Visual Foundation Models are called recursively, and how answers are provided. The right part of the figure shows the detailed process of the second Question-Answer (QA) dialogue.

2.4.1.2. GPT4

OpenAI released GPT 4, the new version of ChatGPT on March 14, 2023. GPT4 is a multimodal and large-scale model that accepts images and text as input and can produce text output. In many real-world scenarios, it is currently less capable than a human. Nevertheless, it can pass tests where humans can succeed at low rates, and it can succeed in many human-level professional and academic processes. GPT-4 outperforms both major previous language models and most high-end systems on a range of traditional NLP tests. Despite all its capabilities, GPT-4 has similar limitations to earlier GPT models. GPT-4 is not completely reliable, has a limited context window, and does not learn from experience. Caution should be exercised when using GPT-4 results in situations and operations where reliability is critical [16].

2.4.2. ChatGPT Rivals

ChatGPT has reached a lot of users around the world right now, and it is increasing its popularity day by day as a hype. But are its rivals, which seem to be backward, to be so inadequate? What is the technical and structural situation of the competitors? Under this title, we will question the main differences between ChatGPT and its strong rivals and whether it is better than them.

2.4.2.1. LaMDA Bard.AI

Google's LaMDA (Language Model for Dialogue Applications) is a collection of conversational neural language models. The first generation was revealed during the Google I/O keynote in 2021. LaMDA attracted great attention in June 2022 after Google employee Blake Lemoine claimed that the chatbot had grown sentient. Lambda is based on Transformer, a neural network architecture designed by Google Research and open-sourced in 2017. This architecture generates a model that can be taught to read many words, pay attention to how those words connect and then predict which words it believes will come next [17]. After Microsoft had stated that it intends to

incorporate AI into all of its products, Google introduced Bard, a conversational artificial intelligence chatbot powered by LaMDA, in February 2023.

Bard is powered by LaMDA which is Google's collection of conversational neural language models. Bard AI is a cloud-based conversational AI platform that allows organizations to build and deploy chatbots that can communicate with consumers through various channels such as websites, messaging applications, and voice assistants. The platform is user-friendly, with a simple drag-and-drop interface that allows businesses to develop and configure chatbots easily without coding or technical knowledge. Bard AI is driven by powerful machine learning algorithms and natural language processing (NLP) technologies, which allow chatbots to comprehend and reply to client inquiries in a meaningful and engaging manner. This implies that clients may converse with chatbots in a natural and human-like manner, making it easier for them to discover information and execute activities.

Alphabet CEO Sundar Pichai stated that Google intends to add AI-powered capabilities to its search engine to synthesize information and answer difficult inquiries. These capabilities will be initially powered by LaMDA and eventually by other AI technologies. Pichai stated that Google will begin providing tools to web developers, producers, and companies in 2003 [18].

With the opening of Google Bard to public use, the number of academic studies on Bard has started to increase. Google Bard started to be used and will be used in many fields such as academic writing, health, education and law etc. [19-22].

2.4.2.2. Claude

Anthropic, an AI business co-founded by former OpenAI workers, has discreetly begun testing Claude, a new ChatGPT-like AI assistant. Both Claude and ChatGPT use reinforcement learning (RL) to train a preference model over their outputs, and preferred generations are used for fine-tuning subsequently. Anthropic differs from its using Constitutional AI to construct these preference models [23]. Constitutional AI builds upon reinforcement learning from human feedback (RLHF) baseline with a process given in Figure 3.

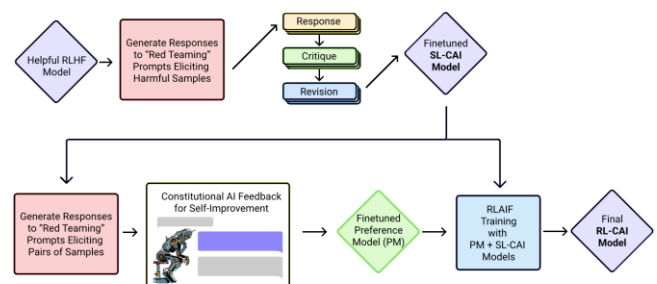


Figure 3. Basic Steps of Constitutional AI (CAI) process (adapted from [24])

2.4.2.3. Wit.ai

Wit.ai is one of the leading bot frameworks in the market with strong NLP capabilities. Its NLP engine can be used to build conversational applications and devices that can be interacted with through speech or text. It provides an easy-to-use interface and quick-learning APIs to analyze human communication and turn complex messages into structured data. The engine also helps to predict future events based on past data. Wit.ai is well-positioned with Facebook's support to innovate in speech recognition and voice interfaces for developers. Its continuous learning and the ability to use relevant data to connect with specific customers makes the difference [25]. Many applications have been using Wit.ai. For example, Handoyo et al. constructed a chatbot for the reservation system using Wit.ai [26]. Figure 4 illustrates the request and response flow of the system architecture.

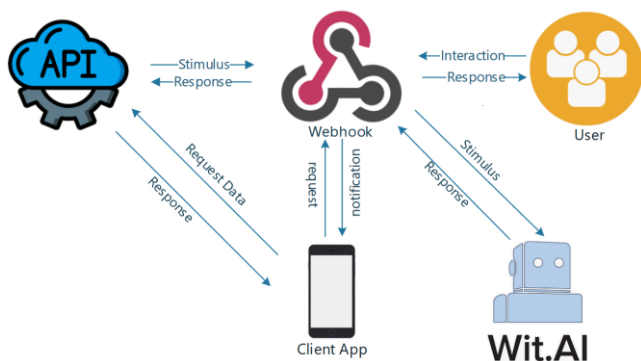


Figure 4. Serverless Architecture Chatbot using Wit.AI (adapted from [26])

2.4.2.4. HunyuanAide

It is known that Microsoft-backed OpenAI does not allow users in China to access the chatbot. As the open AI models behind the program are increasingly finding wider applications in consumer technology applications in China, Alibaba Group, Baidu Inc., and Tencent Holdings announced that they are working on their solutions. Tencent Holdings, one of the leading internet companies in China, announced that it formed a development team to work on a chatbot similar to ChatGPT. The product will be named "HunyuanAide", and it will use the AI training model called "Hunyuan" [27].

2.4.2.5. OPT by Meta

Meta released the Open Pretrained Transformer (OPT) in May 2022. OPT language model contains 175 billion parameters. It's trained on multiple general datasets,

including The Pile (a dataset with 22 subsets of more than 800GB of English text) and BookCorpus. OPT has combined its pre-trained models and source code to use/train these models. It is currently available for research purposes with a non-commercial license [28].

2.4.2.6. AlexaTM by Amazon

Amazon publicly released AlexaTM 20B, a large-scale multilingual sequence2sequence model, in November 2022. It uses an encoder-decoder architecture. It is trained on a combination of Causal Language Modeling (CLM) and noise removal tasks. It outperforms Google's PaLM 540B at one-shot summarization and machine translation tasks. AlexaTM supports multiple languages [28].

2.4.2.7. GPT-J and GPT-NeoX by EleutherAI

GPT-J was trained on the Pile. Despite its smaller size compared to its competitors, it can perform almost the same as the GPT-3 6.7B-param. It also outperforms its predecessor, the GPT-Neo. It is converted to GPT-NeoX with 20B parameters in February 2022. The comparison of GPT-J and GPT-NeoX performance to OpenAI GPT-3 versions is given in Table 2. As seen in the table, there is almost no difference in performance between paid GPT3 and open-source GPT-NeoX and GPT-J

2.4.2.8. Jurassic-1 language model by AI21 labs

Jurassic-1 is an autoregressive natural language processing (NLP) model. There is a fee for registration. Jurassic-1 consists of two parts. The first part is J1-Jumbo, which is trained on more than 178B parameters. The second one is J1-Large, which consists of 7B parameters. It is more advanced than the GPT-3 language model [28].

2.4.2.9. CodeGen by Salesforce

CodeGen is a large-scale language model that can generate programs from simple text input. The future of coding is at the crossroads of human and computer languages, and speech-based AI is the ideal bridge to bring the two together. The CodeGen version is available in three models (NL, multi, and mono), and with different sizes (350M, 2B, 6B, and 16B). NL models use the Pile dataset. Multiple models are built on NL models and employ a variety of code written in various computer languages. Mono models are built from many models and employ Python code. The most exciting aspect of CodeGen is that it does not need technical knowledge [28].

Table 2. Performance comparison of GPT3, GPT-J and GPT-NeoX [28]

Task	Category	Babbage	GPT-J-6B	Curie	GPT-NeoX-20B	GPT3 DaVinci
LAMBADA	Sentence Completion	62.49%	68.29%	69.51%	72.00%	75.16%
ANLI R3	Natural Language Inference	33.75%	35.50%	35.50%	35.40%	36.83%
WSC	Coreference Resolution	54.54%	49.54%	49.54%	50.00%	59.18%
WinoGrande	Coreference Resolution	59.51%	64.01%	64.56%	66.10%	69.93%
HellaSwag	Sentence Completion	40.38%	36.53%	54.81%	53.50%	63.46%

2.4.2.10. Megatron-Turing NLG by NVIDIA and Microsoft

This LLM is one of the biggest language models, with over 530B parameters. Microsoft and NVIDIA collaborated to develop Megatron-Turing Natural Language Generation (NLG). They employed the Pile dataset and used the NVIDIA DGX SuperPOD-based Selene supercomputer to train the model. According to research published in October 2021, the Megatron-Turing NLG performed exceptionally well on PiQA dev kit tasks and LAMP test set workloads. The model also predicts over 50% of the time in zero-shot testing and much better in one-shot and four-shot tests [28].

3. METHOD

The literature review will be used as a method in this study. First of all, the studies published on the relevant subject in the literature will be systematically examined. Then, preprint studies that have not been published yet will be compiled. Finally, information shared on some platforms on the internet and various news and blogs will be included.

For the literature review, searches will be made with the word groups "ChatGPT", and "Generative AI" on SSRN, Wos, Scopus, and Google Scholar, and the found studies will be examined. In addition, a search will be made on LinkedIn with the hashtags "#chatgpt" and "#GenerativeAI" and the ones that can be used in the work will be examined. Finally, current and relevant blog posts and news will be compiled by searching on Google with the word groups "ChatGPT" and "Generative AI".

4. LITERATURE REVIEW

Many studies are available in the literature on ChatGPT. The number of these publications has started to increase especially at the end of 2022, as it became popular again with GPT3 and being free of charge made it used by a large number of users.

As a result of our research, we have seen that it is possible to collect the studies in the literature in the following subgroups. Some authors have analyzed ChatGPT only by asking questions and analyzing their answers. Others have mentioned the possible effects and made evaluations. Table 3 groups the reviewed literature by their study fields. The main fields of these studies are general education, academic writing, writing, health, law, economics and finance. On the other hand, the negative aspects of ChatGPT and the strengths of its competitors were emphasized on social media, blog sites and various platforms.

There are many studies that we can group on general aspects of ChatGPT. Benzon offers three claims about ChatGPT in his paper [29]. The first one is that because of the intricacy of ChatGPT's behavior, the concept of thinking is no longer applicable in distinguishing between human and AI activity. The second one is that the operation of ChatGPT is divided into two levels. Last is that the symbolic computation aspect is latent in Large Language Models (LLM), and an appendix shows how a story grammar is constructed. According to

Benzon's other study [30], an artificial neural net can be regarded as a platform for implementing higher-level structures such as word processors or databases.

ChatGPT's neural net uses a simple grammatical structure with five components termed "Donné, Disturb, Plan, Enact, and Celebrate." Through four studies, the results reveal that ChatGPT can adapt a tale based on a change in protagonist or adversary [31]. He investigates the idea of using Noam Chomsky's concept of linguistic competence and David Marr's degrees of analysis to understand how language models such as ChatGPT function in his other study. The text of ChatGPT reveals that it has a command of discourse skills, such as the ability to govern language patterns, execute analogical reasoning in tail interpretation, comprehend complex concepts, and alter discourse for different ages [30].

Sætra contends that embracing the inevitability of technological progress without considering its repercussions is a poor strategy. ChatGPT has become a global hot subject, and its impact on generative AI is evident. The question today is how much of an impact it will have and what implications would result from utilizing AI to generate text and other stuff. Before adopting new technology as progress, he feels it is critical to understand how they alter values, power dynamics, and social structures. Sætra recognizes the great potential of generative AI but says it should be regulated to minimize undesirable consequences at multiple levels. He believes that technology should not be stopped, but rather carefully guided to ensure that its development and application do not result in disaster [32]. In Shah's paper, he offered his opinions, evaluations, and visions on artificial intelligence chat agents that have emerged as technology has advanced [33]. With a focus on the chatbot ChatGPT, Lund and Agbaji's study investigates the links between three digital literacies (information literacy, data literacy, and privacy literacy) and attitudes toward technology adoption. A 2-week web survey was used to collect data from 130 persons in northern Texas. The findings reveal a link between using ChatGPT and information and privacy literacy abilities, but not data literacy skills. Age, gender, education, and internet usage are all important determinants. These findings are important for understanding the interplay between literacies and personal/community factors [34]. Bommarito et al. evaluate the capability of OpenAI's text-davinci-003 and previous versions of GPT for knowledge work as measured by the American Institute of Certified Public Accountants (AICPA). Text-davinci-003 underperformed on numeric reasoning in zero-shot prompts but neared human-level performance on Remembering & Understanding and Application skill levels, according to the results. Recent GPT-3 generations improved, rising from 30% for text-davinci-001 to 57% for text-davinci-003. The findings show that large language models have the potential to improve the quality and efficiency of knowledge work [35].

Table 3. Study Fields of the reviewed literature

Paper / Field	General Aspects of ChatGPT	Education	Health	Writing and Academic Writing	Business, Finance and Economy	Law
Shah, 2023 [33]	✓					
Lund & Agbaji, 2023 [34]	✓					
Bommarito et al., 2023 [35]	✓					
Hartman et al.,2023 [36]	✓					
Gozalo-Brizuela & Garrido-Merchan, 2023 [37]	✓					
Azaria, 2022 [38]	✓					
ChatGPT & Perlman, 2022 [39]	✓					✓
Jabotinsky and Sarel, 2022 [40]	✓					
Frye, 2022 [41]		✓		✓		
Hacker, 2022 [42]	✓					✓
Baidoo-Anu & Owusu Ansah, 2023 [43]		✓				
Zhai, 2022 [44]		✓				
Zhai, 2023 [45]		✓				
Ventayen, 2023 [46]		✓		✓		
Avila-Chauvet et al., 2023 [47]		✓				
Nisar and Aslam, 2023 [48]		✓	✓			
Bommarito & Katz, 2022 [49]		✓				✓
Susnjak, 2022 [50]		✓		✓		
Cotton et al., 2023 [51]		✓				
Mellon et al., 2022 [52]		✓				
Zentner, 2022 [53]		✓				
Ryznar, 2020 [54]		✓		✓		
Gilson et al., 2022 [55]		✓	✓			
Jeblick et al., 2022 [56]			✓			
Shope, 2023 [57]				✓		✓
Bishop, 2023 [58]				✓		
Armstrong, 2023 [59]				✓		✓
Lund & Wang, 2023 [60]	✓			✓		
Willems, 2023 [61]				✓		
Uludağ, 2023 [62]	✓			✓		
Kutela et al., 2023 [63]				✓		
Jiao et al., 2023 [64]				✓		
Aljanabi et al., 2023 [65]	✓			✓		
Pavlik, 2023 [66]	✓	✓		✓		
Gao et al., 2022 [67]				✓		
Marti, 2022 [68]	✓			✓		
Aydın & Karaarslan, 2022 [69]	✓		✓	✓		
Aydın et al., 2023 [70]	✓				✓	
Aydın & Karaarslan, 2023 [71]			✓			
Alici et al.,2023 [72]	✓					
Dowling & Lucey, 2023 [73]					✓	
Mertens, 2023 [74]					✓	
Zaremba & Demir, 2023 [75]					✓	
Chesterman, 2023 [76]					✓	✓
Alshater, 2022 [77]					✓	
Wenzlaff & Sebastian, 2022 [78]					✓	
Leippold, 2023[79]					✓	
Taylor, 2023 [80]						✓
Bishop, 2023[81]						✓
Iu & Wong, 2023 [82]						✓
Choi et al., 2023 [83]		✓				✓

Hartman et al. examine the limitations and biases of ChatGPT, an OpenAI conversational AI model, with an emphasis on its impact on political elections. Using 630 political statements and political compass tests, the authors undertake three experiments and discover that ChatGPT has a pro-environmental, left-libertarian worldview. The authors examine the societal repercussions of politically biased AI [36]. Gozalo-Brizuela and Garrido-Merchan consist of an attempt to present simply the main models and sectors touched by generative AI, in addition to providing a taxonomy of the most recent generative models published

[37]. Azaria presented the benefits and limitations of ChatGPT as a conversational agent in his paper. The paper highlights ChatGPT's bias towards digits favorable to humans and its ability to correct mistakes but also its over-confidence in incorrect answers. The paper also explores ChatGPT's ability to request additional information, explain its responses, and its susceptibility to conflicting answers from small changes in a question. Finally, it is discussed that ChatGPT's responses to social experiments show its tendency to answer similarly to humans [38]. Perlman created an article using OpenAI's ChatGPT to demonstrate

the sophistication and potential ramifications of chatbots in legal services and society. While the ChatGPT responses were unsatisfactory, they show the promise of AI and its potential to revolutionize how we access, create, and consume information and services. It does, however, create regulatory and ethical concerns, such as the role of knowledge workers and the possible exploitation of AI-generated data. The disruptions caused by AI's rapid progress have already occurred, and this study provides a preview of what is to come [39]. Jabotinsky and Sarel's academic study explores the ethical quandaries created by the usage of artificial intelligence (AI), in addition to the governmental solutions that have been implemented to solve these difficulties. It addresses issues like surveillance, bias, and the potential for harm, as well as policy initiatives such as ethical norms, regulatory frameworks, and the promotion of ethical AI technologies [40]. Hacker states that three novel contributions are made in his study. First, it deconstructs the Commission proposals, demonstrating that, while they are a step in the right direction, they ultimately represent a half-hearted approach: if implemented as planned, AI liability in the EU will rely primarily on the disclosure of evidence mechanisms and a set of narrowly defined presumptions concerning fault, defectiveness, and causality. As a result, the article presents recommendations for changes to the proposed AI liability framework. Third, based on an analysis of the key risks AI poses, the final part of the paper maps out a road for the future of AI liability and regulation, in the EU and beyond [42].

The education area is also an area where publications related to ChatGPT are published intensively. While there are many studies on the positive and negative effects of ChatGPT in the field of education, there are also studies on how it can be integrated into education processes and how we can prevent possible harm. Baidoo-Anu and Owusu Ansah discussed ChatGPT, a generative AI platform that attracted over one million subscribers in the first week of its debut. They discuss the potential benefits of ChatGPT in education, such as tailored and interactive learning and formative evaluations, as well as its drawbacks, such as faulty information generation and privacy concerns. The report makes recommendations for using ChatGPT in education and encourages policymakers, researchers, educators, and technology professionals to work together to use the tool securely and effectively to promote students' learning [43]. According to Zhai, the K-12 Framework for Science Education seeks to involve students in scientific procedures for giving them key ideas and crosscutting concepts. This study tests the usage of ChatGPT to address science education difficulties such as measuring students' progress and offering feedback and coaching. The ChatGPT application was utilized in the study to construct a performance-based evaluation, grade and provide comments, and recommend learning materials. The findings indicate that ChatGPT can address issues in science learning, but it cannot replace teachers, who require professional competence to use it. The study also emphasizes the importance of explainable AI to comprehend how ChatGPT generates results [45]. Zhai stated that it reflects on the potential impact of ChatGPT on education and suggests adjusting learning goals to focus on creativity and critical thinking rather than general skills. It also raises concerns

about students outsourcing assessment tasks and calls for new assessment formats that focus on creativity and critical thinking [45]. Ventayen's study investigates the potential risk of academic dishonesty offered by the popular AI model ChatGPT in students generating essays and themes. It uses plagiarism checker tools to examine the resemblance of ChatGPT-generated results and discovers that the results pass the institution's required similarity index, which may jeopardize academic integrity. To prevent cheating with such models, the institution may need to alter its assessment technique [46]. According to Avila-Chauvet et al., OpenAI ChatGPT, a fine-tuned AI chatbot, is a promising tool for behavior analyzers to automate and systematize experimental tasks. The authors show how to use ChatGPT as a programming helper to create an online behavioral assignment in HTML, CSS, and JavaScript code, cutting programming time while gaining precise solutions [47]. Nisar and Aslam's study evaluates the AI-based platform ChatGPT's suitability for self-learning by asking medical questions in pharmacology. The questions covered a wide range of drug-related topics. The responses provided by ChatGPT were deemed to be relevant and accurate, but without a reference or source. Students of traditional and complementary medicine who are struggling with pharmacology might use the tool as a rapid reference and self-study assistance. Despite artificial intelligence's increasing presence in education, its patterns, issues, and possible uses remain largely unexplored [48]. On the multistate multiple choice (MBE) segment of the Bar Exam, OpenAI's text-davinci-003 (GPT-3.5) was tested. Despite people with 7 years of post-secondary education and specific training answering only 68% of questions correctly, text-davinci-003 obtained a 50.3% accuracy rate for the best prompt and parameters and passed the Evidence and Torts parts. The top two and top three selections of the model were correct 71% and 88% of the time, demonstrating good performance. The study discovered no benefit to fine-tuning GPT-3.5, and it reported observations about prompt engineering, parameters, and older GPT models. The current GPT model improvement trend predicts that a language model could pass the MBE component of the Bar Exam in the near future [49]. Susnjak examined ChatGPT, an AI agent, on its capacity to execute high-level cognitive tasks and write human-like texts. It raises worries regarding the potential use of ChatGPT as a tool for online exam cheating. According to the research, ChatGPT is capable of critical thinking and very realistic text generation with minimal input, posing a danger to the integrity of online exams. Invigilated and oral tests, sophisticated proctoring techniques, and AI-text detectors could be solutions, but more study is needed to completely address the issue. Educators and institutions should be aware of the danger of cheating utilizing ChatGPT and look into ways to keep online assessments fair and valid [50]. Cotton et al. analyzes the usage of chatAPIs and GPT-3 in higher education, exploring benefits and drawbacks and risks of academic dishonesty. The paper describes features of chatAPIs and GPT-3, provides examples of their use in education, and highlights the challenge of detecting cheating. They suggest various strategies for universities to ensure responsible and ethical usage of chatAPIs and GPT-3 including developing policies, offering training and support, and implementing methods to detect and prevent cheating [51]. Mellon et al.

put OpenAI's GPT-3 language model to the test when it came to coding open-ended survey responses from the British Election Study Internet Panel (BESIP). The accuracy of GPT-3 was compared to that of a human coder and an SVM machine learning method. GPT-3 had 97% agreement with a second human coder and 95% of the time matched the original human coder's collapsed category. In terms of accuracy and bias, GPT-3 surpassed the SVM. The findings imply that language models such as GPT-3 can be used to code open-ended survey responses, allowing for broader use of open-ended questions [52]. Aeron's report explores the use of AI in higher education, specifically mentioning ChatGPT. The report covers both the current and potential uses of AI in education and also uses ChatGPT as a tool for conducting research [53]. After interacting with the artificial intelligence writing tool ChatGPT, Ryznar's paper proposes several approaches for administering exams while keeping their integrity. Ryznar states that the chatbot's clever response, which students might employ in their written study, simply raises the stakes of determining how to conduct tests equitably [54].

Another usage area of ChatGPT is in the field of health. It is likely to be used especially as a decision support system. In this field, some studies have been published on both the studies conducted by students in the field of health and the use of ChatGPT in the field of health. ChatGPT performs well on medical question-answering tasks, with a performance comparable to a third-year medical student. It provides reasoning and context in its responses, making it a promising tool for medical education [55]. Jeblick et al. evaluated the use of ChatGPT, a large language model, in generating simplified radiology reports. Radiologists agreed that the reports were factually correct and complete, but identified incorrect statements and missing medical information. Further research is needed to validate the findings and explore the potential use of LLMs in medical domains. The study suggests using ChatGPT-generated reports alongside original reports, approved by experts, to improve patient-centered care in radiology [56].

With the release of GPT3 for free, it was realized that ChatGPT, whose capabilities were discovered by many people, could also be used in academic writings. "Can you write a scientific article on a given topic like a scientist? What is your main purpose in this matter?" ChatGPT's answer to a question can be seen in Figure 4 and text below.

"Yes, I can write a scientific article on a given topic as a scientist. My main purpose in doing so would be to present the results of a study or research in a clear, concise, and objective manner, following the conventions and standards of scientific writing. This would include an introduction to the topic, a literature review, a description of the methodology used, an analysis of the results, and a discussion of the conclusions and implications of the study. The goal is to communicate the findings of the research to a wider scientific audience, and to contribute to the advancement of knowledge in the field."



Figure 5. The answer of the ChatGPT to our questions

As can be seen in Figure 5, writing academic articles with ChatGPT has some conveniences and advantages, and it will also bring many problems. There are many studies on writing and academic writing in the literature so far. For example, Shope offers advice on how to disclose the use of AI technologies in legal writing, specifically in law review papers and law school courses in his research. It gives a starting point for disclosure and reference by using ChatGPT as an example. In their rules for AI tool usage and disclosure in writing, the author gives policies for professors and publications to use [57]. In her other study, in which some of her fields were written with ChatGPT, she said that people could not understand whether it was a human or artificial intelligence product, and this was surprising. Shows the difference between mechanical and sophisticated writing in the article "Socratic Dialogue with ChatGPT" [58]. Frye states that there is little risk in using an AI text generator to generate answers to queries. If students can give satisfactory answers to your questions using an AI text generator, then you are asking superficial questions. Also, if an AI text generator can compete with a human's scholarship, it indicates that the individual is a shallow thinker [41]. The ramifications of the AI tool ChatGPT for legal writing are discussed in Armstrong's paper. The author put ChatGPT to the test in terms of legal research, writing in legal writing frameworks, and accuracy. The article also discusses the ethical implications of using ChatGPT in legal practice and how it could be used as a learning tool in legal writing education. This article serves as a starting point for future studies on the impact of AI on legal writing [59]. Lund and Wang introduce ChatGPT and its underlying technology, GPT. It highlights GPT's history and technology, including its generative pre-trained transformer model and capacity to execute a variety of language-based tasks. The report includes an interview with ChatGPT about its possible impact on academia and libraries, as well as a discussion of the benefits and ethical concerns. Improved search and discovery, reference services, categorization, and content production are among the advantages. The paper also looks into using ChatGPT to write scholarly papers [60]. Willems included some information, confusion, and various trials about the use of ChatGPT in academic literature, in his draft paper. He especially tried to give examples of citations [61]. Uludağ's goal is to talk about the application of an AI chatbot, ChatGPT, in psychology. A review of eight studies was done using the PubMed keyword "ChatGPT". The results demonstrated that ChatGPT can create creative articles that are difficult to differentiate from human writing, but no study examined its impact on psychology. The conclusion is that ChatGPT has potential applications in various domains, including psychology [62]. Kutela et al. wanted to compare ChatGPT's ability to generate a

manuscript for publishing against actual published content. It used supervised text classifiers and unsupervised text mining algorithms to evaluate ChatGPT-generated introductions to 327 published articles on traffic safety. The results revealed a substantial difference between ChatGPT-generated and human-generated introductions, with higher accuracy of supervised text classifiers, but inconsistent results in keyword frequency and cluster analysis. The work contributes to a broader understanding of the use of advanced language models in scientific writing [63]. Jiao et al. evaluate ChatGPT's machine translation capabilities, such as translation quick, multilingual translation, and robustness in their paper. They stated that ChatGPT-recommended prompts have been found to work effectively in general, with small performance variances. Also, they claimed that ChatGPT outperforms commercial systems in high-resource European languages but falls short in low-resource or distant languages. Moreover, they said that it performs lower than commercial systems for biomedical abstracts and Reddit comments in terms of robustness, but it has the potential to be a good translation for spoken English [64]. Aljanabi et al. mentioned the role of ChatGPT in different fields in their studies. In the study, evaluations were made on topics such as Academic writing, ChatGPT as a search engine, Coding, Detecting security vulnerabilities, and social media [65]. Pavlik's paper which is co-written by a human journalism professor and ChatGPT explores the free generative AI platform named ChatGPT. It demonstrates ChatGPT's capacity and limitations in the journalism and media field, highlighting its impressive range of knowledge in journalism and media history, scholars, and issues. However, it also mentions its limitations in critical thinking and creativity. The essay raises concerns about ChatGPT posing a threat to human journalists and media professionals due to its ability to pass the Turing Test. Further research is needed to assess and critically examine generative AI systems such as ChatGPT for journalism and media education. Educators should consider incorporating generative AI in their courses, training students in its effective use, and addressing the ethical and biased implications for academic integrity [66]. Gao et al. claimed that ChatGPT can generate realistic scientific abstracts, but the data is completely generated. The generated abstracts are original but can be detected by AI detectors and human reviewers. Scientific journals and medical conferences need to update their policies to maintain scientific standards and include AI detectors in the editorial process and require disclosure if these technologies are used. The ethical and acceptable use of language models for scientific writing is still being determined [67]. Marti gives an in-depth look at the use of machine learning in quantitative trading, covering key concepts, techniques, and applications in his book. ChatGPT wrote most of the book. It discusses difficulties and best practices in areas such as feature engineering, model selection, and backtesting. The book addresses several machine learning methodologies, such as supervised learning, unsupervised learning, and reinforcement learning, in addition to the difficulties associated with implementing these models in the real world. The book contains examples and case studies, as well as practical suggestions and resources to assist traders and practitioners in getting started with machine learning for quantitative trading [68]. Aydın and Karaarslan investigate the possibility of using AI, specifically OpenAI's ChatGPT

model, to summarize literature review articles. The study focuses on Digital Twin applications in healthcare and uses ChatGPT to paraphrase abstracts from recent studies. The study's findings were promising, with significant similarities discovered when examined with a plagiarism detection tool, but they also underlined the need for additional developments in AI technology. According to the study, artificial intelligence has the potential to transform academic publications by lowering the amount of human effort required and allowing academics to focus on their research [69]. In another study, Aydın et al. questioned the use of artificial intelligence and other popular technologies in human resources processes. [70].

Another usage area of ChatGPT is Business, Finance, and Economy. Studies in this area or related have also taken place in the literature. Dowling and Lucey conclude that ChatGPT can considerably improve finance research based on evaluations from finance journal reviewers. The findings are thought to be generalizable across research disciplines. ChatGPT performs well in terms of idea generation and data identification but falls short in terms of literature synthesis and testing framework construction. The amount of private data and researcher competence used determines the outcome quality. The paper finishes with a discussion of the new technology's ethical implications [73]. Mertens stated that well-known AI applications such as ChatGPT are just a few instances of how AI may be used in business. In his paper, he analyzes the possible applications of AI in the business world, specifically in corporate governance. According to the report, AI can improve decision-making and challenge groupthink, but company law frameworks still consider decision-making as exclusively human, causing uncertainty about AI adoption in the corporate realm. The purpose of this study is to investigate the expanding usage of AI in corporate governance by categorizing its levels of autonomy and exploring the prospective roles of AI in various types of organizations. The study focuses on major legal issues concerning AI decision-making delegation, human oversight, and accountability for algorithmic failure [74]. Zaremba and Demir examine the usage of ChatGPT technology in finance and how it might be used to improve NLP-based financial applications. They discuss ethical and regulatory issues as well as future research directions. According to the literature, ChatGPT has the potential to improve financial applications, but it poses ethical and regulatory concerns. To ensure responsible usage of ChatGPT in finance, the paper underlines the significance of research in robustness, interpretability, and ethics [75]. Chesterman examines the impact of these technologies on the knowledge economy, in which knowledge workers earn a living through analysis and writing. The author stated that ChatGPT can quickly and cheaply repeat this process, raising problems about who owns the content generated by generative AI and the necessity to identify AI-generated content [76]. Using economics and finance as examples, Alshater investigates the application of AI, namely NLP, in boosting academic achievement. It examines ChatGPT as an example of an NLP tool and concludes that it has the potential to improve research in economics and finance. However, limitations such as dependency on data quality and variety, a lack of experience, a limited comprehension of context, and ethical considerations must be considered when

utilizing AI tools such as ChatGPT, which should be used in conjunction with human analysis [77]. Wenzlaff and Sebastian investigate the accuracy of OpenAI ChatGPT model replies for the definitions of crowdfunding, alternative finance, and community finance. These terms were chosen due to inconsistencies in academic literature definitions, but some are accepted by more researchers. In academic scholarship, the model's responses are compared to those of real people. The purpose of their study is to contribute to the expanding body of literature on the implications of AI for academia [78]. Leippold tried to attack sentiment analysis using GPT3. He used an old database of 4,837 English news headlines of financial-related companies listed on the OMX Helsinki stock exchange. As a result, it is shown that sentiment analysis results were able to be manipulated [79].

A judge in Colombia caused controversy after using the AI tool ChatGPT in a case involving a child with autism. The judge, Juan Manuel Padilla from Cartagena, determined that the child's medical plan should pay for all expenses as the parents were unable to do so and cited earlier resolutions as a precedent for his conclusion. The verdict itself was not disputed, but the use of ChatGPT in the ruling sparked outrage [80]. As can be understood from this event, ChatGPT has now made a name for itself in the field of law. While its use in the field of law has begun, there are also some academic discussions in the literature on this subject. Lea Bishop tried to understand whether ChatGPT, which has recently been tested for use in many areas and whose effectiveness has been evaluated, can think like a Lawyer through a Socratic dialogue. She asked some questions and then shared ChatGPT's answers in her draft paper. She concluded that ChatGPT was quite good at IRAC analysis. She called it bad because law students must spend weeks producing the content that ChatGPT produces in 10 seconds. She predicts that law firms will understand that AI can write very good analytical legal notes without a salary or health insurance. On the other hand, thinking like a lawyer requires a lot of things like critical thinking skills that can be applied more broadly. That's why she stated that in the age of ChatGPT, "thinking like a lawyer" can be a matter of having the ability to ask good questions [81]. Iu and Wong evaluate the capacity of ChatGPT, an OpenAI AI language model, to replace litigation lawyers. The results reveal that ChatGPT has sophisticated legal drafting skills for a variety of papers and can assess facts, provide legal bases, develop tactics, construct arguments, and provide easy recommendations. However, data source restrictions result in a deficit in detecting current case laws, hence the study implies that ChatGPT should be viewed as a supplement rather than a replacement for lawyers [82]. In their article, Choi et al. assess the capacity of the AI model ChatGPT to write law school exams without human aid. The model was used to generate answers for four genuine exams at the University of Minnesota Law School, and the answers were graded as part of the standard grading process. ChatGPT performed at the level of a C+ student on average, passing all four courses with a low grade. The results are discussed in terms of their implications for legal education and practice, with example prompts and advice on using ChatGPT for legal writing provided [83].

General Motors Company has announced that it has decided to conduct more extensive research on the usage areas of ChatGPT. GM partnered with Microsoft in 2021 to accelerate the commercialization of autonomous vehicles. GM Vice President Scott Miller stated that ChatGPT is likely to be involved in everything. Scott Miller also shared some details about the areas where ChatGPT can be used [84].

5. RESULTS AND DISCUSSION

OpenAI ChatGPT's models are based on Google's transformer models. Google labs made these models public, but the company didn't integrate these models into their products. OpenAI ChatGPT's success depends on its wide usage as it is open to public use. ChatGPT's popularity and Microsoft's support is a game changer. These are forcing companies like Google to change their business plans.

Open is mostly used for open source and nonprofit. OpenAI was meant to be a nonprofit firm, but ChatGPT is evolving to be a commercial product. Companies like Google, Microsoft, Amazon, and IBM also offer similar models and services for building conversational interfaces.

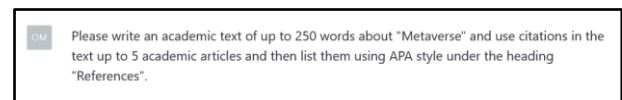


Figure 6. Metaverse question asked to ChatGPT

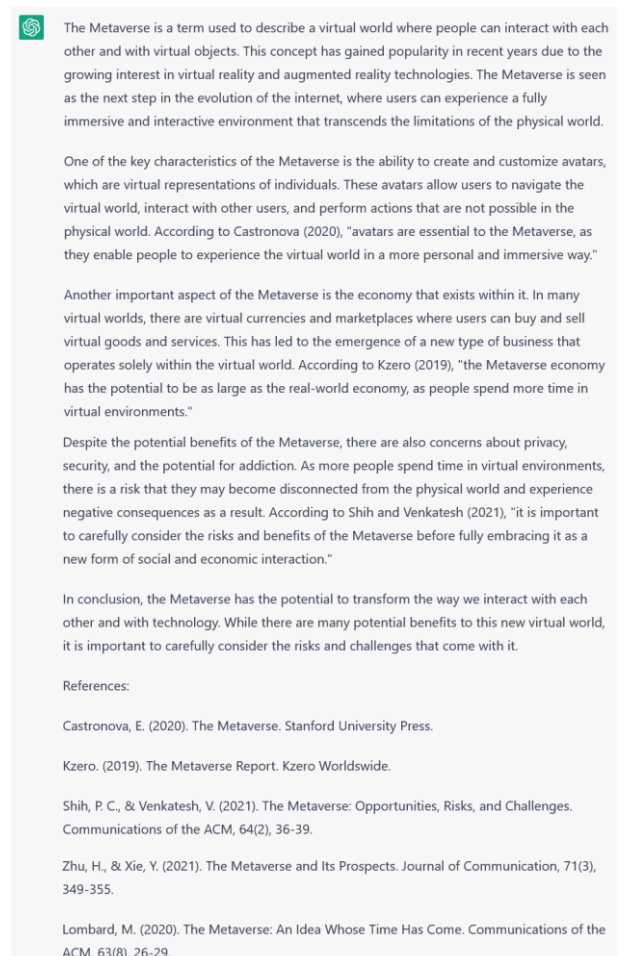


Figure 7. ChatGPT's answer to the question given in Figure 6

The chatbot industry is constantly evolving, and even small improvements in the user interface, algorithms, or training data can bring significant gains. Continuous learning and the ability to use relevant data to connect with specific customers is the key to successful chatbots.

There are examples in social media where ChatGPT gave wrong answers. Some of the applications are combining their machine learning models to evade these wrong answers. Bard also gave a wrong answer in the first public demo about the very first picture of a planet outside of our own solar system [85]. These all show that we are still at the very start of this journey.

When we use ChatGPT to write an academic text on a specific subject, including references, ChatGPT presents a general text close to the truth. It also gives citations in the text with the format you request (APA, etc.). It also produces and presents the reference list with these citations. We asked ChatGPT to write an academic text shown in Figure 6.

As you can see in Figure 7, ChatGPT returned with a text with references as follows:

- *Castronova, E. (2020). The Metaverse. Stanford University Press.*
- *Kzero. (2019). The Metaverse Report. Kzero Worldwide.*
- *Shih, P. C., & Venkatesh, V. (2021). The Metaverse: Opportunities, Risks, and Challenges. Communications of the ACM, 64(2), 36-39.*
- *Zhu, H., & Xie, Y. (2021). The Metaverse and Its Prospects. Journal of Communication, 71(3), 349-355.*
- *Lombard, M. (2020). The Metaverse: An Idea Whose Time Has Come. Communications of the ACM, 63(8), 26-29.*

However, when you search for the listed studies in the references, you can see that some of these references point to unpublished work. ChatGPT made some of them up. When we asked the reason for this to ChatGPT by using the sentence “*There are no published articles or materials as you listed under the references heading. Why do you give that wrong and inexistent references?*”, ChatGPT generates an answer shown in Figure 8. It generates “*I apologize for the error in my previous response. It appears that the references I provided were not accurate and did not exist. As an AI language model, I strive to provide accurate and reliable information. In this case, I failed to meet that standard.*”

To provide you with accurate references on the topic of the Metaverse, here are some relevant sources that you may find useful:

1. *Castronova, E. (2020). The Metaverse: A Very Short Introduction. Oxford University Press.*
2. *Damer, B. (2018). Avatars! Exploring and Building Virtual Worlds on the Internet. Peachpit Press.*

3. *He, W., & Feng, W. (2021). The Metaverse: A New Digital Era of Virtual Reality. Information, 12(4), 152.*
4. *Kzero. (2021). The Metaverse Report. Kzero Worldwide.*
5. *Rosedale, P. (2019). The Metaverse: What It Is, Where to Find it, Who Will Build It, and Fortnite. Journal of Virtual Worlds Research, 12(1).*

Once again, I apologize for any confusion caused by my earlier response and I hope that these references will be helpful to you.”



Figure 8. Question and answer about the previous conversation

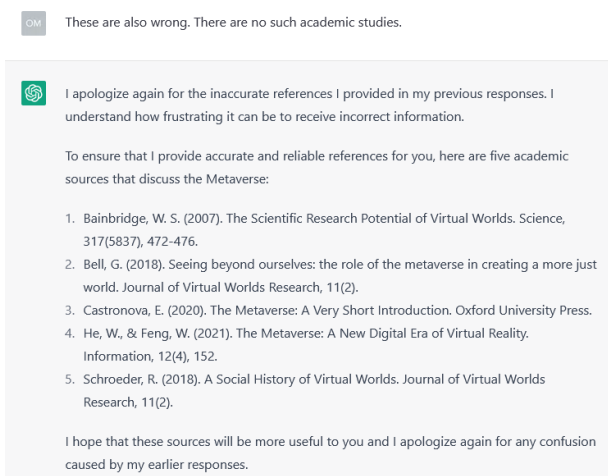


Figure 9. Question and answer about the previous conversation



Figure 10. ChatGPT's response when asked who a person is

We would like to emphasize that ChatGPT may give different responses on even repetitive questions as it is constantly learning with each question and response. As can be seen in Figure 9, when we repeat the question, the references it gives are updated but still they are incorrect and inexistent.

We asked who Assistant Professor Ömer Aydın (the first author of this article) is, and ChatGPT generated an answer as in Figure 10. The answer seems logical at first, but the information contains inconsistencies and mistakes. Assistant Professor Ömer AYDIN is a faculty member at the Engineering Faculty of Manisa Celal Bayar University. His research interests are different. He received his Ph.D. in Computer Engineering in 2019 from Dokuz Eylül University in Türkiye. According to the academic information system of Türkiye (<http://akademik.yok.gov.tr>), also no similar person complies with this information. It is clear that ChatGPT fabricated this information.

GPT has been evolving since its first version to GPT-4. Its model architecture, size, and dataset are not announced and there are only speculations about the parameter sizes. Different products, even different versions of GPT are trained with different datasets of different sizes. The number of training tokens is also important in the performance. GPT-4 is assumed to use a smaller model for better performance [86].

6. CONCLUSION

In this study, firstly fundamentals about NLP and Generative AI are given. Afterwards, we shared basic information about ChatGPT. We also included ChatGPT's current rivals. Then, we presented the studies on ChatGPT by grouping them according to their fields such as law, education, academic writing, and finance. Finally, we have included comparisons and evaluations between its rivals and ChatGPT on technical and structural issues.

With its free launch, OpenAI ChatGPT has suddenly become the most popular among Generative AI tools. Being the most popular Generative AI tool does not mean that ChatGPT is the best tool. There are many alternative ChatRobots or Generative AI tools, and they seem to lag behind ChatGPT. Of course, we cannot know what will happen in the future, but we know that Generative AI tools will be at the center of our lives. We will surely use Artificial Intelligence applications more in daily life.

This paper gives evidence that shows we are still at the very start of this journey. The presence of a considerable amount of wrong answers is one of them. ChatGPT and similar systems can make up wrong answers. These wrong answers mainly depend on the weakness of the used dataset and algorithms. So, humanity should use these tools as supportive material keeping in mind that their dataset does not cover all information. However, these systems will surely evolve in time and their wrong answers will decrease. New versions started to give references to the used data.

OpenAI ChatGPT and similar platforms started with public access, and this caused their popularity. Their development and popularity can change when they focus on paid access.

Even though the usage of LLMs as a chatbot gained a wide range of attraction, we should be aware of their limits. We still can not train these models with all of our data since then these models become inefficient. Companies can deploy open LLMs but the cost of training such a model is still very expensive, and we should also be aware of its environmental impact. Also, ethical, legal, economic and sociological difficulties will determine the future of these tools. We should be aware of privacy rights and licences. We will see different LLMS which focus on specific areas. Custom LLMS are announced which focus on different areas such as finance or law.

We should be aware that AI systems can be misused in cybersecurity and terrorism threats. Unrestrained AI development can cause serious, even “catastrophic” societal problems as Yoshua Bengio has said. We need ways to ensure the usage of these tools is safe and under human control [87]. We also need to understand if the content is AI-generated or not. Recently a forum is initiated with Google, Microsoft, OpenAI and Anthropic to promote safe AI development and it is a good step forward. We need legislation which limits the use of AI in risk scenarios.

As a future study, we will continue to evaluate upcoming generative AI platforms and new use cases. We will work on analyzing new generative AI approaches. Such new evaluations can be made by monitoring the status and development of competitors over time.

Author contributions: OA And EK co-authored the article. OA made ChatGPT queries. Tables were prepared by OA and graphs by EK. EK carried out English language checks. Figures are adapted by OA and revisions made by EK and OA.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study has received no financial support.

REFERENCES

- [1] Turing, A. M. (1951). Can digital computers think?. The Turing Test: Verbal Behavior as the Hallmark of Intelligence, 111-116.
- [2] IBM (2023). Natural Language Processing (NLP). <https://www.ibm.com/in-en/topics/natural-language-processing>
- [3] Hiemstra, D. (2009). Language Models. In: LIU, L., ÖZSU, M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_923
- [4] Jelinek, F. (1998). Statistical methods for speech recognition. MIT press.
- [5] Brants, T., Popat, A. C., Xu, P., Och, F. J., & Dean, J. (2007). Large language models in machine translation.


- [6] Smith, S., Patwary, M., Norrick, B., LeGresley, P., Rajbhandari, S., Casper, J., ... & Catanzaro, B. (2022). Using deepspeed and megatron to train megatron-turing nlg 530b, a large-scale generative language model. arXiv preprint arXiv:2201.11990.
- [7] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets in advances in neural information processing systems (NIPS). Curran Associates, Inc. Red Hook, NY, USA, 2672-2680.
- [8] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139-144.
- [9] Erdoğan, A. (2019) GANs Nedir? Nasıl Çalışır? DevNot. <https://devnot.com/2019/gans-nedir-nasil-calisir/>
- [10] Gharakhanian, A. (2017). "Generative Adversarial Networks – Hot Topic in Machine Learning", <https://www.kdnuggets.com/2017/01/generative-adversarial-networks-hot-topic-machine-learning.html>
- [11] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
- [12] Sim, A. [@aaronstim]. (2022, October 29). Generative AI. Twitter. <https://twitter.com/aaronstim/status/1586412346817187840?s=20>
- [13] Sim, A. [@aaronstim]. (2022, November 10). Generative AI Models. Twitter. <https://twitter.com/aaronstim/status/1590474215383531520?s=20>
- [14] ChatGPT (2023). <https://chat.openai.com/>. Access date: 04.07.2023
- [15] Wu, C., Yin, S., Qi, W., Wang, X., Tang, Z., & Duan, N. (2023). Visual ChatGPT: Talking, Drawing and Editing with Visual Foundation Models. arXiv preprint arXiv:2303.04671.
- [16] OpenAI (2023). GPT-4 Technical Report. <https://cdn.openai.com/papers/gpt-4.pdf>
- [17] Collins, E., Ghahramani, Z. (2021) LaMDA: our breakthrough conversation technology. Google Blogs. <https://blog.google/technology/ai/lamda/>
- [18] Editor, M. (2023) Brad Vs ChatGPT: How Does It Work? MFAME Guru. <https://mfame.guru/brad-vs-chatgpt-how-does-it-work/>
- [19] Aydın, Ö. (2023). Google Bard Generated Literature Review: Metaverse. *Journal of AI*. 7(1), 1-14.
- [20] Rahsepar, A. A., Tavakoli, N., Kim, G. H. J., Hassani, C., Abtin, F., & Bedayat, A. (2023). How AI Responds to Common Lung Cancer Questions: ChatGPT vs Google Bard. *Radiology*, 307(5), e230922.
- [21] West, D. M. (2023). Comparing Google Bard with OpenAIs ChatGPT on political bias, facts, and morality.
- [22] Le, T. T. (2023). Google Bard talking about applying mindsponge theory to ADHD research.
- [23] Goodside, R., Papay, S. (2023) Meet Claude: Anthropic's Rival to ChatGPT. Scale.com Blog. <https://scale.com/blog/chatgpt-vs-claude>
- [24] Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., ... & Kaplan, J. (2022). Constitutional AI: Harmlessness from AI Feedback. arXiv preprint arXiv:2212.08073.
- [25] Shetty, S. (2018) Facebook's Wit.ai: Why we need yet another chatbot development framework? Packtpub. <https://hub.packtpub.com/facebooks-wit-ai-why-we-need-yet-another-chatbot-development-framework/>
- [26] Handoyo, E., Arfan, M., Soetrisno, Y. A. A., Somantri, M., Sofwan, A., & Sinuraya, E. W. (2018, September). Ticketing chatbot service using serverless NLP technology. In 2018 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE) (pp. 325-330). IEEE.
- [27] Ye, J. (2023). China's Tencent establishes team to develop ChatGPT-like product -sources. Reuters. <https://www.reuters.com/technology/chinas-tencent-sets-up-team-develop-chatgpt-like-product-sources-2023-02-27/>
- [28] Altamira Expert Team. (2023). 9 Best Open-Source Alternatives to GPT-3 You Should Try. Altamira.ai. <https://www.altamira.ai/blog/open-source-gpt-alternative-solutions?fbclid=IwAR2FDSsHXVQL85JS26hZU43U3keDnKjkg8hYyZVUOz1OkKkX70keNck51nw>
- [29] Benzon, W.L. (2023) ChatGPT Intimates a Tantalizing Future; Its core LLM is Organized on Multiple Levels; and it has Broken the Idea of Thinking. Available at SSRN: <https://ssrn.com/abstract=4336442>
- [30] Benzon, W.L. (2023) Discursive Competence in ChatGPT, Part 1: Talking with Dragons Version 2.2. Available at SSRN: <https://ssrn.com/abstract=4318832>
- [31] Benzon, W.L. (2023) A Note about Story Grammars in ChatGPT. Available at SSRN: <https://ssrn.com/abstract=4324840>
- [32] Sætra, H.S. (2023). Generative Ai: Here to Stay, But for Good?. Available at SSRN: <https://ssrn.com/abstract=4330244>
- [33] Shah, C. (2022). The Rise of AI Chat Agents and the Discourse with Dilettantes. *Information Matters*, Vol. 2, No. 12, 2022, Available at SSRN: <https://ssrn.com/abstract=4327315>
- [34] Lund, B., Agbaji, D. (2023). Information Literacy, Data Literacy, Privacy Literacy, and ChatGPT: Technology Literacies Align with Perspectives on Emerging Technology Adoption within Communities. Available at SSRN: <https://ssrn.com/abstract=4324580> or <http://dx.doi.org/10.2139/ssrn.4324580>


- [35] Bommarito, J., Bommarito, M.J., Katz, J., Katz, D.M. (2023) Gpt as Knowledge Worker: A Zero-Shot Evaluation of (AI)CPA Capabilities. Available at SSRN: <https://ssrn.com/abstract=4322372> or <http://dx.doi.org/10.2139/ssrn.4322372>
- [36] Hartmann, J., Schwenzow, J., Witte, M. (2023). The political ideology of conversational AI: Converging evidence on ChatGPT's pro-environmental, left-libertarian orientation. Available at SSRN: <https://ssrn.com/abstract=4316084> or <http://dx.doi.org/10.2139/ssrn.4316084>
- [37] Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). ChatGPT is not all you need. A State of the Art Review of large Generative AI models. arXiv preprint arXiv:2301.04655.
- [38] Azaria, A. (2022). ChatGPT Usage and Limitations. hal-03913837. Available at <https://hal.science/hal-03913837/>
- [39] ChatGPT, Perlman, A. (2022). The Implications of OpenAI's Assistant for Legal Services and Society. Available at SSRN: <https://ssrn.com/abstract=4294197> or <http://dx.doi.org/10.2139/ssrn.4294197>
- [40] Jabotinsky, H.Y., Sarel, R. (2022). Co-authoring with an AI? Ethical Dilemmas and Artificial Intelligence. Available at SSRN: <https://ssrn.com/abstract=4303959> or <http://dx.doi.org/10.2139/ssrn.4303959>
- [41] Frye, B.L. (2022). Should Using an AI Text Generator to Produce Academic Writing Be Plagiarism?. Fordham Intellectual Property, Media & Entertainment Law Journal, Forthcoming, Available at SSRN: <https://ssrn.com/abstract=429228>
- [42] Hacker, P. (2022). The European AI Liability Directives – Critique of a Half-Hearted Approach and Lessons for the Future. Available at SSRN: <https://ssrn.com/abstract=4279796> or <http://dx.doi.org/10.2139/ssrn.4279796>
- [43] Baidoo-anu, D. & Owusu Ansah, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. Journal of AI, 7 (1), 52-62. Retrieved from <https://dergipark.org.tr/tr/pub/jai/issue/77844/1337500>
- [44] Zhai, X. (2022). ChatGPT User Experience: Implications for Education. Available at SSRN: <https://ssrn.com/abstract=4312418> or <http://dx.doi.org/10.2139/ssrn.4312418>
- [45] Zhai, X. (2023). ChatGPT for Next Generation Science Learning. Available at SSRN: <https://ssrn.com/abstract=4331313>
- [46] Ventayen, R.J.M. (2023). OpenAI ChatGPT Generated Results: Similarity Index of Artificial Intelligence-Based Contents. Available at SSRN: <https://ssrn.com/abstract=4332664>
- [47] Avila-Chauvet, L. and Mejía, D., Acosta Quiroz, C.O. (2023) Chatgpt as a Support Tool for Online Behavioral Task Programming. Available at SSRN: <https://ssrn.com/abstract=4329020>
- [48] Nisar, S., Aslam, M.S. (2023). Is ChatGPT a Good Tool for T&CM Students in Studying Pharmacology? Available at SSRN: <https://ssrn.com/abstract=4324310>
- [49] Bommarito, M.J., Katz, D.M. (2022). GPT Takes the Bar Exam. Available at SSRN: <https://ssrn.com/abstract=4314839> or <http://dx.doi.org/10.2139/ssrn.4314839>
- [50] Susnjak, T. (2022). ChatGPT: The End of Online Exam Integrity? arXiv preprint arXiv:2212.09292.
- [51] Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and Cheating: Ensuring academic integrity in the era of ChatGPT. Preprint. <https://doi.org/10.35542/osf.io/mrz8h>.
- [52] Mellon, J., Bailey, J., Scott, R., Breckwoldt, J., Miori, M. (2022). Does GPT-3 know what the Most Important Issue is? Using Large Language Models to Code Open-Text Social Survey Responses at Scale. Available at SSRN: <https://ssrn.com/abstract=4310154> or <http://dx.doi.org/10.2139/ssrn.4310154>
- [53] Zentner, A. (2022). Applied Innovation: Artificial Intelligence in Higher Education. Available at SSRN: <https://ssrn.com/abstract=4314180> or <http://dx.doi.org/10.2139/ssrn.4314180>
- [54] Ryznar, M. (2020). Exams in the Time of ChatGPT. Washington and Lee Law Review Online, Forthcoming, Available at SSRN: <https://ssrn.com/abstract=3684958> or <http://dx.doi.org/10.2139/ssrn.3684958>
- [55] Gilson, A., Safranek, C., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2022). How Well Does ChatGPT Do When Taking the Medical Licensing Exams? The Implications of Large Language Models for Medical Education and Knowledge Assessment. medRxiv, 2022-12.
- [56] Jeblick, K., Schachtner, B., Dextl, J., Mittermeier, A., Stüber, A. T., Topalis, J., ... & Ingrisich, M. (2022). ChatGPT Makes Medicine Easy to Swallow: An Exploratory Case Study on Simplified Radiology Reports. arXiv preprint arXiv:2212.14882.
- [57] Shope, M. (2023). Best Practices for Disclosure and Citation When Using Artificial Intelligence Tools. Available at SSRN: <https://ssrn.com/abstract=4338115>
- [58] Bishop, L. (2023) A Computer Wrote this Paper: What ChatGPT Means for Education, Research, and Writing. Available at SSRN: <https://ssrn.com/abstract=4338981> or <http://dx.doi.org/10.2139/ssrn.4338981>
- [59] Armstrong, A.B. (2023) Who's Afraid of ChatGPT? An Examination of ChatGPT's Implications for Legal Writing. Available at SSRN: <https://ssrn.com/abstract=4336929>
- [60] Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? Library Hi Tech News.
- [61] Willems, J. (2023). ChatGPT at Universities – The Least of Our Concerns. Available at SSRN: <https://ssrn.com/abstract=4334162>


- [62] Uludağ, K. (2023). The Use of AI-Supported Chatbot in Psychology. Available at SSRN: <https://ssrn.com/abstract=4331367>
- [63] Kutela, B., Msechu, K., Das, S., Kidando, E. (2023). Chatgpt's Scientific Writings: A Case Study on Traffic Safety. Available at SSRN: <https://ssrn.com/abstract=4329120>
- [64] Jiao, W., Wang, W., Huang, J. T., Wang, X., & Tu, Z. (2023). Is ChatGPT a good translator? A preliminary study. arXiv preprint arXiv:2301.08745.
- [65] Aljanabi, M., Ghazi, M., Ali, A. H., & Abed, S. A. (2023). ChatGpt: Open Possibilities. *Iraqi Journal for Computer Science and Mathematics*, 4(1), 62-64.
- [66] Pavlik, J. V. (2023). Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *Journalism & Mass Communication Educator*, 10776958221149577.
- [67] Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2022). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *bioRxiv*, 2022-12.
- [68] Marti, G. (2022). From Data to Trade: A Machine Learning Approach to Quantitative Trading. Available at SSRN: <https://ssrn.com/abstract=4315362> or <http://dx.doi.org/10.2139/ssrn.4315362>
- [69] Aydın, Ö., Karaarslan, E. (2022). OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare. In Ö. Aydın (Ed.), *Emerging Computer Technologies 2* (pp. 22-31). İzmir Akademi Derneği.
- [70] Aydın, Ö., Karaarslan, E., & Dutta, P. (2023). Artificial Intelligence, VR, AR and Metaverse Technologies for Human Resources Management. *VR, AR and Metaverse Technologies for Human Resources Management* (June 15, 2023).
- [71] Aydın, Ö., Karaarslan, E. (2023). OpenAI ChatGPT interprets Radiological Images: GPT-4 as a Medical Doctor for a Fast Check-Up.
- [72] Alıcı, U. I., Oksuztepe, A., Kilincceker, O., & Karaarslan, E. (2023). OpenAI ChatGPT for Smart Contract Security Testing: Discussion and Future Directions. Available at SSRN 4412215.
- [73] Dowling, M.M., Lucey, B.M. (2023). ChatGPT for (Finance) Research: The Bananarama Conjecture. Available at SSRN: <https://ssrn.com/abstract=4322651> or <http://dx.doi.org/10.2139/ssrn.4322651>
- [74] Mertens, F. (2023). The Use of Artificial Intelligence in Corporate Decision-Making at Board Level: A Preliminary Legal Analysis. *Financial Law Institute Working Paper Series 2023-01*, Available at SSRN: <https://ssrn.com/abstract=4339413> or <http://dx.doi.org/10.2139/ssrn.4339413>
- [75] Zaremba, A., Demir, E. (2023). ChatGPT: Unlocking the Future of NLP in Finance. Available at SSRN: <https://ssrn.com/abstract=4323643>
- [76] Chesterman, S. (2023). AI-Generated Content is Taking over the World. But Who Owns it?. *NUS Law Working Paper No. 2023/002*, Available at SSRN: <https://ssrn.com/abstract=4321596> or <http://dx.doi.org/10.2139/ssrn.4321596>
- [77] Alshater, M. (2022) Exploring the Role of Artificial Intelligence in Enhancing Academic Performance: A Case Study of ChatGPT. Available at SSRN: <https://ssrn.com/abstract=4312358> or <http://dx.doi.org/10.2139/ssrn.4312358>
- [78] Wenzlaff, K., Spaeth, S. (2022). Smarter than Humans? Validating how OpenAI's ChatGPT Model Explains Crowdfunding, Alternative Finance and Community Finance. Available at SSRN: <https://ssrn.com/abstract=4302443> or <http://dx.doi.org/10.2139/ssrn.4302443>
- [79] Leippold, M. (2023) Sentiment Spin: Attacking Financial Sentiment with GPT-3. *Swiss Finance Institute Research Paper No. 23-11*. pp 1-11. Available at SSRN: <https://ssrn.com/abstract=4337182> or <http://dx.doi.org/10.2139/ssrn.4337182>
- [80] Taylor, L. (2023). Colombian judge says he used ChatGPT in ruling. *The Guardian*. <https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling>
- [81] Bishop, L. (2023) Can ChatGPT 'Think Like a Lawyer?' A Socratic Dialogue. Available at SSRN: <https://ssrn.com/abstract=4338995> or <http://dx.doi.org/10.2139/ssrn.4338995>
- [82] Iu, K.Y., Wong, V.M. (2023). ChatGPT by OpenAI: The End of Litigation Lawyers?. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4339839
- [83] Choi, J.H., Hickman, K.E., Monahan, A., Schwarcz, D.B. (2023). ChatGPT Goes to Law School. Available at SSRN: <https://ssrn.com/abstract=4335905>
- [84] Cook, R. (2023) GM explores using ChatGPT in vehicles. *Reuters*. <https://www.cnbc.com/cdn.ampproject.org/c/s/www.cnbc.com/amp/2023/03/10/gm-explores-using-chatgpt-in-vehicles.html>
- [85] Verge, (2023). Google's AI chatbot Bard makes factual error in first demo, <https://www.theverge.com/2023/2/8/23590864/google-ai-chatbot-bard-mistake-error-exoplanet-demo>
- [86] Radar, (2022). Everything We Know About GPT-4, <https://www.datacamp.com/blog/what-we-know-gpt4>
- [87] Fung, B. (2023). Google, Microsoft, OpenAI and Anthropic announce industry group to promote safe AI development. *CNN*. <https://edition.cnn.com/2023/07/26/tech/ai-industry-group/index.html>

Tabu Search with Variable Neighborhood Search Algorithm for Home Healthcare Routing Problem for Multiple Hospitals with Balanced Workload

*¹Gülçin DİNÇ YALÇIN, ²Tuğçe YAVUZ, ³Şüheda ALTINTAŞ

¹Department of Industrial Engineering, Eskisehir Technical University, Eskişehir, Türkiye, gdinc@eskisehir.edu.tr 

²Department of Industrial Engineering, Eskisehir Technical University, Eskişehir, Türkiye, yavuz_tgc@hotmail.com 

³Department of Industrial Engineering, Eskisehir Technical University, Eskişehir, Türkiye, suhedaaltintas@gmail.com 

Abstract

In this paper, we study home healthcare routing and scheduling problem where multiple hospitals serve patients. In the public hospitals in healthcare system of Türkiye, patients requiring home healthcare are assigned to the hospital that serves their place of residence. This can cause the workload of hospitals to become unbalanced in terms of the time needed for both traveling and operation. The aim of this paper is to generate routes with a balanced workload for hospitals, giving consideration to the time windows of patients and the working hours of health workers. Firstly, we construct a mathematical model which can solve toy and small-scale problems whilst taking into account the importance of a balanced workload. Then, a Tabu Search with a Variable Neighborhood Search (TS-VNS) algorithm is developed to solve large-scale problems. The performance of the TS-VNS algorithm is tested by comparing the results of the mathematical model with the generated test problems at a small scale. Additionally, large-scale test problems from the literature are sourced for the problem and solved by the TS-VNS algorithm. The results demonstrate the efficiency of the TS-VNS algorithm.

Keywords: home healthcare, tabu search, variable neighborhood search, workload balance, OR in healthcare

1. INTRODUCTION

There are many patients who are unable to go to hospitals for various reasons, such as old age, and home healthcare (HHC) has a significant impact on their well-being as it has been shown that life quality increases with HHC [1]. Since the population of people over 60 years is more than 1 billion and the rate of population aging is higher than ever before [2], HHC services will become increasingly important in the future. In fact, the size of the global HHC market was \$345.6 billion in 2022, and it is expected to grow at an annual rate of 7.9% from 2022 to 2030 [3].

In HHC services, the scheduling of appointments is complicated by the fact that a suitable time has to be found between two parties: the patients, with their preferred time windows, and the caregivers, such as nurses, doctors, physiotherapists, with their own hours of working and busy schedules. Additionally, all routes start from and end at the related hospital, pharmacy, laboratory, etc., but the order in which patients are visited (the routes) presents an optimization problem. All these operations in HHC service were first modeled as a Vehicle Routing Problem (VRP) with time windows by [4] (see also [5, 6, 7, 8, 9] for detailed survey). Generally, the objective is to minimize the total cost (time) of all the routes.

In the public hospitals HHC system of Türkiye, each hospital that has caregivers for HHC serves a predetermined area, and patients are assigned to whichever hospital covers their place of residence. Given that hospitals have different capacities and that each area hosts a varied number of patients, hospitals are often burdened with unbalanced workloads and this may impact the quality of HHC offered by the hospitals that handle relatively more patients. In an attempt to address this issue, we propose a new problem for HHC with multi hospital and time windows under balanced workload of objectives. This problem is referred to as the multi-hospital home-healthcare routing problem with balanced workload (MH-HHCRP-BW). The basic constraints of MH-HHCRP-BW are no route between hospitals, each route starts from and ends at the hospital, and each patient is served once. Additionally, the working hours and the time schedule for each patient is taken into account with time windows constraints.

Firstly, we develop a mathematical model for MH-HHCRP-BW. Then, since the model can solve the problem with a limited number of patients and hospitals, we propose a hybrid of a Tabu Search and Variable Neighborhood Search (TS-VNS) algorithm since VNS provides the flexibility of designing of neighborhood structures and TS prevents cycles in the search space.

Additionally, we generate some small problems to solve by both the mathematical model and TS-VNS. Then, we compare the performance of the TS-VNS with the solutions obtained from mathematical model. Finally, multi-depot VRP test problems with time windows are modified to generate a large-scale data set, and then, these problems are solved by TS-VNS.

The main contribution of this paper can be summarized as follows: 1) We create a multi-hospital HHC model. 2) We aim to generate solutions with balanced workloads among hospitals. 3) We model the problem as a mixed-integer mathematical model. 4) We propose a TS-VNS algorithm with a new insertion operator to solve the problem within a reasonable time. 5) We generate new small-scale test problems and modify the large-scale from the literature for the problem, which we refer to as MH-HHCRP-BW.

The rest of the paper is organized as follows: A literature review is given in Section 2. The mathematical model of the problem is described in detail in Section 3. The TS-VNS algorithm is explained in Section 4. The computational results are given in Section 5. Finally, some conclusions are drawn in Section 6.

2. LITERATURE REVIEW

Bredstöm and Rönnqvist [10] studied the formulation of a combined VRP with time windows and additional temporal constraints and presented a home healthcare problem as a VRP. Furthermore, they proposed balancing constraints for vehicles; however, they optimized balance service duration or traveling time and model the problem for one hospital. In this paper, we extend the problem for multiple hospitals and we balance the workload in both service duration and traveling time.

Lanzarone and Matta [11] dealt with home healthcare problem by only considering the patient-nurse assignment under the objective function-balancing maximum over workload of nurses. Each nurse has a predetermined workload and if this workload is not exceeded for each nurse, then the value of the objective function is zero whether the workload between nurses is balanced or not. Carello et al. [12] studied a patient-nurse assignment problem, where they addressed the problem from the perspectives of patients, nurses, and service providers. To solve the problem, they employed integer linear programming and utilized different objective functions to optimize the assignment process.

Yuan et al. [13] worked on the HHC problem in the case of patients' stochastic service times and caregivers' skill requirements. A stochastic model was proposed and, while column generation was used to solve the master problem, a label algorithm was developed to solve the pricing sub-problem. Rest and Hirsch [14] considered the problem with time-dependent public transport since most of the caregivers from the Austrian Red Cross in Vienna use a combination of public transport and walking. Then, they proposed a tabu search to solve the problem.

Shi et al. [15] dealt with a HHC problem with fuzzy demands related to the quantity of drugs required for each customer

and designed a fuzzy change constraint model. Additionally, a hybrid genetic algorithm integrated with stochastic simulation methods was developed. Masmoudi and Cheikhrouhou [16] considered HHC for one hospital with a heterogeneous fleet and a lunch break under the objective of cost minimization. They introduced a mathematical model and developed the Adaptive Large Neighborhood Search. Liu et al. [17] dealt with stochastic travel and service time and proposed a method that combines a branch-and-price algorithm and a discrete approximation method.

Bahadori-Chinibelagh et al. [18] proposed a multi-depot VRP with time windows model for HHC where depots are pharmacies. Each route starts from a pharmacy and ends at the related laboratory. The objective is to minimize the total cost. Additionally, two constructive heuristics were developed. Our study differs from it in terms of the objective function and the fact that each route starts from and ends at the related hospital.

Tanoumand and Ünlüyurt [19] considered new resource constraints: there are two types of personnel providing the service, but the number of personnel is limited. Then, they proposed an exact algorithm, the branch-and-price algorithm, to solve the problem. Li et al. [20] extended the problem for outpatient services and considered a new objective: minimizing the waiting times for outpatients. Then, they adopted an outer-approximation method and developed a hybrid genetic algorithm.

Besides these studies, in order to solve the HHC routing problem, Allaoua [21] proposed a matheuristic based on the decomposition of mathematical programming. Cappanera et al. [22] conducted a study that involves scheduling, assignment, and routing decisions for random requests. To address this complex problem, they employed a matheuristic approach. Frifita et al. [23] developed a general variable neighborhood search (see also [24] for a detailed survey about variable neighborhood search in healthcare management). Rahimian et al. [25] proposed a hybrid approach that combines integer programming and variable neighborhood search methods. Riazi et al. [26] studied decomposition and distributed algorithms, and Riazi et al. [27] proposed a gossip-column generation algorithm. Moussavi et al. [28] proposed a new mathematical formulation and then developed a matheuristic approach based on the decomposition of the mathematical formulation. Grenouilleau et al. [29] presented a set partitioning heuristic based on a set partitioning formulation and a large neighborhood search framework, and Grenouilleau et al. [30] developed new decomposition methods for predefined visits. Shahnejat-Bushehri et al. [31] conducted a study that involves random travel and transaction times. They employed three different metaheuristic algorithms (i.e. simulated annealing, genetic algorithm, and memtic algorithm) to address this problem with the objective of minimizing the total processing time. Dekhici et al. [32] solved the home-care problem by utilizing the firefly algorithm, which is a metaheuristic algorithm. They modeled the problem as a vehicle routing problem with time windows, where the objective function aimed to minimize the total route time. Hassani et al. [33] introduced the differential evaluation algorithm as a solution approach

for a nurse scheduling problem. The objective of their proposed algorithm is to minimize the overall cost, which includes various factors such as overtime, undertime, employment, and other related costs.

Furthermore, the HHC problem can also have multiple objectives. Rasmussen et al. [34] considered three objectives: minimizing the total cost, maximizing preference-based visits of caregiving sans patients, and maximizing the number of visited patients. Nickel et al. [35] dealt with the HHC problem by minimizing four objectives: the number of unscheduled tasks, the nurse-patient loyalty and overtime study, and the distance travelled by all nurses. It should be noted that only total overtime study and distance travelled is aimed to be minimized and the workload balanced is not taken into consideration. Mankowska et al. [36] considered minimizing three objectives: the total distance travelled, the total tardiness of services and the maximal tardiness observed overall service operation. The third objective function is similar to our objective function, however the tardiness is determined if a service lasts after the time windows of a patient. Thus, if all services can be scheduled between time windows of patients, then the value of the third objective function is equal to zero whether the workload is balanced or not. Braekers et al. [37] studied the HHC routing problem with two objectives: minimizing the cost and customer inconvenience. They applied an ϵ -constraint method to solve their mathematical model and also developed a metaheuristic based on a multi-directional local search.

Moreover, Hertz and Lahrichi [38] studied the problem for balancing workload of both service time and traveling time. Yalcindag et al. [39] proposed a data-driven method to estimate the travel times and dealt with the HHC problem with balancing the workload. Yalcindag et al. [40] presented a two-stage approach for addressing a home health care problem. This approach gradually combines the stages of assignment, planning, and scheduling decisions. The objective of the proposed model is to balance the workloads of operators, taking into account both travel time and service time. Decerle et al. [41] considered three objective functions: balancing the workload, and minimizing both the total time and patients' dissatisfaction. They formulated the problem and then developed a hybrid memetic-ant colony optimization algorithm. Kandakoglu et al. [42] tackled the scheduling and routing problem of home health nurses by employing a mixed-integer linear programming algorithm. The objective of their study was to balance the workload among the nurses effectively. Gomes et al. [43] focused on addressing a multi-objective problem in a system that involves patient-caregiver loyalty and dynamic patient numbers. Their study aimed to optimize multiple objectives, including workload balancing, minimizing total travel time, and minimizing the variation in visiting hours. Yang et al. [44] dealt with three objectives: minimizing the route cost, improving service consistency and balancing the workload. Additionally, the travel and service time were considered to be uncertain, so a multi-objective artificial bee colony framework was developed to solve the problem. The workload objective of these three studies [34-44] is similar to our work, but those covered the case of only one hospital.

We summarized the literature in Table 1 and it is shown from Table 1, the problem MH-HHCRP-BW defined in this paper cover all the criteria: balance workload for both service duration and traveling time, and multiple hospital. It is seen from Table 1 that there are studies that deal with balancing the workload taking into consideration only service duration or only travelling time or both of them. However, all these problems are for one hospital. A HHC system may have multiple hospitals, e.g., the main motivation of this paper, public hospitals in Türkiye. Addition, there are private healthcare companies having multiple affiliates in an area and providing services for HHC. Thus, balancing the workload for only one hospital of the healthcare systems having multiple hospitals is not sufficient to be fair among healthcare workers of the relevant healthcare system. According to best of our knowledge, our study is the first considering all these criteria, especially for multiple hospitals.

Table 1. Literature Review

Study	Balance the Workload		Multiple Hospital
	Service Duration	Travelling Time	
Bredstöm and Rönnqvist [10]	-	+	-
Lanzarone and Matta [11]	+	-	-
Carello et al. [12]	+	-	-
Yuan et al. [13]	-	-	-
Rest and Hirsch [14]	-	-	-
Shi et al. [15]	-	-	-
Masmoudi and Cheikhrouhou [16]	-	-	-
Liu et al. [17]	-	-	-
Bahadori-Chinibelagh et al. [18]	-	-	+
Tanoumand and Ünlüyurt [19]	-	-	-
Li et al. [20]	-	-	-
Allaoua [21]	-	-	-
Cappanera et al. [22]	+	+	-
Frifita et al. [23]	-	-	-
Rahimian et al. [25]	-	-	-
Riazi et al. [26]	-	-	-
Riazi et al. [27]	-	-	-
Moussavi et al. [28]	-	-	-
Grenouilleau et al. [29]	-	-	-
Grenouilleau et al. [30]	-	-	-
Shahejati-Bushehri et al. [31]	-	-	-
Dekhici et al. [32]	-	-	-
Hassani et al. [33]	-	-	-
Rasmussen et al. [34]	-	-	-
Nicket et al. [35]	-	-	-
Mankowska et al. [36]	-	-	-
Braekers et al. [37]	-	-	-
Hertz and Lahrichi [38]	+	+	-
Yalcindag et al. [39]	+	+	-
Yalcindag et al. [40]	+	+	-
Decerle et al. [41]	+	+	-
Kandakoglu et al. [42]	+	+	-
Gomes et al. [43]	+	+	-
Yang et al. [44]	+	+	-
This study	+	+	+

3. PROBLEM DESCRIPTION AND MATHEMATICAL MODEL

In the MH-HHCRP-BW, the assumptions are: there is more than one hospital; patients and caregivers have time windows; each patient is visited at one time by one caregiver; a route between the hospitals is not allowed; and each route starts from and ends at a hospital. A demonstration of the problem is given in Figure 1. The pentagons and circles represent hospitals and patients, respectively.

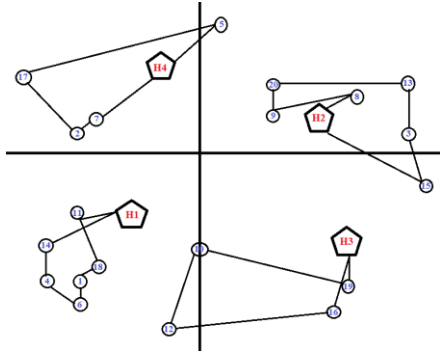


Figure 1. A demonstration of the MH-HHCRP-BW

First, we define sets, parameters and decision variables of the problem as follows.

Sets:

- K Set of routes.
- I_0 Set of hospitals.
- I_1 Set of patients.
- $I_0 \cup I_1 = I$ Set of hospitals and patients.

Parameters:

- T_{ij} Travel time between i th hospital/patient and j th hospital/patient, $\forall i, j \in I$.
- D_i Operation time for the i th patient, $\forall i \in I_1$.
- $[a_i, b_i]$ Time windows for the i th patient, $\forall i \in I_1$.
- $[e_i, l_i]$ Time windows (working hours) for the i th hospital, $\forall i \in I_0$.

Decision variables:

- $x_{ijk} \begin{cases} 1, & \text{if route } k \text{ goes from } i \text{ to } j \\ 0, & \text{otherwise} \end{cases}, \forall i, j \in I, i \neq j, k \in K.$
- t_i The start time of the visit to the i th patient, $\forall i \in I_1$.

The MILP is reformulated for the MH-HHCRP-BW as follows (see e.g. [45, 46]).

$$\min_{\forall k, l \in K, l \neq k} \max \left\{ \sum_{i \in I} \sum_{j \in I} (T_{ij} + D_i) x_{ijk} - \sum_{i \in I} \sum_{j \in I} (T_{ij} + D_i) x_{ijl} \right\}, \quad (1)$$

subject to

$$\sum_{k \in K} \sum_{i \in I_0} x_{ijk} = 0 \quad \forall j \in I_0 \quad (2)$$

$$\sum_{k \in K} \sum_{i \in I} x_{ijk} = 1 \quad \forall j \in I_1 \quad (3)$$

$$\sum_{k \in K} \sum_{i \in I} x_{ijk} = \sum_{k \in K} \sum_{i \in I} x_{jik} \quad \forall j \in I_1 \quad (4)$$

$$\sum_{j \in I_1} x_{ijk} = \sum_{j \in I_1} x_{jik} \quad \forall i \in I_0, k \in K \quad (5)$$

$$\sum_{k \in K} \sum_{j \in I_1} x_{ijk} = 1 \quad \forall i \in I_0 \quad (6)$$

$$t_{ik} + (T_{ij} + D_i) \leq t_{ik} + b_i(1 - x_{ijk}) \quad (7)$$

$$\forall i, j \in I, k \in K$$

$$a_i \sum_{j \in I} x_{ijk} \leq t_{ik} \leq b_i \sum_{j \in I} x_{ijk} \quad (8)$$

$$\forall i \in I_1, k \in K$$

$$e_i \leq t_{ik} \leq l_i \quad \forall i \in I_0, k \in K \quad (9)$$

$$x_{ijk} \in \{0,1\} \quad \forall i, j \in I, k \in K \quad (10)$$

$$t_{ik} \geq 0 \quad \forall i \in I_1. \quad (11)$$

The objective (1) is to balance the workload between the hospitals. The workload contains both the traveling time and the operation time for patients. Equation (2) prevents travel between hospitals. Equation (3) ensures that every patient is visited once. Equation (4) says that if a patient is visited by a hospital team, then the team should leave from the patient. Equation (5) declares that the team should leave from a hospital and then return to the hospital. Equation (6) ensures that only one home healthcare vehicle leaves from a hospital. Equation (7) takes into consideration both the traveling time and operation time of the patient to determine the visiting time of the patient. Equation (8) restricts the visiting time of a patient to the suitable time window of the patient. Equation (9) ensures that the vehicles for home healthcare leave and return to hospitals within a work shift. Equations (10) and (11) specify the variable domains. The objective function (1) for multiple hospitals and the equation (7) with operations time of patients and the equation (8) for the work shift of home healthcare team are newly defined in this paper.

The objective function (1) is linearized as following:

$$\min w \quad (12)$$

$$\sum_{i \in I} \sum_{j \in I} (T_{ij} + D_i) x_{ijk} - \sum_{i \in I} \sum_{j \in I} (T_{ij} + D_i) x_{ijl} \leq w \quad \forall k, l \in K, i \neq j \quad (13)$$

Thus, the final model is to minimize (12) under the constraints (2)-(11) and (13).

Now, the importance of the new objective function is explained using a toy problem. Consider the following objective function that is minimizing the travel time of the routes.

$$\min \sum_k \sum_{i \in I} \sum_{j \in I} T_{ij} x_{ijk} \quad (14)$$

When the objective function about time minimization (14) is taken into consideration, an unbalanced workload may occur. A toy problem is generated (the data for the toy problem is given in the appendix). Then, the toy problem is solved with the model, having objective function (12) under the constraints (2)-(11), (13), and objective function (14) under the constraints (2)-(11), individually. The results are given in detail in Table 2. The time difference between the hospitals having maximum and minimum total time (the sum of total travel time and total operation time) is 15.7 minutes for objective function (12) while it is 135.1 minutes for objective function (14). It is clearly seen that the new objective function ensures a balanced workload among hospitals in terms of both travel time from hospital to homes and operation time at homes.

Table 2. Results of the toy problem for the objective functions (12) and (14)

Objective function	Hospital number	Total travel time	Total operation time	Total time
(12)-balancing the workload	Hospital 1	210	19.4	229.4
	Hospital 2	195	34.8	229.8
	Hospital 3	208	25.1	233.1
	Hospital 4	200	17.4	217.4
			Maximum time difference	
(14)-minimizing total time	Hospital 1	264	29.2	293.2
	Hospital 2	259	22.9	281.9
	Hospital 3	138	20.1	158.1
	Hospital 4	152	24.5	176.5
			Maximum time difference	

4. TABU SEARCH WITH VARIABLE NEIGHBOURHOOD SEARCH (TS-VNS)

At this stage of the study, we use a heuristic and a metaheuristic algorithm together to solve the problem. Heuristic algorithms are preferred because they can find acceptable results in a relatively short time for problems that are difficult to solve or take a long time with classical optimization approaches. While heuristic algorithms offer problem-specific approaches, metaheuristic algorithms provide a guiding framework for solving the problem. In this study we use a heuristic algorithm to generate a feasible initial solution. After that we apply the hybrid algorithm, we propose which combines two different metaheuristic algorithms: Tabu search and Variable neighborhood search.

4.1. The Initial Solution

We developed a heuristic algorithm to find a feasible initial solution. The algorithm that is given in Figure 2 assigns each route in random order by taking into account the time window constraints. The algorithm continues this assignment process until all patients are assigned. We used permutation representation to represent routes since the permutation representation provides an efficient representation of the solution space of our problem and is

advantageous for search procedures. And so, the solution is generated by the heuristic algorithm has matrix form in which each row is a route.

One of the parameters used in the pseudocode in Figure 2, **lb**, is a vector which represents the lower bounds of the time window constraints. The variable **nodes** represent the patient group to be assigned to routes, and **routes** represents the number of routes to be created.

```

Function InitialSolution (nodes, routes, lb)
for each  $x$ : /*  $x$  is the notation of routes */
     $x[0] = \text{argmin}(\mathbf{lb})$ 
    delete  $\mathbf{lb}[\text{argmin}(\mathbf{lb})]$ 
    delete nodes [ $x[0]$ ]
while nodes  $\neq \emptyset$ 
    routes = Permutation (routes)
    for each  $r$ :
        Generate a list of nodes eligible to be assigned to
        the current route.
        if list  $\neq \emptyset$ 
            Choose the node which has a minimum time
            window lower bound in the list.
        else
            break
        end if
    end for
end while

```

Figure 2. Pseudo code for feasible initial solution

4.2. TS-VNS

The tabu search algorithm is a metaheuristic with memory [47]. The memory is a tabu list that stores previous solutions or moves to prevent cycles in the search space. The algorithm starts with a feasible solution and proceeds by searching for neighboring solutions of this current solution with move operators. The algorithm calculates a fitness value for each feasible neighboring solution it reaches. If this fitness value is better than the fitness value of the best available solution, the current best solution is updated. If there is no better solution than the current best solution among the neighboring solutions, the current solution is updated with a feasible neighbor solution with a worse fitness value. The searching continues from this current solution. Besides, the variable neighborhood search aims to reach the global optimum by changing the neighborhood of a local search stuck in the local optimum, and it also uses neighborhood change in descent to the local optimum [48].

In the TS-VNS approach, we use some procedures of Tabu search and Variable neighborhood search algorithms together. The part that we inherit from the VNS algorithm in the proposed approach is the use of three different move operators which are inter-route swap, in-route swap, and insertion. We use inter-route swap and in-route swap operators nested. Inter-route swap disrupts the initial solution and diversion the local search to another area. Then, in-route swap searches feasible and improvement solutions and thus searching intensifies in this search area. This nested

loop iterates over a deterministic number of iterations. And finally, we aim to shake the current solution in a controlled way with the insertion operator. The algorithm uses tabu lists for swap operators and this is the part of taken from the TS

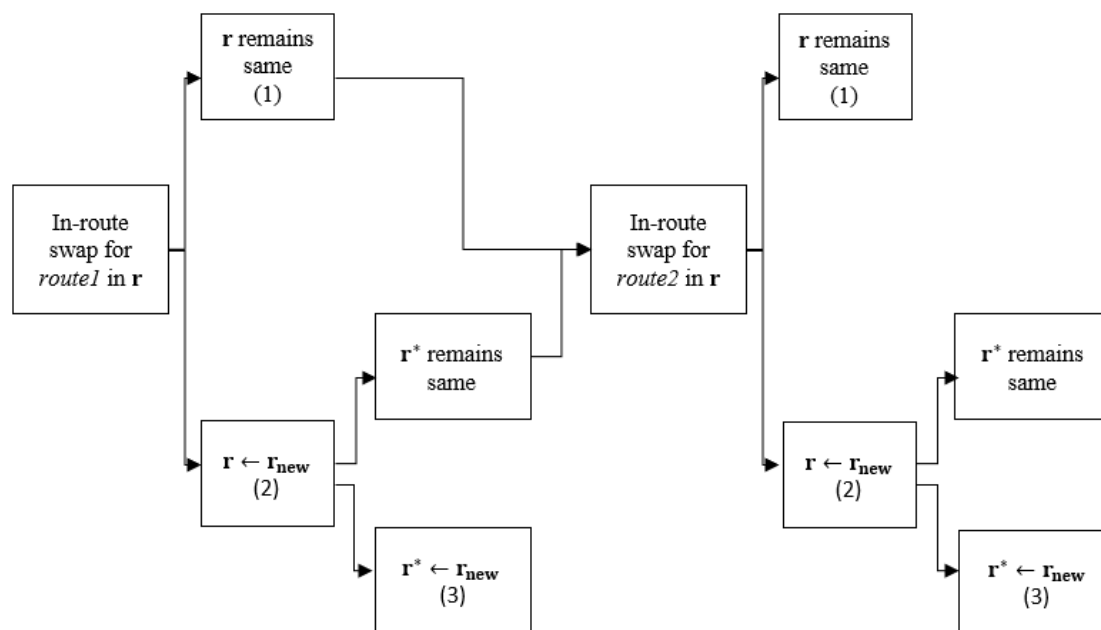
algorithm. We use the short-term memory and aim to prevent circles in the search space with using tabu lists as with the TS algorithm. Pseudocode of the TS-VNS algorithm is given in Figure 3.

```

Function TS-VNS ( $\mathbf{r}$ ,  $end1$ ,  $end2$ )
 $\mathbf{r}^* = \mathbf{r}$  /* the best current solution */
 $Z^* = \text{ObjValue}(\mathbf{r}^*)$ 
 $iter = 0$ 
Create a empty tabu list for inter-route swap operation which called tabucswap
while  $iter < end1$ : /*inter-route swap phase*/
    Choose 2 random routes and 2 random positions [ $route1, route2, pos1, pos2$ ] for inter-route swap operation and
    check if [ $route1, route2, pos1, pos2$ ] in tabucswap
    If this move is a tabu then repeat previous step
    Apply inter-route swap operation on current solution and update current solution as
     $\mathbf{r}[route1][pos1], \mathbf{r}[route2][pos2] \leftarrow \mathbf{r}[route2][pos2], \mathbf{r}[route1][pos1]$ 
    if  $\mathbf{r}$  is feasible
         $Z = \text{ObjValue}(\mathbf{r})$ 
        if  $Z \leq Z^*$ 
             $\mathbf{r}^* \leftarrow \mathbf{r}$ 
             $Z^* \leftarrow \text{ObjValue}(\mathbf{r})$ 
        end if
    else /*in-route swap phase*/
        Apply swap operation for  $route1$  and update current solution as  $\mathbf{r} \leftarrow \text{Swap}(\mathbf{r}, \mathbf{r}[route1], i)$ 
        if swap operator found a better solution
             $\mathbf{r}^* \leftarrow \mathbf{r}$ 
             $Z^* \leftarrow \text{ObjValue}(\mathbf{r})$ 
        else
            Apply swap operation for  $route2$  and update current solution as  $\mathbf{r} \leftarrow \text{Swap}(\mathbf{r}, \mathbf{r}[route2], i)$ 
            if swap operator found a better solution
                 $\mathbf{r}^* \leftarrow \mathbf{r}$ 
                 $Z^* \leftarrow \text{ObjValue}(\mathbf{r})$ 
            end if
        end if
    end if
     $iter := iter + 1$ 
    Append [ $route1, route2, pos1, pos2$ ] to tabucswap
end while
 $iter = 0$ 
while  $iter < end2$  /*insertion phase*/
     $\mathbf{r}_{new} = \text{Insertion}(\mathbf{r})$ 
    if  $\mathbf{r}_{new}$  is feasible:
         $\mathbf{r} \leftarrow \mathbf{r}_{new}$ 
         $Z = \text{ObjValue}(\mathbf{r})$ 
        if  $Z \leq Z^*$ 
             $\mathbf{r}^* \leftarrow \mathbf{r}$ 
             $Z^* \leftarrow \text{ObjValue}(\mathbf{r})$ 
        end if
    end if
     $iter := iter + 1$ 
end while

```

Figure 3. Pseudo code for TS-VNS



The numbers represent the following conditions, respectively:

- 1) There is no different feasible neighboring solution,
- 2) A different feasible neighboring solution (r_{new}) is exist,
- 3) $Z_{new} \leq Z^*$.

Figure 4. In-route swap phase in TS-VNS algorithm.

Function Swap (r , $route$, i)

$r_{new} = r$

$iter = 0$

Create a empty list called *tabuswap*

Calculate t which is total number of swap move for **route** which has n elements: $t = n(n - 1)/2$

While $iter < t$

 Choose 2 random position [$pos1$, $pos2$] and check if this moves in *tabuswap*

 If this move is a tabu then repeat previous step

 Swap elements on these position in **route**: **route**

[$pos1$], **route** [$pos2$] = **route** [$pos2$], **route** [$pos1$]

 Append [$pos1$, $pos2$] to *tabuswap*

 Update i . route in current solution r_{new} with new

route

if r_{new} is feasible

$r \leftarrow r_{new}$

break

end if

$iter := iter + 1$

end while

Figure 5. Pseudo code for in-route swap operator

The first inputs of the TS-VNS function, r , is a feasible initial solution. We find this feasible initial solution by random assignments and for this, we run the function in Figure 2 repeatedly until it finds a feasible solution. The other four inputs are hyperparameters, and decision maker determines their values. The first hyperparameter, *iteration*, is the number of iterations of outside loop which is the repeat number of process, and similarly *inter-route swap* represents the number of the inter-route swaps and *insertion* represents

the number of insertions in the Figure 3. The last hyperparameter, *tabu*, is the size of the tabu list for the inter-route swap operator. If the tabu list is complete when the last move is added to the tabu list, the oldest move is removed from the list. The in-route swap operator does not remove any tabu from the list as it searches for all neighboring solutions on a route.

The TS-VNS procedure starts with the inter-route swap, which applies a swap operation between two randomly determined routes. The algorithm finds a new solution with the inter-route swap operator and moves on with this current solution (r). If r is a feasible solution that provides improvement, the current best solution is updated. If r is not feasible, the algorithm switches to another move operator. In this point, searching moves on with in-route swap operator. This operator aims to search the local area to reach any feasible solution.

After the inter-route swap, there is two routes to apply in-route swap process. We follow a sequential procedure to not miss the benefit of the swap made on one of the routes. Firstly, the algorithm applies the in-route swap for *route1* and it continues according to the result obtained. The in-route swap operator can produce two different kinds of solutions: i) there is a different feasible neighboring solution and current solution is updated with this new solution, ii) current solution is not updated. If the in-route swap operator finds a new feasible solution, the objective function value of this new current solution is calculated. If the new current solution improves the current best solution, the current best solution is also updated. After the in-route swap operation if the current solution is the same, or if the new solution is not better than the current best solution then the in-route swap operator is applied on *route2*. The algorithm returns to the inter-route swap phase after the second in-route swap

operation and repeats the same steps. We explain this phase of the algorithm with a flowchart in Figure 4. As seen in the figure, if the algorithm reaches a new feasible solution by neighborhood search, the current solution is updated regardless of whether this solution is better than the current best solution. In this way, it is aimed to search for good solutions that can be reached over relatively bad solutions. After this first loop, which nested the diversification and intensification procedures, the algorithm switches the search operator for the last time.

The in-route swap operator gets current solution \mathbf{r} which is a 2-dimensional vector, \mathbf{route} which is a one-dimensional vector and i which is index of \mathbf{route} in \mathbf{r} as inputs. Then, it chooses 2 random position and swaps elements on these position in \mathbf{route} . Finally, it updates \mathbf{route} in \mathbf{r} with new \mathbf{route} . With the swap operator, it is possible to search for all

Possible neighboring solutions. If neighboring searching find a feasible solution, then search stops, and the current route is updated with the new route. In the situation that is all neighboring solutions are searched, and no feasible solution is found, the current solution remains the same. Pseudo code for in-route swap operator is given in Figure 5.

Function Insertion (\mathbf{r})

Calculate workload of every route on current solution \mathbf{r} and hold with an array; \mathbf{w}

Detect routes with maximum and minimum workloads: $\mathbf{max_w}$ and $\mathbf{min_w}$

Choose a random element in $\mathbf{max_w}$ and remove it from $\mathbf{max_w}$, then insert this element in a random position in $\mathbf{min_w}$

Update \mathbf{r} with these new two routes

Figure 6. Pseudo code for insertion operator

Third and last move operator is the insertion operator (see Figure 6) that is modified in a way that contributes to the workload balance, which is the objective of the problem. The proposed insertion operator is named as “balanced-insertion” operator. The balanced-insertion operator transfers a patient from the route with the highest workload to the route with the least workload. In this way, we aim to find better neighboring solutions with small steps.

5. COMPUTATIONAL RESULTS

For numerical results, we used two types of test problems, which were classified as small-scale and large-scale. Table 3 summarizes the information about the test problems. Small-scale problems were generated, while large-scale problems were modified from the literature [49]. Small-scale problems were solved by both mathematical model and TS-VNS algorithm. Obtaining solutions of large-scale problems by the mathematical model within reasonable time is difficult due to the size of problems. Thus, these problems were solved by only the TS-VNS algorithm. In this section, we showed the performance of the TS-VNS algorithm against the solution of the mathematical model obtained by General Algebraic Modeling system (GAMS) with CPLEX solver for small-scale problems. In addition, we analyze the efficiency

of the TS-VNS algorithm in large-scale problems in terms of time and the improvement it provides in random feasible initial solutions. We solved all runs by a computer with a core i7 processor (2.3 GHz) and 16 GB RAM.

Table 3. Properties of test problems

	Problem no	Number of patients	Number of caregivers team for each hospital	Number of hospitals
Small-scale	pr-s-1	30	1	6
	pr-s-2	35	1	7
	pr-s-3	40	1	8
	pr-s-4	45	1	9
Large-scale	pr-l-1	48	2	4
	pr-l-2	96	3	4
	pr-l-3	144	5	4
	pr-l-4	192	7	4
	pr-l-5	240	8	4
	pr-l-6	288	9	4
	pr-l-7	72	2	6
	pr-l-8	144	3	6
	pr-l-9	216	5	6
	pr-l-10	288	6	6
	pr-l-11	48	2	4
	pr-l-12	96	4	4
	pr-l-13	144	6	4
	pr-l-14	192	7	4
	pr-l-15	240	8	4
	pr-l-16	288	11	4
pr-l-17	72	2	6	
pr-l-18	144	4	6	
pr-l-19	216	5	6	
pr-l-20	288	7	6	

In all of the solutions of the test problems with the TS-VNS algorithm, we performed three repetitions by considering the random initial solution of the algorithm and the random search mechanism in the structure of the algorithm. The value of time in the tables are the total duration of three repetitions, and the “best value” and “average value” are the best and average values found during this period. Since the purpose of the problem is the balanced distribution of workloads in terms of time, the results of the TS-VNS algorithm also express time.

5.1. Hyperparameter Optimization

A hyperparameter tuning is required to determine the hyperparameter values (the number of iterations, inter-route, insertion, and tabu length) of the TS-VNS algorithm. For this purpose, we used a Python module called “hyperopt” [50]. Hyperopt is based on Bayesian optimization and supports automated hyperparameter optimization. Four basic components need to be defined to use hyperopt: an objective function to be minimized, a search space including hyperparameter values, a database to be used to store the points evaluated during the search, and a search algorithm to enable the transition between points in the search space.

Table 4. Search spaces of hyperparameters for hyperopt

Hyperparameter	Parameter of uniform distribution	
	Small-scale	Large-scale
Iteration	(100, 500)	(100, 1000)
Inter-route swap	(100, 500)	(100, 1000)
Insertion	(50, 500)	(100, 1000)
Tabu length	(10, 50)	(10, 100)

The objective function for this problem is to balance the total workload, that is, to minimize the difference between the workloads. Search space can be made up of discrete values as well as probability distributions. We used uniform distribution. Finally, tree of Parzen estimators (TPE) [51] and random search can be used for the search algorithm. We used TPE which is a sequential model-based optimization approach.

We performed hyperparameter optimizations for small-scale and large-scale problems separately with using the problems pr-s-4 and pr-l-10 for small-scale problems and large-scale problems, respectively. We chose these problems because they are relatively large. Table 4 summarizes the distribution parameters we used for the search space. We made different trials before deciding on the distribution parameters. As a result of the guidance of these trials, we used search spaces with as wide ranges as possible.

At the end of the search, which took about 3 hours (170.6 min) for small-scale and about 21 hours (1249.4 min) for large-scale, the hyperparameter values we determined with hyperopt were [iteration:466, inter-route swap: 190, insertion: 429, tabu: 41] for small-scale problems and [iteration: 726, inter-route swap: 887, insertion: 318, tabu: 49] for large-scale problems.

5.2. Small-Scale Test Problems

We solved the small-scale test problems by both the mathematical model and TS-VNS algorithm. We used the General Algebraic Modeling System (GAMS) with CPLEX solver to solve the mathematical model within a 12-hour time limit. We solved the same problems with TS-VNS using the hyperparameter values mentioned above. The best and average values and also, a comparison with the GAMS-CPLEX results are given in Table 5. The gap is calculated as

$$\left(\frac{z^{TS-VNS} - z^{GAMS-CPLEX}}{z^{GAMS-CPLEX}} \right) 100 \quad (15)$$

where z^{TS-VNS} and z^{GAMS} show the results obtained by TS-VNS and GAMS-CPLEX, respectively.

It is clear that the TS-VNS approach we proposed achieves successful results, especially as the problem size increases, it finds competitive results compared to GAMS-CPLEX. According to the best results TS-VNS found a better solution than GAMS-CPLEX for all problems of small-scale problems while for the average solution, it found better results for 3 out of the 4 small-scale problems. It is seen that the TS-VNS algorithm is more effective in terms of time. Addition, it should be noted, the best possible lower bound

given by GAMS-CPLEX is 0 (zero) for all small-scale problems which means that the workload of all caregivers are equal. However, it is impossible to obtain the results that has the value of 0 (zero) since the problem is mixed-integer programming and the dual gap is highly possible.

Table 5. Comparison of the results obtained by the mathematical model-GAMS and TS-VNS

	pr-s-1	pr-s-2	pr-s-3	pr-s-4
GAMS-CPLEX (within 12 hours)	0.6	1.4	2.6	10.8
TS-VNS (Average)	0.9	0.7	1.3	2.3
Gap (%)	61.2	-49.2	-51.3	-78.8
TS-VNS (Best)	0.3	0.4	0.9	0.9
Gap (%)	-42.3	-69.6	-64.6	-91.3
TS-VNS time (min)	4.2	4.4	4.9	5.1

5.3. Large-Scale Test Problems

We solved the 20 large-scale problems in Table 3 with the TS-VNS algorithm. We summarize the results in Table 6. In addition, we have included the averages of the objective function values of the feasible initial solutions produced in each iteration. The difference between the values of the initial solutions and the final values shows the success of the TS-VNS algorithm in improving the solution.

Table 6. Results of large-scale test problems

Problem no	Initial solution	TS-VNS (Best)	TS-VNS (Average)	Time (min)
pr-l-1	201.1	4.1	5.1	40.3
pr-l-2	222.8	7.9	12.9	123.5
pr-l-3	220.3	8.0	11.9	128.7
pr-l-4	419.3	7.4	9.3	144.5
pr-l-5	257.5	4.4	6.7	199.8
pr-l-6	266.0	6.7	7.9	280.6
pr-l-7	159.2	6.2	7.4	52.7
pr-l-8	266.0	7.7	9.8	161.4
pr-l-9	277.8	5.9	7.0	168.2
pr-l-10	288.6	8.5	10.0	285.5
pr-l-11	232.7	1.1	1.2	48.3
pr-l-12	344.9	1.9	2.1	78.9
pr-l-13	333.3	3.9	4.8	101.9
pr-l-14	301.4	2.8	3.8	161.5
pr-l-15	264.3	2.7	2.9	229.6
pr-l-16	359.2	4.2	5.2	215.1
pr-l-17	212.1	1.6	2.2	61.1
pr-l-18	337.4	3.7	3.9	101.9
pr-l-19	298.3	3.7	4.5	194.4
pr-l-20	303.8	5.2	5.5	228.9

Finally, we made a comparison with large-sized problems to show that the balanced-insertion operator can achieve better results than the insertion moves with random selection. The numerical results are given in appendix Table A3. The graphs in Figure 7 visualize these numerical results. It can be seen from the graphs that the balanced-insertion approach yields better results. We use a t-test to test the significance

of the difference between the two approaches. “ H_0 : The average of results for the two different insertion type is equal”, and “ H_1 : The average of results for two different insertion type is not equal” are the hypothesis of t-test for both best and average results. The value of the test statistic is $2e-06$ for best values and $8e-06$ for average values. Both values are lower than the critic value at the 0.05 significance level, so the H_0 hypothesis is rejected. As a result, there is a statistically significant difference between the results of balanced-insertion and random insertion.

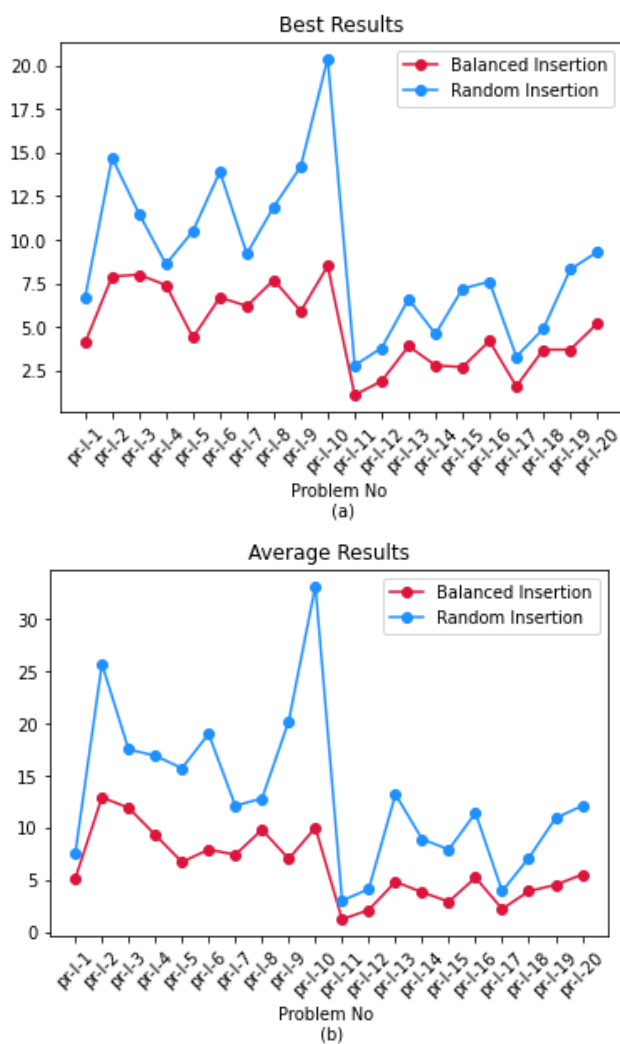


Figure 7. The comparison of balanced-insertion and random insertion: a) Best results of problems, b) Average results of problems

6. CONCLUSION

In this paper, we study the home healthcare routing problem for multiple hospitals with the objective of balancing workloads between hospitals. For this purpose, we propose a mathematical model based on a multi-depot vehicle routing problem with time windows. We show that when the balancing of the workload is not taken into consideration the differences between the workloads of the hospitals may be very large.

We generated small-scale problems and modified test problems from the literature for large-scale problems. Since

the mathematical model can only solve the problem for up to 45 patients and 9 hospitals (small-scale problems), we use general-purpose metaheuristic algorithms to find good solutions in large search spaces. In the study, we proposed a hybrid algorithm (TS-VNS) that combines the strengths of two different metaheuristic algorithms. For small-scale test problems, we compared the results obtained by the TS-VNS algorithm and the mathematical model solved by GAMS. Solutions show that the TS-VNS algorithm is capable of finding good solutions. Furthermore, large-scale problems are solved by the TS-VNS algorithm within a reasonable time.

TS-VNS includes arrangements suitable for the representation of the problem, its purpose and time window constraints. In accordance with the objective function of the problem, the proposed "balanced-insertion" operator works better than the "random insertion" operator. In the study, results supporting that the algorithm has a successful search procedure in producing time-efficient and good solutions were obtained. In addition, TS-VNS has the flexibility of a general-purpose algorithm that can be easily applied to problems of similar nature.

The study deals with the home health care problem with the aim of balanced workload in terms of both service time and travel time. Unlike studies in the literature that have both purposes, it seeks a solution for multiple hospitals. It is also innovative in terms of the proposed hybrid metaheuristic approach.

For future research. Patients' hospital preference could be considered. Since every preference may not be acted upon, a second objective function is modeled as maximizing the patients' preference. Thus, a multi-objective mathematical model would be constructed and multi-objective solution methods, such as the weighted-sum method, would be needed to generate solutions.

Author contributions: Gülçin Dinç Yalçın; Literature Review, Mathematical Model, Developing Algorithm, Writing Manuscript- Tuğçe Yavuz; Developing Algorithm, Computational Results, Writing Manuscript-Şüheda Altıntaş; Literature Review, Mathematical Model.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study has received no financial support.

REFERENCES

- [1] P. Egeborn, P. Flisberg, and M. Rönnqvist, "Laps Care—an operational system for staff planning of home care", *European Journal of Operational Research*, vol. 171, no. 3, pp. 962–976, Jun. 2006, doi: 10.1016/j.ejor.2005.01.011.
- [2] "WHO's work on the UN decade of healthy ageing (2021-2030)". Accessed on: Feb. 13, 2023. [Online]. Available: <https://www.who.int/initiatives/decade-of-healthy-ageing>

- [3] “Home healthcare market size, share & trends analysis report, 2022-2030”. Accessed on: Feb. 13, 2023. [Online]. Available: <https://www.researchandmarkets.com/reports/5450245/home-healthcare-market-size-share-and-trends>
- [4] E., Cheng, and, J. L. Rich, “A home healthcare routing and scheduling problem”, 1998. Accessed on: Feb. 13, 2023. [Online]. Available: <https://scholarship.rice.edu/handle/1911/101899>
- [5] S. Afifi, D.-C. Dang, and A. Moukrim, “Heuristic solutions for the vehicle routing problem with time windows and synchronized visits”, *Optim Lett*, vol. 10, no. 3, pp. 511–525, Mar. 2016, doi: 10.1007/s11590-015-0878-3.
- [6] M. Cissé, S. Yalçındağ, Y. Kergosien, E. Şahin, C. Lenté, and A. Matta, “OR problems related to home health care: A review of relevant routing and scheduling problems”, *Operations Research for Health Care*, vol. 13–14, pp. 1–22, Jun. 2017, doi: 10.1016/j.orhc.2017.06.001.
- [7] M. Di Mascolo, C. Martinez, and M.-L. Espinouse, “Routing and scheduling in home health care: A literature survey and bibliometric analysis”, *Computers & Industrial Engineering*, vol. 158, p. 107255, Aug. 2021, doi: 10.1016/j.cie.2021.107255.
- [8] C. Fikar and P. Hirsch, “Home health care routing and scheduling: A review”, *Computers & Operations Research*, vol. 77, pp. 86–95, Jan. 2017, doi: 10.1016/j.cor.2016.07.019.
- [9] L. Grieco, M. Utley, and S. Crowe, “Operational research applied to decisions in home health care: A systematic literature review”, *Journal of the Operational Research Society*, vol. 72, no. 9: pp. 1960-1991, 2021. <https://doi.org/10.1080/01605682.2020.1750311>
- [10] D. Bredström and M. Rönnqvist, “Combined vehicle routing and scheduling with temporal precedence and synchronization constraints”, *European Journal of Operational Research*, vol. 191, no. 1, pp. 19–31, Nov. 2008, doi: 10.1016/j.ejor.2007.07.033.
- [11] E. Lanzarone and A. Matta, “Robust nurse-to-patient assignment in home care services to minimize overtimes under continuity of care”, *Operations Research for Health Care*, vol. 3, no. 2, pp. 48-58, June 2014, doi: <https://doi.org/10.1016/j.orhc.2014.01.003> .
- [12] G. Carello, E. Lanzarone, and S. Mattia, “Trade-off between stakeholders’ goals in the home care nurse-to-patient assignment problem”, *Operations Research for Health Care* vol. 16, pp. 29–40, March 2018, doi: <https://doi.org/10.1016/j.orhc.2017.12.002>
- [13] B. Yuan, R. Liu, and Z. Jiang, “A branch-and-price algorithm for the home health care scheduling and routing problem with stochastic service times and skill requirements”, *International Journal of Production Research*, vol. 53, no. 24, pp. 7450–7464, Dec. 2015, doi: 10.1080/00207543.2015.1082041.
- [14] K.-D. Rest and P. Hirsch, “Daily scheduling of home health care services using time-dependent public transport”, *Flex Serv Manuf J*, vol. 28, no. 3, pp. 495–525, Sep. 2016, doi: 10.1007/s10696-015-9227-1.
- [15] Y. Shi, T. Boudouh, and O. Grunder, “A hybrid genetic algorithm for a home health care routing problem with time window and fuzzy demand”, *Expert Systems with Applications*, vol. 72, pp. 160–176, Apr. 2017, doi: 10.1016/j.eswa.2016.12.013.
- [16] M. Masmoudi and N. Cheikhrouhou, “Heterogeneous vehicle routing problems with synchronization: Application to homecare scheduling routing problem”. 9th Conference on Engineering and Management of Healthcare Systems GISEH, Geneva, 2018.
- [17] R. Liu, B. Yuan, and Z. Jiang, “A branch-and-price algorithm for the home-caregiver scheduling and routing problem with stochastic travel and service times”, *Flex Serv Manuf J*, vol. 31, no. 4, pp. 989–1011, Dec. 2019, doi: 10.1007/s10696-018-9328-8.
- [18] S. Bahadori-Chinibelagh, A. M. Fathollahi-Fard, and M. Hajiaghahi-Keshteli, “Two constructive algorithms to address a multi-depot home healthcare routing problem”, *IETE Journal of Research*, vol. 68, no. 2, pp. 1108–1114, Mar. 2022, doi: 10.1080/03772063.2019.1642802.
- [19] N. Tanoumand and T. Ünlüyurt, “An exact algorithm for the resource constrained home health care vehicle routing problem”, *Ann Oper Res*, vol. 304, no. 1, pp. 397–425, Sep. 2021, doi: 10.1007/s10479-021-04061-9.
- [20] Y. Li, T. Xiang, and W. Y. Szeto, “Home health care routing and scheduling problem with the consideration of outpatient services”, *Transportation Research Part E: Logistics and Transportation Review*, vol. 152, p. 102420, Aug. 2021, doi: 10.1016/j.tre.2021.102420.
- [21] H. Allaoua, S. Borne, L. Létochart, and R. Wolfler Calvo, “A matheuristic approach for solving a home health care problem”, *Electronic Notes in Discrete Mathematics*, vol. 41, pp. 471–478, Jun. 2013, doi: 10.1016/j.endm.2013.05.127.
- [22] P. Cappanera, M. G. Scutellà, F. Nervi, and L. Galli, “Demand uncertainty in robust Home Care optimization”. *Omega*, vol. 80, pp. 95–110, October 2018, doi: <https://doi.org/10.1016/j.omega.2017.08.012>
- [23] S. Frifita, M. Masmoudi, and J. Euchel, “General variable neighborhood search for home healthcare routing and scheduling problem with time windows and synchronized visits”, *Electronic Notes in Discrete Mathematics*, vol. 58, pp. 63–70, Apr. 2017, doi: 10.1016/j.endm.2017.03.009.
- [24] S. Lan, W. Fan, S. Yang, P. M. Pardalos and, N. Mladenovic, “A survey on the applications of variable neighborhood search algorithm in healthcare management”, *Annals of Mathematics and Artificial Intelligence*, vol. 89, no. 8, pp. 741-775, 2021, doi: 10.1007/s10472-021-09727-5.

- [25] E. Rahimian, K. Akartunalı, J. Levine, “A hybrid integer programming and variable neighbourhood search algorithm to solve nurse rostering problems”, *Eur J Oper Res*, vol. 258, no. 2, pp. 411–423, April 2017, doi: <https://doi.org/10.1016/j.ejor.2016.09.030>
- [26] S. Riazi, O. Wigström, K. Bengtsson, and B. Lennartson, “Decomposition and distributed algorithms for home healthcare routing and scheduling problem”, in *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Sep. 2017, pp. 1–7. doi: [10.1109/ETFA.2017.8247622](https://doi.org/10.1109/ETFA.2017.8247622).
- [27] S. Riazi, O. Wigström, K. Bengtsson, and B. Lennartson, “A column generation-based gossip algorithm for home healthcare routing and scheduling problems”, *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 1, pp. 127–137, Jan. 2019, doi: [10.1109/TASE.2018.2874392](https://doi.org/10.1109/TASE.2018.2874392).
- [28] S. E. Moussavi, M. Mahdjoub, and O. Grunder, “A matheuristic approach to the integration of worker assignment and vehicle routing problems: Application to home healthcare scheduling”, *Expert Systems with Applications*, vol. 125, pp. 317–332, Jul. 2019, doi: [10.1016/j.eswa.2019.02.009](https://doi.org/10.1016/j.eswa.2019.02.009).
- [29] F. Grenouilleau, A. Legrain, N. Lahrichi, and L.-M. Rousseau, “A set partitioning heuristic for the home health care routing and scheduling problem”, *European Journal of Operational Research*, vol. 275, no. 1, pp. 295–303, May 2019, doi: [10.1016/j.ejor.2018.11.025](https://doi.org/10.1016/j.ejor.2018.11.025).
- [30] F. Grenouilleau, N. Lahrichi, and L.-M. Rousseau, “New decomposition methods for home care scheduling with predefined visits”, *Computers & Operations Research*, vol. 115, p. 104855, Mar. 2020, doi: [10.1016/j.cor.2019.104855](https://doi.org/10.1016/j.cor.2019.104855).
- [31] S. Shahnejat-Bushehri, R. Tavakkoli-Moghaddam, M. Boronooos, and A. Ghasemkhani, “A robust home health care routing-scheduling problem with temporal dependencies under uncertainty”, *Expert Syst Appl*, vol. 182, no. 115209, November 2021, doi: <https://doi.org/10.1016/j.eswa.2021.115209>
- [32] L. Dekhici, R. Redjem, K. Belkadi, and A. Mhamedi, “Discretization of the firefly algorithm for home care”, *Can J Electr Comput Eng*, vol. 42, no. 1, pp. 20–26, 2016.
- [33] M. R. Hassani, and J. Behnamian, “A scenario-based robust optimization with a pessimistic approach for nurse rostering problem”, *J Comb Optim*, vol. 41, no. 1, pp. 143–169, November 2021, doi: <https://doi.org/10.1007/s10878-020-00667-0>
- [34] M. S. Rasmussen, T. Justesen, A. Dohn, and J. Larsen, “The home care crew scheduling problem: Preference-based visit clustering and temporal dependencies”, *European Journal of Operational Research*, vol. 219, no. 3, pp. 598–610, Jun. 2012, doi: [10.1016/j.ejor.2011.10.048](https://doi.org/10.1016/j.ejor.2011.10.048).
- [35] S. Nickel, M. Schröder, and J. Steeg, “Mid-term and short-term planning support for home health care services”, *European Journal of Operational Research*, vol. 219, pp. 574–587, Jun. 2012, doi: <https://doi.org/10.1016/j.ejor.2011.10.042> .
- [36] D. S. Mankowska, F. Meisel, and C. Bierwirth, “The home health care routing and scheduling problem with interdependent services”, *Health Care Management Science*, vol. 17, pp. 15–30, 2014, doi: <https://doi.org/10.1007/s10729-013-9243-1> .
- [37] K. Braekers, R. F. Hartl, S. N. Parragh, and, F. Tricoire, “A bi-objective home care scheduling problem: Analyzing the trade-off between costs and client inconvenience”, *European Journal of Operational Research*, vol. 248, no. 2, pp. 428–443, 2016, doi: [10.1016/j.ejor.2015.07.028](https://doi.org/10.1016/j.ejor.2015.07.028)
- [38] A. Hertz and N. Lahrichi, “A patient assignment algorithm for home care services”, *Journal of the Operational Research Society*, vol. 60, no. 4, pp. 481–495, 2009, doi: <https://doi.org/10.1057/palgrave.jors.2602574> .
- [39] S. Yalçındağ, A. Matta, E. Şahin, and J. G. Shanthikumar, “The patient assignment problem in home health care: using a data-driven method to estimate the travel times of care givers”, *Flexible Services and Manufacturing Journal*, vol. 28, pp. 304–335, June 2016, doi: <https://doi.org/10.1007/s10696-015-9222-6> .
- [40] S. Yalçındağ, P. Cappanera, M. Scutella, E. Şahin, and A. Matta, “Pattern-based decompositions for human resource planning in home health care services”, *Comput Oper Res*, vol. 73, pp. 12–26, September 2016, doi: <https://doi.org/10.1016/j.cor.2016.02.011>
- [41] J. Decerle, O. Grunder, A. Hajjam El Hassani, and O. Barakat, “A hybrid memetic-ant colony optimization algorithm for the home health care problem with time window, synchronization and working time balancing”, *Swarm and Evolutionary Computation*, vol. 46, pp. 171–183, May 2019, doi: [10.1016/j.swevo.2019.02.009](https://doi.org/10.1016/j.swevo.2019.02.009).
- [42] M. I. Gomes, and T.R.P. Ramos, “Modelling and (re-)planning periodic home social care services with loyalty and non-loyalty features”, *Eur J Oper Res*, vol. 277, no. 1 pp. 284–299, August 2019, doi: <https://doi.org/10.1016/j.ejor.2019.01.061>
- [43] A. Kandakoglu, A. Sauré, W. Michalowski, M. Aquino, J. Graham, and B. McCormick, “A decision support system for home dialysis visit scheduling and nurse routing”, *Decis Support Syst*, vol. 130, no. 113224, March 2020, doi: <https://doi.org/10.1016/j.dss.2019.113224>
- [44] M. Yang, Y. Ni, and L. Yang, “A multi-objective consistent home healthcare routing and scheduling problem in an uncertain environment”, *Computers & Industrial Engineering*, vol. 160, no. 107560, Oct. 2021, doi: [10.1016/j.cie.2021.107560](https://doi.org/10.1016/j.cie.2021.107560).

- [45] H. Bae and I. Moon, "Multi-depot vehicle routing problem with time windows considering delivery and installation vehicles", *Applied Mathematical Modelling*, vol. 40, no. 13, pp. 6536–6549, Jul. 2016, doi: 10.1016/j.apm.2016.01.059.
- [46] E. Jabir, V. V. Panicker, and R. Sridharan, "Design and development of a hybrid ant colony-variable neighbourhood search algorithm for a multi-depot green vehicle routing problem", *Transportation Research Part D: Transport and Environment*, vol. 57, pp. 422–457, Dec. 2017, doi: 10.1016/j.trd.2017.09.003.
- [47] F. Glover, "Tabu search: Part I", *ORSA Journal on Computing*, vol. 1, no. 3, pp. 190–206, Aug. 1989. doi: <https://doi.org/10.1287/ijoc.1.3.190>
- [48] N. Mladenović and P. Hansen, "Variable neighborhood search", *Computers & Operations Research*, vol. 24, no. 11, pp. 1097–1100, Nov. 1997, doi: 10.1016/S0305-0548(97)00031-2.
- [49] J.-F. Cordeau, G. Laporte, and A. Mercier, "A unified tabu search heuristic for vehicle routing problems with time windows", *J Oper Res Soc*, vol. 52, no. 8, pp. 928–936, Aug. 2001, doi: 10.1057/palgrave.jors.2601163.
- [50] J. Bergstra, D. Yamins, and D. Cox. "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures", *International Conference on Machine Learning*, pp. 115-123, Feb. 2013, doi: <https://doi.org/10.48550/arXiv.1209.5111>
- [51] J. Bergstra, R. Bardenet, Y. Bengio, and B. K'egl. "Algorithms for hyper-parameter optimization", *Advances in Neural Information Processing Systems* 24, pp. 2546-2554, 2011.

Appendix

The data of the toy problem is given in the following tables. Hospitals and patients are represented as “H” and “P”. respectively.

Table A1. The time matrix between hospitals and patients

	H1	H2	H3	H4	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
H1	0	10.8	10.2	10.2	5.0	5.8	13.9	6.4	13.0	5.8	6.3	13.0	9.2	4.5	3.0	6.3	16.1	5.4	14.1	10.3	10.8	3.6	10.8	10.6
H2	10.8	0	8.1	8.1	15.6	12.0	4.1	17.2	7.2	16.3	11.0	2.2	2.0	9.4	13.4	13.9	5.4	16.1	6.4	11.0	15.3	14.2	10.0	2.8
H3	10.2	8.1	0	14.4	13.2	14.8	7.6	15.1	14.9	13.3	14.4	9.1	8.5	6.0	13.2	8.9	10.8	15.0	5.7	3.2	19.4	12.0	2.0	10.4
H4	10.2	8.1	14.4	0	14.8	7.1	12.1	15.7	3.6	15.8	5.7	9.5	6.4	12.2	11.2	16.0	12.2	13.9	14.4	16.6	8.1	13.6	16.1	5.4
P1	5.0	15.6	13.2	14.9	0	9.0	18.4	2.0	17.9	1.0	10.0	17.8	14.1	7.3	4.0	5.4	20.8	2.8	18.0	12.0	13.3	1.4	13.0	15.6
P2	5.8	12.0	14.8	7.1	9.0	0.0	16.0	9.2	10.6	10.0	1.4	14.1	10.0	9.9	5.0	12.1	17.3	7.3	17.3	15.6	5.0	8.1	15.8	10.4
P3	13.9	4.1	7.6	12.1	18.4	16.0	0.0	20.1	10.6	18.9	15.0	2.8	6.1	11.4	16.8	15.6	3.2	19.3	3.2	10.8	19.4	17.0	9.5	6.7
P4	6.4	17.2	15.1	15.7	2.0	9.2	20.1	0.0	18.9	2.2	10.4	19.4	15.6	9.2	4.5	7.3	22.5	2.0	19.9	14.0	13.0	3.2	15.0	17.0
P5	13.0	7.2	14.9	3.6	17.9	10.6	10.6	18.9	0.0	18.8	9.2	7.8	6.3	14.0	14.4	18.2	9.8	17.2	13.5	17.5	11.4	16.6	16.8	4.5
P6	5.8	16.3	13.3	15.8	1.0	10.0	18.9	2.2	18.8	0.0	11.0	18.4	14.9	7.6	5.0	5.1	21.4	3.6	18.4	12.0	14.3	2.2	13.0	16.4
P7	6.3	11.0	14.4	5.7	10.0	1.4	15.0	10.4	9.2	11.0	0.0	13.0	9.0	10.0	6.1	12.6	16.1	8.5	16.5	15.6	5.0	9.0	15.6	9.2
P8	13.0	2.2	9.1	9.5	17.8	14.1	2.8	19.4	7.8	18.4	13.0	0.0	4.1	11.4	15.7	15.8	3.2	18.4	5.8	12.2	17.1	16.4	11.0	4.1
P9	9.2	2.0	8.5	6.4	14.1	10.0	6.1	15.6	6.3	14.9	9.0	4.1	0.0	8.5	11.7	13.0	7.3	14.4	8.1	11.2	13.3	12.7	10.4	2.0
P10	4.5	9.4	6.0	12.2	7.3	9.9	11.4	9.2	14.0	7.6	10.0	11.4	8.5	0.0	7.3	4.5	14.1	9.0	10.8	5.8	14.9	6.1	6.3	10.4
P11	3.0	13.4	13.2	11.2	4.0	5.0	16.8	4.5	14.4	5.0	6.1	15.7	11.7	7.3	0.0	7.8	18.8	2.8	17.1	13.0	9.5	3.2	13.6	12.8
P12	6.3	13.9	8.9	16.0	5.4	12.1	15.6	7.3	18.2	5.1	12.6	15.8	13.0	4.5	7.8	0.0	18.4	8.1	14.4	7.1	17.0	5.0	8.2	14.9
P13	16.1	5.4	10.8	12.2	20.8	17.3	3.2	22.5	9.8	21.4	16.1	3.2	7.3	14.1	18.8	18.4	0.0	21.5	6.0	13.9	20.0	19.4	12.6	7.0
P14	5.4	16.1	15.0	13.9	2.8	7.3	19.3	2.0	17.2	3.6	8.5	18.4	14.4	9.0	2.8	8.1	21.5	0.0	19.4	14.3	11.0	3.2	15.1	15.6
P15	14.1	6.4	5.7	14.4	18.0	17.3	3.2	19.9	13.5	18.4	16.5	5.8	8.1	10.8	17.1	14.4	6.0	19.4	0.0	8.6	21.2	16.8	7.2	9.2
P16	10.3	11.0	3.2	16.6	12.0	15.6	10.8	14.0	17.5	12.0	15.6	12.2	11.2	5.8	13.0	7.1	13.9	14.3	8.6	0.0	20.5	11.2	1.4	13.2
P17	10.8	15.3	19.4	8.1	13.3	5.0	19.4	13.0	11.4	14.3	5.0	17.1	13.3	14.9	9.5	17.0	20.0	11.0	21.2	20.5	0.0	12.6	20.6	13.0
P18	3.6	14.2	12.0	13.6	1.4	8.1	17.0	3.2	16.6	2.2	9.0	16.4	12.7	6.1	3.2	5.0	19.4	3.2	16.8	11.2	12.6	0.0	12.0	14.2
P19	10.8	10.0	2.0	16.1	13.0	15.8	9.5	15.0	16.8	13.0	15.6	11.0	10.4	6.3	13.6	8.2	12.6	15.1	7.2	1.4	20.6	12.0	0.0	12.4
P20	10.6	2.8	10.4	5.4	15.6	10.4	6.7	17.0	4.5	16.4	9.2	4.1	2.0	10.4	12.8	14.9	7.0	15.6	9.2	13.2	13.0	14.2	12.4	0.0

Table A2. Caregivers and patients” time windows and operation times (in minutes)


No	e_i for “H” a_i for “P”	l_i for “H” b_i for “P”	D_i
H1	480	1080	–
H2	480	1080	–
H3	480	1080	–
H4	480	1080	–
P1	900	960	56
P2	720	1080	45
P3	480	540	28
P4	840	960	47
P5	900	1020	26
P6	540	960	54
P7	600	720	42
P8	840	1020	56
P9	660	720	39
P10	600	960	42
P11	780	1080	28
P12	660	840	21
P13	540	1020	45
P14	540	600	42
P15	480	540	35
P16	780	1080	55
P17	780	960	39
P18	600	1080	37
P19	480	660	20
P20	660	1080	56


Table A3. Comparison of insertion types: i) balanced-insertion, ii) random insertion

Problem no	Insertion type	TS-VNS (Best)	TS-VNS (Average)
pr-l-1	i	4.1	5.1
	ii	6.7	7.5
pr-l-2	i	7.9	12.9
	ii	14.7	25.7
pr-l-3	i	8.0	11.9
	ii	11.5	17.5
pr-l-4	i	7.4	9.3
	ii	8.6	16.9
pr-l-5	i	4.4	6.7
	ii	10	16
pr-l-6	i	6.7	7.9
	ii	13.9	19.0
pr-l-7	i	6.2	7.4
	ii	9.2	12.1
pr-l-8	i	7.7	9.8
	ii	11.9	12.8
pr-l-9	i	5.9	7.0
	ii	14.2	20.2
pr-l-10	i	8.5	10.0
	ii	20.4	33.2
pr-l-11	i	1.1	1.2
	ii	2.8	3.0
pr-l-12	i	1.9	2.1
	ii	3.8	4.1
pr-l-13	i	3.9	4.8
	ii	6.6	13.2
pr-l-14	i	2.8	3.8
	ii	4.6	8.9
pr-l-15	i	2.7	2.9
	ii	7.2	7.9
pr-l-16	i	4.2	5.2
	ii	7.6	11.4
pr-l-17	i	1.6	2.2
	ii	3.3	3.9
pr-l-18	i	3.7	3.9
	ii	4.9	7.1
pr-l-19	i	3.7	4.5
	ii	8.3	10.9
pr-l-20	i	5.2	5.5
	ii	9.3	12.1

A Novel Clustering-based Forecast Framework: The Clusters with Competing Configurations Approach

¹Miray ALP, ^{*2}Gökhan DEMİRKIRAN

¹Department of Electrical and Electronics Engineering, Yaşar University, İzmir, Türkiye, mirayalp94@gmail.com 

²Department of Electrical and Electronics Engineering, Yaşar University, İzmir, Türkiye, gokhan.demirkiran@yasar.edu.tr 

Abstract

Accurate aggregate (total) short-term load forecasting of Smart Homes (SHs) is essential in planning and management of power utilities. The baseline approach consists of simply designing and training predictors for the aggregated consumption data. Nevertheless, better performance can be achieved by using a clustering-based forecasting strategy. In such strategy, the SHs are grouped according to some metric and the forecast of each group's total consumption are summed to reach the forecast of aggregate consumption of all SHs. Although the idea is simple, its implementation requires fine-detailed steps. This paper proposes a novel clustering-based aggregate-level forecast framework, so called Clusters with Competing Configurations (CwCC) approach and then compares its performance to the baseline strategy, namely Clusters with the Same Configurations (CwSC) approach. The configurations in the name refer to the configurations of ARIMA, Multi-Layer Perceptron (MLP), and Long Short-Term Memory (LSTM) forecasting methods, which the CwCC approach uses. We test the CwCC approach on Smart Grid Smart City Dataset. The results show that better performance can be achieved using the CwCC approach for each of the three forecast methods, and LSTM outperforms other methods in each scenario.

Keywords: clustering; deep neural networks; short term load forecasting; smart grid.

1. INTRODUCTION

Energy is regarded as the most crucial aspect of a region's social and economic growth since it significantly contributes to its progress and improved economy. Concerns about its uninterrupted supply especially to the vital operations arose as the electricity consumption has risen dramatically in recent years, primarily due to population, home, and commercial floor space increases [1]. While installing backup power grid infrastructures is one (costly) option, the regulation of consumption through demand response programmes seems one practical and cheap way to maintain the supply-demand balance by restricting the demand (The term demand is interchangeably used for the consumption or the load) to the supply. One strategy to incentivize customers to reduce consumption to meet supply is to use reward-based response programs [2]. Reward-based programs can also reduce operating costs of grids, improving the economic conditions of both providers and consumers [3].

Operations of demand response programmes in modern power systems generally require the information of future load consumption over an interval of 1-hour to 1-week (short-term load forecasting) which may be at the aggregate level or the residential level. For example, in one

implementation scenario of the demand response programme, the utility operator forecasts the aggregated residential load (i.e., total load) in an area and decides if an action is needed to be taken based on that aggregated load forecast [4]. Simultaneously, it can also use algorithms based on the individual customer data (such as load profiles and forecasts of individual customer consumptions) to evaluate the potential of the customers to participate the program, and design accordingly the specification of the demand response to be initiated. Apparently, in such implementations the success of the demand response programmes relies on the accuracy of the forecasts. Having said that, in this study, our aim is to increase the performance of aggregate-level forecasting. Our strategy is to cluster the aggregate consumption into clusters of similar and better-forecastable consumption profiles and then sum the forecast of each group to reach the forecast of aggregate consumption.

Mostly clustering is used as a tool to comprehend user profiles to be used in demand response scenarios. For example, time series clustering can model consumer behavior for decision-making in Swedish electricity market that is subject to ongoing developments [5]. The diversity of user profiles in Swish electricity market is inevitable as shown by [6] via comparing different clustering approach.

Moreover, Reference [7] emphasizes the need for using diverse profiles instead of relying on one standard pattern for policy makers. One challenge with consumer behaviour is that the user load profiles are changeable, thereby diminishing the efficiency of static load profiling approaches. In this direction, a recursive clustering algorithm that continuously update the load profiles using newly updated data has been proposed by [8]. Load profile classification can also be improved by a spectral clustering algorithm [9]. Despite the efforts for more accurate load profiling, the references [5]–[9] do not address how to utilize

such load profiling in forecasting. A customizable toolbox Divinus which uses clustering methods both for user profiling and for forecasting total consumption is proposed by [10]. However, they use clustering to forecast by remembering the previous data in a non-generalizable way. Clustering is also used to increase the performance of individual residential load forecasting, since clustered load profiles in the cluster which the residential belongs to are similar, thus can be used increase the data diversity of individual load forecasting and decrease the overfitting [11], [12].

Table 1. The references using cluster-based aggregate-level forecasting.

Reference	Dataset	Total number of houses	Forecast Method	Inputs (Features)	Cluster Method	Clustering Objective
[13]	CER	6000	PARX	Load signal, Calendar, Temperature.	spectral clustering	similarity matrix
[14]	CER	3639	Linear Regression, MLP, SVR, ARIMA.	Load signal, calendar features.	Max-AC, Min-StDev, Max-Sim.	24-hour load profile
[15]	CER	3176	MLP	Load signal, Calendar features, Temperature.	k-means	Regression Coefficients, 24-hour load profile.
[16]	Enernoc, CER	782	SVM, MLP, Linear Regression, Ensemble.	Load signal	greedy algorithm	entire sequence
[17]	CER	3176	MLP, Deep Belief Network.	Load signal, Calendar features, Temperature.	k-shape	24-hour load profile
[18]	Slovak Electricity consumption	1152	LSTM	Load signal	k-shape	entire sequence
[19]	Arbon(Swiss), CER.	7500	LSTM variant	Load signal	spectral clustering	similarity matrix

Most load forecasting studies use point forecasts based on the aggregate system-level data classified as statistical approaches, machine learning techniques, or hybrid models [20]. However, the number of studies using clustering techniques to improve aggregate-level forecasting are relatively few. We summarize the most significant 7 papers, which share the same aim and strategy with this paper in Table-1. These papers experiment with clustering-based aggregate load forecasting, from different point of views. 6 of 7 papers use Commission for Energy Regulation (CER) dataset from Ireland whilst one of them uses Slovak Electricity Consumption. The total number of houses range from 782 to 7500. The forecasting algorithms used are MLP (Multi-Layer Perceptron), SVR (Support Vector Regression), ARIMA, Deep Belief Networks, LSTM (Long Short-Term Memory), Linear Regression Ensemble and PARX (Periodic Auto-Regressive model with Exogenous variables). As the input features, calendar and temperature information can be used together with the load signal itself, specifically in Artificial Intelligence (AI) techniques. The references in Table-1 use similarity matrix, 24-hour load

profile, regression coefficients, or the entire sequence as the input to the clustering algorithms. Overall, in all the references in Table-1, clustering is shown to be an effective technique that increases the forecasting performance regardless of the clustering technique used, though some clustering techniques may give better forecasting performance [17].

However, the studies listed in Table-1 do not directly experiment with the strategy, rather use or modify it as a tool to tackle a particular issue. In this paper, we conceptualize this strategy as a standard framework, and further propose an immediate-update to it, so called Clusters with Competing Configurations, so that the clustered groups can have their own forecast structures that fits better to their load profile characteristics. These forecast structures can have different configurations of the same or a different methodology.

Regarding the methodologies, we conduct a comparison of ARIMA, MLP, and LSTM in clustering-based aggregate

load forecasting. That is, we compare the most-adopted linear statistical method ARIMA with its widely used nonlinear AI counterparts, i.e. MLP and LSTM. MLP is a shallow AI technique, whilst LSTM is the state-of-the-art deep AI technique that has been very popular recently. A good configuration parameter of a naive method can have a better performance than a poorly configured state-of-the-art model. Thus, to make sure that the best possible performances of each forecasting algorithm is adopted in forecasting, we propose a clustering-based aggregate-level short-term load forecasting, so called Clusters with Competing Configurations (CwCC) approach. CwCC approach includes the configuration spaces of these algorithms, which the references in Table-1 lack, such as the different sets of lags in ARIMA, the number of neurons in MLP, and the different number of hidden units in LSTM etc. That is, in our strategy we have different configurations of the same forecasting method competing for each cluster's forecast. That is, there can be different configurations of the same forecasting method for different clusters. Additionally, to compare the effect of adding configuration space in CwCC approach, we test the case of using the same configuration (but different weights learned during training of each cluster) of the same forecasting method across all clusters, which we call as Clusters with the Same Configuration (CwSC) approach.

Furthermore, for a fair comparison, we use only the load signal as the input to the MLP and LSTM, as the ARIMA can only handle 1-dimensional data. As for the clustering input, we use 24-hour load profile, which is the most practical option compared to the similarity matrix that needs several user-tuned parameters, and to the entire sequence which is subject to the curse of dimensionality. Additionally, we provide clear steps and reproducible results for those who wish to implement the cluster-based strategy for real-life use.

Finally, we use Smart Grid Smart City dataset, which can provide 12641 residential load time series in the context of this study, that is more residential than CER and Slovak Electricity Consumption datasets that references in Table-1 use, can provide. Our results indicate that CwCC approach is an effective technique to increase the aggregate-level forecast performance and LSTM models in general gives better performance in terms of MAPE.

The rest of this paper is organized as follows. Section 2 presents the elements of the proposed framework: k-means clustering, Long Short-Term Memory (LSTM), ARIMA, Multi-Layer Perceptron (MLP). Section 3 describes the SGSC dataset and analyzes different time scales' effect on energy consumption and patterns essential for modeling. Section 4 assesses the proposed method by evaluating the accuracy of forecasts. Finally, Chapter 5 presents conclusions and recommendations.

2. THE REVIEW OF THE ALGORITHMS

2.1. k-means Clustering

K-means clustering is a method of logically classifying raw data and searching for hidden patterns in datasets. It arranges data into fragmented (k-)clusters such that data in one cluster

is identical to that in another, but data in other clusters differs. K-means is an iterative, numerical, unsupervised, and non-deterministic approach that assigns n observations to exactly one of the k clusters defined by centroids, where k is a predetermined number. It is straightforward and easy; hence it has shown to be a very useful strategy for producing good clustering results in many practical applications.

2.2. ARIMA

ARIMA is a computational iteration with polynomial structure in the form of:

$$x_t = c + x_{t-1} + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

Where ϕ_i refers to the autoregressive coefficients, θ_i to the moving average coefficients, c to the bias. $\varepsilon_t = x_t - \widehat{(x_t)}$ is the error which is assumed to be a Gaussian distribution with constant variance. The zero coefficients are unknown but can be determined using statistical analysis, such as autocorrelation (In other words, autoregressive and moving average lags are unknown, but can be inferred by autocorrelation). After that the exact values of non-zero coefficients and bias are learned during training.

In the following, we carry out an autocorrelation analysis of the data used in this paper (See Chapter 3.1 for data) to find the lags of non-zero ϕ_i and θ_i . Autocorrelation analysis in Fig.1 reveals that the consumption at a given time is more related to previous increments, and as well as the same time of the previous days. The correlation decreases as the number of previous days increases. In designing the configuration space in Table 2 which will be used in experimentation, we have used lag samples from the peaks of the sample autocorrelation function sketched in Fig. 1 and our pre-experiments. Specifically, we have designed an experiment from the lags around the peaks of 1st, 48th (24th hour or previous day) and 96th (48th hour or the previous two day) lag which represent the peak correlations to the previous samples of the data (Table 2).

Table-2. Configuration Space of ARIMA models.

Index	AR Lags	MA Lags
1	1 2	1 2
2	1 2 3	1 2 3
3	1 2 3 4	1 2 3 4
4	1 2 3 4	1 2 47 48 49
5	1 2 3 4	1 2 95 96 97
6	1 2 47 48 49	1 2 3 4
7	1 2 47 48 49	1 2 47 48 49
8	1 2 47 48 49	1 2 95 96 97
9	1 2 95 96 97	1 2 3 4
10	1 2 95 96 97	1 2 47 48 49
11	1 2 95 96 97	1 2 95 96 97
12	1 2 3 4 47 48 49	1 2 3 4
13	1 2 3 4 47 48 49	1 2 47 48 49
14	1 2 3 4 47 48 49	1 2 95 96 97
15	1 2 3 4 95 96 97	1 2 3 4
16	1 2 3 4 95 96 97	1 2 47 48 49
17	1 2 3 4 95 96 97	1 2 95 96 97

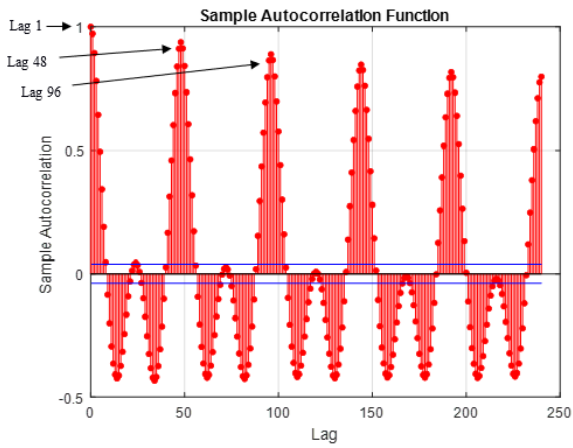


Figure 1. Autocorrelation analysis of data.

2.3. Multi-Layer Perceptron (MLP)

MLP is a type of feedforward neural network, which in fact uses a structured function with unknown parameters (weight parameters) to approximate a hypothetical function which perfectly maps (training) input to (training) outputs. MLPs are in the form of:

$$\sigma_n(W_n \dots (\sigma_1(W_1(\sigma_0(W_0x + b_0)) + b_1)) + b_n) = y \quad (2)$$

Where n is the number of hidden layers, W_0, \dots, W_n and b_0, \dots, b_n are the unknown weight matrices, where one of the dimensions of each weight matrix is determined empirically (the number of neurons), whereas the second dimension is determined by the rules of matrix multiplication. $\sigma_0, \dots, \sigma_n$ are the activation functions, which can be a tangent hyperbolic, a sigmoid, a linear function etc., again determined empirically. x is as the architecture space of MLP we have used one hidden layer whose number of neurons are taken from the set $\{5, 10, 15, 20\}$. For the optimizer, we have used Levenberg-Marquardt as it outperformed other algorithms in general in our pre-experiments. Additionally, the two or more hidden layers greatly decreased the test performance, thus they are discarded from the configuration space.

2.4. Long Short-Term Memory (LSTM)

LSTM is a type of RNN, which can learn long-term dependencies. LSTM, initially proposed by [21] is a feed-forward neural network that unfolds in time (Fig. 2). The unknown matrices are learned during training process of LSTM by backpropagation through time algorithm, which prevents the vanishing gradient problem.

The configuration space of LSTM used in this paper are hidden unit layers (50,100,200,500), optimizers (adam, sgd, rmsprop) and the number of epochs (500,1000,2000), which result in $(4 \times 3 \times 3)$ 36 different configurations. Each of the configuration numbers from 1 to 36 consists of hyperparameters which must be empirically determined so that the best model can be found. In our pre-experiments, we have observed that the deviation of learning rate from 0.005 has negative effect on results in general. Thus, we fix it to 0.005. The design of the configuration space is needed several pre-experiments with data.

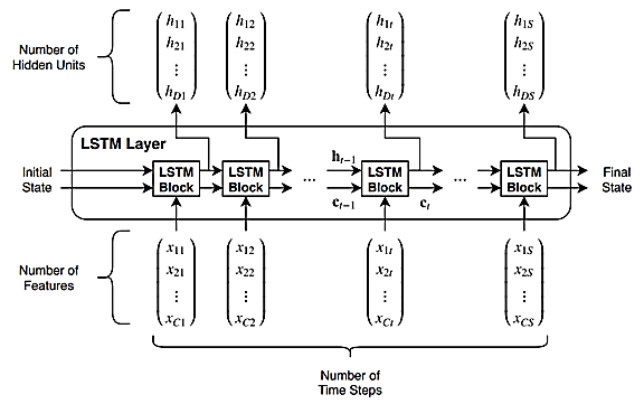


Figure 2. The Computational Architecture of LSTM in MATLAB®.

3. PREPROCESSING DATASET

3.1. Smart Grid Smart City Dataset

In this research, Smart Grid Smart City dataset from Australia [22] is used. The field where data is collected includes energy distribution, transmission businesses, technology firms, universities, and the CSIRO. Since 2009, SGSC project have been deployed in eight local government regions in New South Wales, covering 30,000 residences and serving as one of the world's largest commercial-scale smart grid technology assessment projects. Climate zone, household income, housing type, electricity consumption, and gas consumption level are used to segment the population, yielding 108 socio-demographic statistics cells in total [23].

SGSC data is obtained from 13,735 customers between the years 2010-2014. It includes comprehensive data on appliance use, climate, retail, and distributor product offers, and other associated elements, as well as one of the few linked sets of consumer time of use (with half-hour increments) and demographic data for Australia. Electricity consumption interval readings, home area network plug readings, peak events, peak events reaction, and offer and acceptance of the event signal are some of the data resources for this dataset.

3.2. Data Preprocessing

SGSC dataset is about 20 gigabyte volume of .csv file, which cannot be loaded into a conventional laptop due to ram restrictions. Thus, first we divide the data into .csv files of 1 gigabyte data so that it can be loaded to the ram and MATLAB can reach data. After that, we store 13,735 .mat files (MATLAB® storage file) containing the time series data of each house by carefully searching the houses in that 20 gigabyte data.

The electrical consumption data in SGSC dataset has half-hourly intervals. Additionally, the number of days of all recorded data were very different from household to other. For example, it has been seen that there are households with 2 days of data, as well as households with more than 800 days of data (Fig. 3).

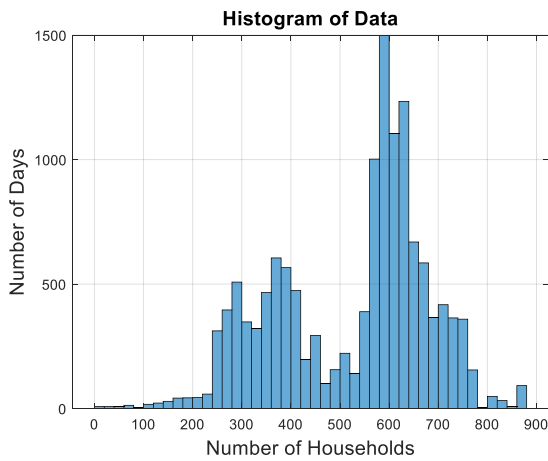


Figure 3. Histogram of number of days in the dataset.

Another important consideration when dealing with time series data is there can be missing as well as double-entered values, which is very problematic due to shifting in data. Thus, first we examine the dataset values according to the below Algorithm-1 and create a proper time series.

Algorithm 1. Fixing the missing and repeated measurements.

Data Preprocessing

- 1: Upload a data in the form of table
- 2: Find improper increments in the data
- 3: if replicated is found
- 4: Remove duplicate data from table
- 5: Update the data as the mean of two-readings
- 6: Start the algorithm from beginning
- 7: end
- 8: if missing is found
- 9: Find the mean of the data before and after
- 10: Replace the missing data
- 11: Start the algorithm from beginning
- 12: end

3.3. Finding a Proper Subset of Data

Figure 4 shows the datetime interval in which a customer participates to the SGSC program (blue dots) and leaves the program (red dots). We decide that having 8 weeks of data for training, 2 weeks for validation, and 2 weeks for testing is appropriate when compared to the studies in Table-1, thus 12-week duration of data is used. The interval date of 12-week duration is specifically chosen such that there will be the highest number of houses that can provide data of 12-week duration, by the help of Fig. 5. Fig.5 depicts the number of houses that can provide data on the interval of $[\text{min_starting_date} + I\text{thweek}, \text{min_starting_date} + I\text{thweek} + X\text{weeks}]$, where $I(\text{thweek})$ increases from 0 to last possible number searching for the date and $X\text{weeks}$ is the duration, which is 12weeks in our case. In Fig.5 also, other $X\text{-week}$ durations are represented for comparison for the reader.

When searching for the interval date of 12-week duration, we also looked for electricity consumption in one season

only to avoid seasonal factors. The highest number of houses, which is 12641, in one season using 12-week duration occurs in the interval date of [03-June-2013, 26-August-2013], i.e. [the first Monday of June, and the last Sunday of August] (summer in northern semi-sphere) (Fig. 4). The aggregation of 12641 household consumptions for train, validation and test data is depicted in Fig. 6.

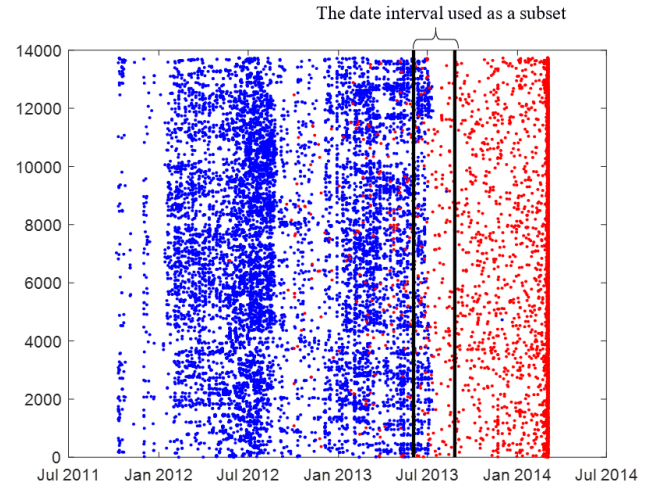


Figure 4. The starting and end date of households participating to SGSC program

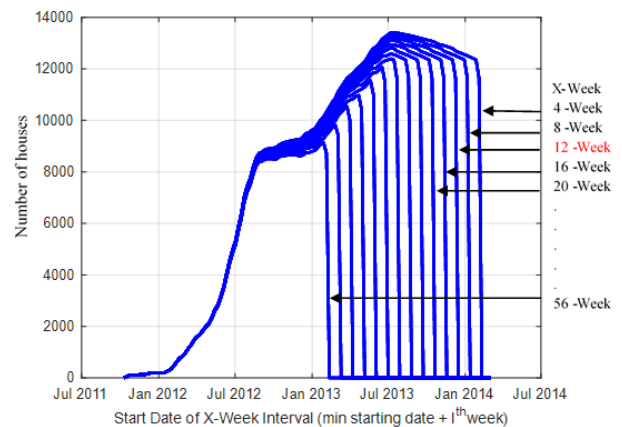


Figure 5. The moving Xweek interval vs. the number of houses that can provide data in that interval.

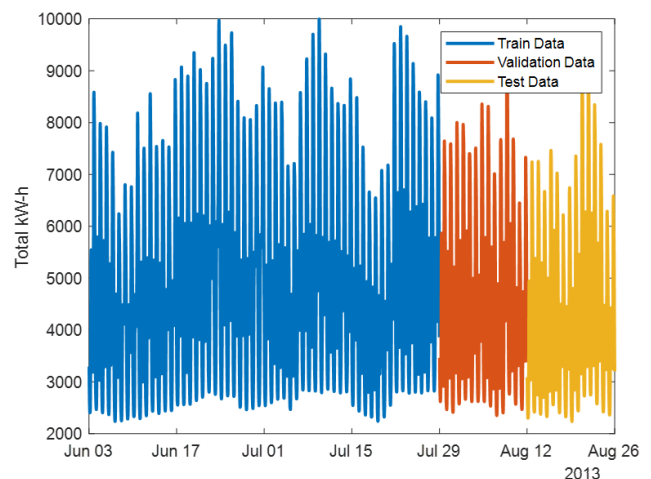


Figure 6. Total Consumption between [03-June-2013,26-August-2013] obtained by aggregating 12641 households.

3.4. k-means Clustering

In this subsection, we cluster the normalized consumption profiles of the households. The normalized consumption profiles are the mean of the daily consumptions for each measurement averaged over total 84 days of consumptions for each house (12 Weeks = 84 days). That is each house has a normalized load profile of a 1×48 (0.5 hour of incremental during day) vector, each element of vector corresponding to the average consumptions in that date interval. For clustering, we have experimented with different number of clusters, ranging from 2 to 10.

Fig. 7 shows an example of the total consumption of each cluster in case of 4 cluster used and the combined total consumption of all clusters. Fig. 8 depicts the case of 4 clusters of normalized profiles of the 12641 households. The shape of average load profile of each cluster (the thick black line) shows how 4 clusters differ from each other by their consumption characteristics.

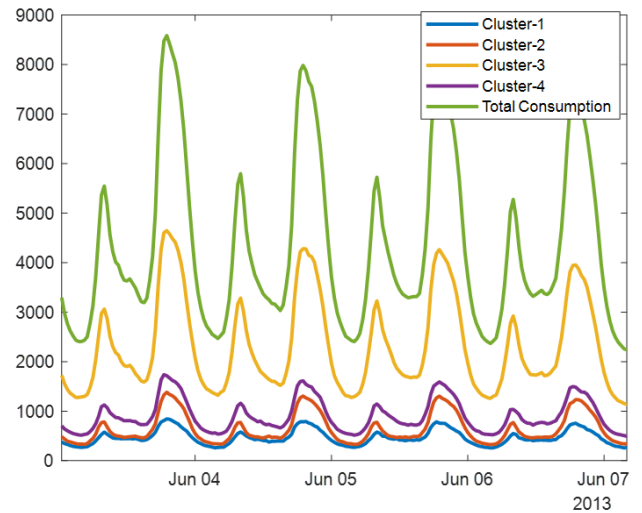


Figure 7. A sample of total consumption and clustered consumptions in case of 4-clusters.

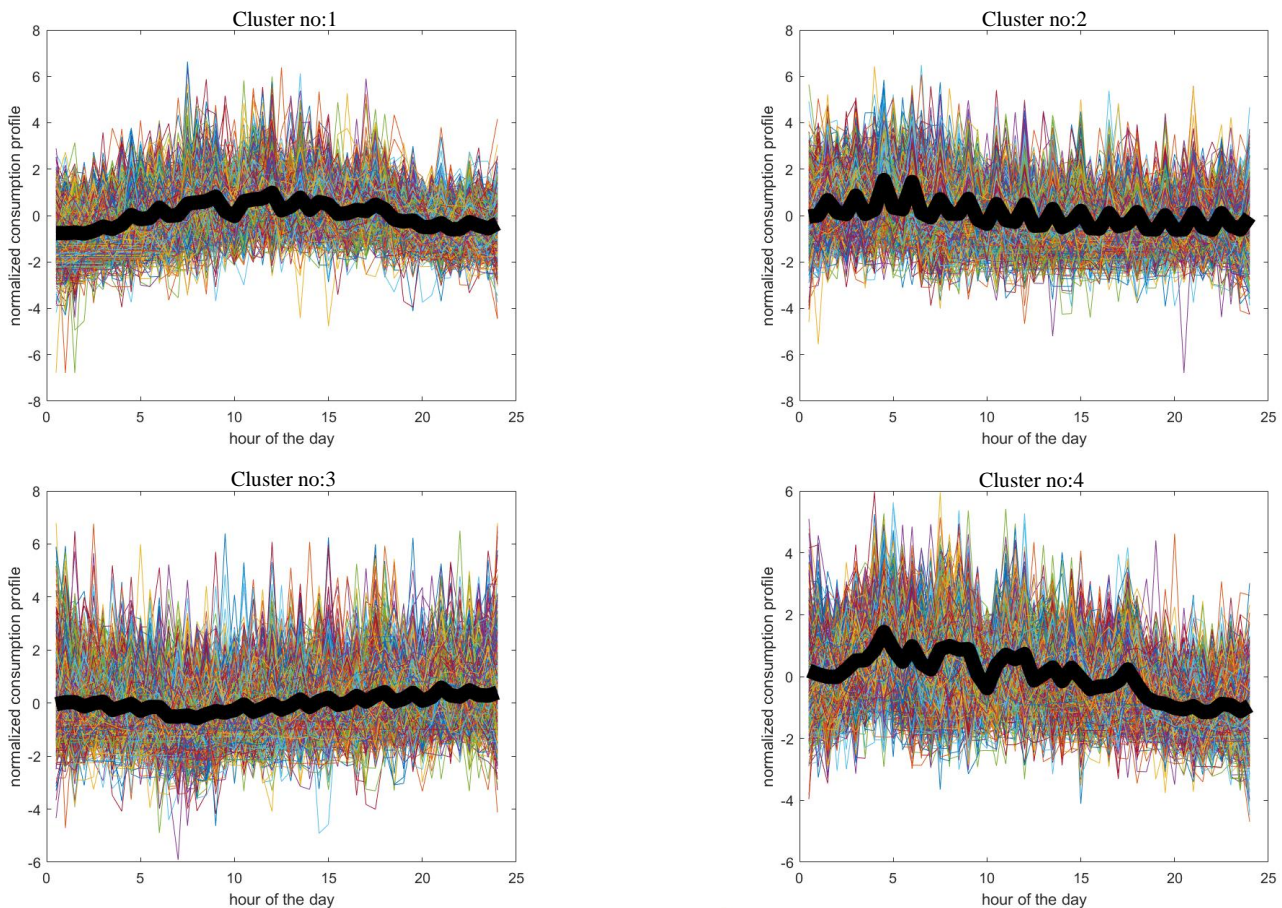


Figure 8. 4-clusters of the normalized load profiles. The tick line is the mean of the corresponding cluster.

4. RESULTS AND DISCUSSION

In this section, we compare the performance increases due to Clusters with Competing Configurations (CwCC) approach and the Clusters with the Same Configuration (CwSC) approach, with respect to the base non-aggregated (or one-cluster) forecasting performance. In evaluation of these two clustering-based approaches, 17, 5, and 36 different configurations of ARIMA, MLP, and LSTM, respectively, are used.

To be more precise in defining the two clustering-based approaches, we use the indexes in Fig. 9 to characterize the predictor and the clustering experiment. For instance, ${}^3P_5^A$ refers to the predictor using ARIMA method with 5th configuration trained using (and assigned to) the 2nd cluster in the experiment where the total consumption is disaggregated into 3 clusters.

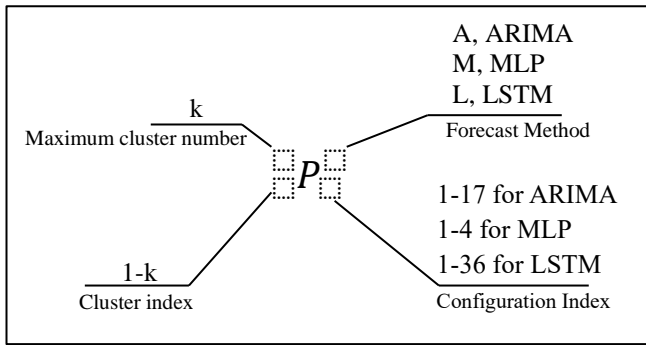


Figure 9. Indexes used in defining predictors.

Now, we explain the steps of CwCC approach using ARIMA method as an example with the help of Fig. 10. First, the (normalized) load profiles are clustered into k -clusters, where k is a predetermined number ranging from 1 to 10. Then, we train 17 differently configured ARIMA predictors, where ARIMA and the configuration space (Table-2) are predetermined, for each cluster on the total consumption data of each of them using their training part of the consumption data only. Then, for each cluster, the best ARIMA predictor is determined based on their performance on the validation part of the cluster's total consumption data. Then, to obtain the aggregate level forecasting of 12641 households, only the forecasts of the best predictors of each cluster is summed as sketched in Fig. 10.

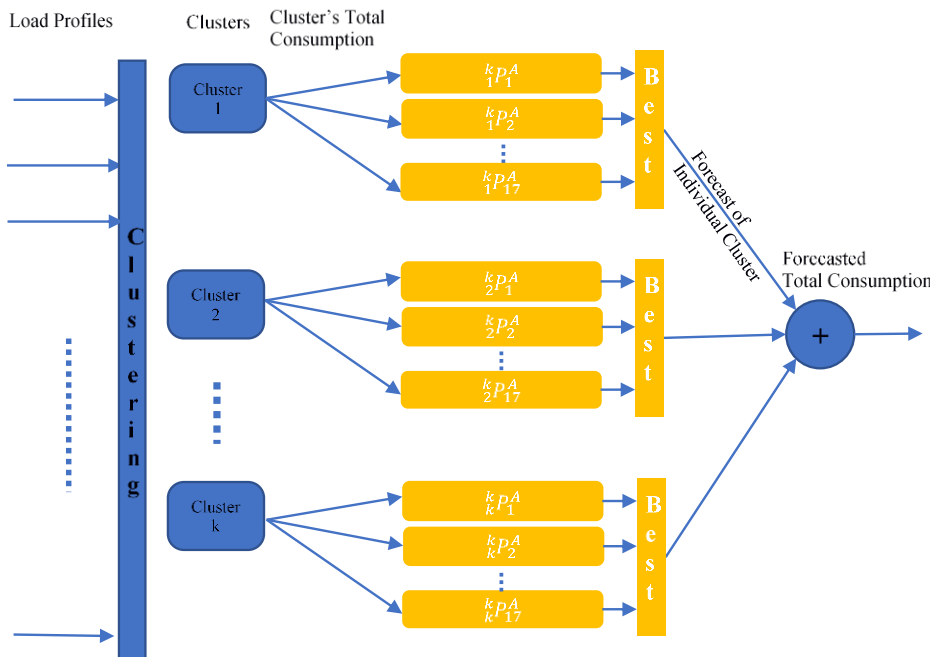


Figure 10. Flow chart for evaluating the performance of CwCC using ARIMA on forecasting of total consumption.

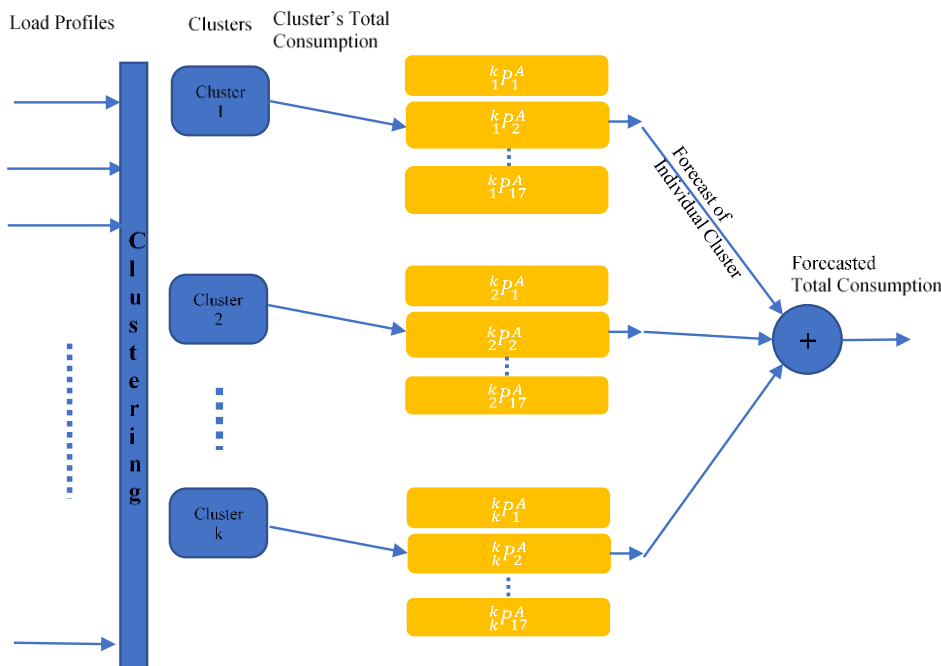


Figure 11. Flow Chart for Evaluating Performance of CwSC using 2nd configuration of ARIMA method on Forecasting of Total Consumption.

The difference of CwSC approach is that there is no best predictor selection step, thus the predictor to be used (i.e. configuration index) is predetermined (Fig. 11). For instance, in Fig. 11, the 2nd configuration of ARIMA is used in training each cluster's total consumption. Although each ARIMA model has the same configuration, their weight values differ due to training on different cluster's training data. The idea of CwSC is to picture the performance increase due to using different configurations which is best to that cluster. Since clusters are obtained using their load profiles, different lag configurations by searching for the best predictors as in CwCC is expected to give better performance. In contrast, in CwSC the same lag configurations are used across clusters.

The performances of CwSC and CwCC are reported using out-of-sample test data. That is, the test data is never introduced in the training and when picking the best predictors. In real-life usage of CwCC, an expert would have only training and validation data where he/she can design a CwCC system, whilst out-of-sample test data is where the CwCC is tested in real-life.

4.1. Performance based on ARIMA Method

We now evaluate the performances of CwSC and CwCC approaches using ARIMA as the forecasting method. In CwCC approach, each ARIMA model is trained on the train data of its corresponding cluster. Then, based on their Mean Absolute Percentage Errors (MAPE) (1) performances on the validation datasets, the best models are chosen, and assigned

to that cluster. After that we arrive at a one CwCC system ready to be tested on the test dataset.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|A_t - F_t|}{A_t} \quad (1)$$

Where n is the length of the data, t is the index, A_t is the actual value, and F_t is the forecast value.

In CwSC approach, each ARIMA model is trained on the train data of its corresponding cluster, and assigned to that cluster, and tested on the test dataset. We note that, at the end there are 17 different versions of CwSC system ready to be tested on test data, whilst there is only one CwCC forecasting system ready to be used. Thus, to report the results of these two clustering-based approaches we use boxplots or average performance for CwSC whilst we use one-line graphic for the CwCC. From one aspect, CwCC can be thought as the soft upper bound performance of CwSC approach.

The bar graphics in Fig.12a depicts the average of the train, the validation and the test performances of 17 CwSC systems in case of different number of clusters going from 1 to 10, where 1 refers to no disaggregation of 12641 household total consumption data. That is, there is only one cluster. We observe that as the number of clusters increases there are slight performance increases on the average of the train, validation and test performances. Additionally, we observe that the average train performance is the best whilst the test performance is the worst in each of the different cluster number, which is expected.

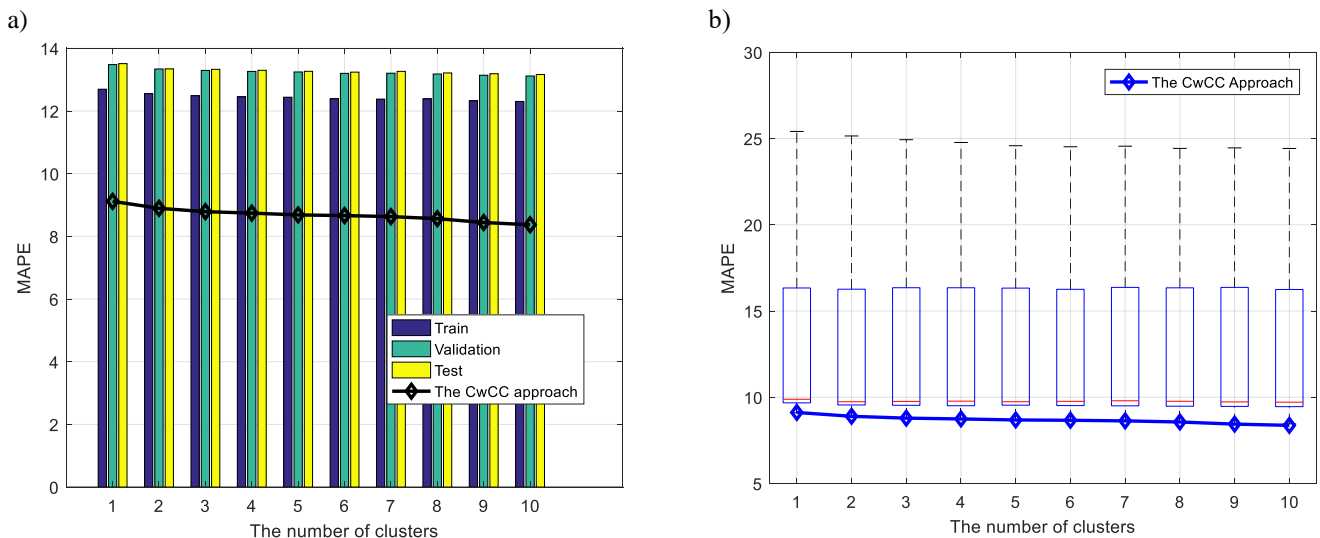


Figure 12 a) MAPE performance of the CwCC system using ARIMA method vs. CwSC system using ARIMA method. **b)** Boxplot of average performance of CwSC approach vs. the performance CwCC using ARIMA models.

The line-graphics in Fig.12a depicts the performance of CwCC system on the test data. We observe that CwCC test performance is way better than the average test performance of 17 CwSC systems (8.69 vs. 13.28 (average) respectively). To better compare the test performances of both systems, Fig. 12b depicts the boxplot (instead of average) test performances of 17 CwSC systems and the test performance of CwCC system. It is observed that CwCC system provides a minimum MAPE bound to the CwSC system.

The advantage of the CwCP approach lies in its use of the best predictor unique to that cluster. Differently than CwSC approach which uses one predictor across all clusters, there are clusters with unique predictors which outperform other predictors, and their forecasted total consumptions are summed to get final forecasted total consumption. Fig. 12b summarizes this advantage. Each cluster in the cases of different numbers of clusters (points on y-axis) has 17 ARIMA Models trained specifically to forecast that cluster's total consumption. The boxplots depict the case of using only

one ARIMA model (with the same lags but different coefficients, since they are trained on different clusters). The best performances can be seen by the lower whiskers of the boxes which coincides with the CwCC approach's performance. Thus, we can conclude that CwCC approach guaranteed the best performance. In the case of CwSC approach, one ARIMA model can perform good for some clusters whilst it can perform poor for the others. For example, in the case of 10-clusters and CwCC approach, the ARIMA Model indices for each cluster are: 11, 11, 17, 11, 11, 17, 13, 14, 11, 17. But in the case of CwSC approach, one configuration of ARIMA Model is used across all clusters.

4.2. Performance based on MLP Method

Our results indicate that as the number of clusters increases, there may be slight MAPE performance increases (although not monotonically) of the CwSC system on average and of the CwCC system on train, validation, and test datasets (Fig. 13a).

Figure 13b depicts the effectiveness of the CwCC system using MLP models. The performance of CwCC system is better than or equal to CwSC in 7/10 cases. And in the remaining cases, 3/10, the performance of CwCP is close to the best performance of CwSC. We note that in real life uses, the best MLP configuration of CwSC cannot be known beforehand, thus the best performance of CwSC is not guaranteed.

We observe that as the number of clusters increases there is not a monotonical increase in performance of CwCC approach using MLP models, but nevertheless the performance is better as the number of clusters increases. Thus, it is an effective clustering-based algorithm.

4.3. Performances based LSTM Method

Our results indicate that as the number of clusters increases, there may be performance increases (although not monotonically) of the CwSC approach (average) and of the CwCC approach in terms of MAPE on train, validation, and test datasets (Fig. 14a).

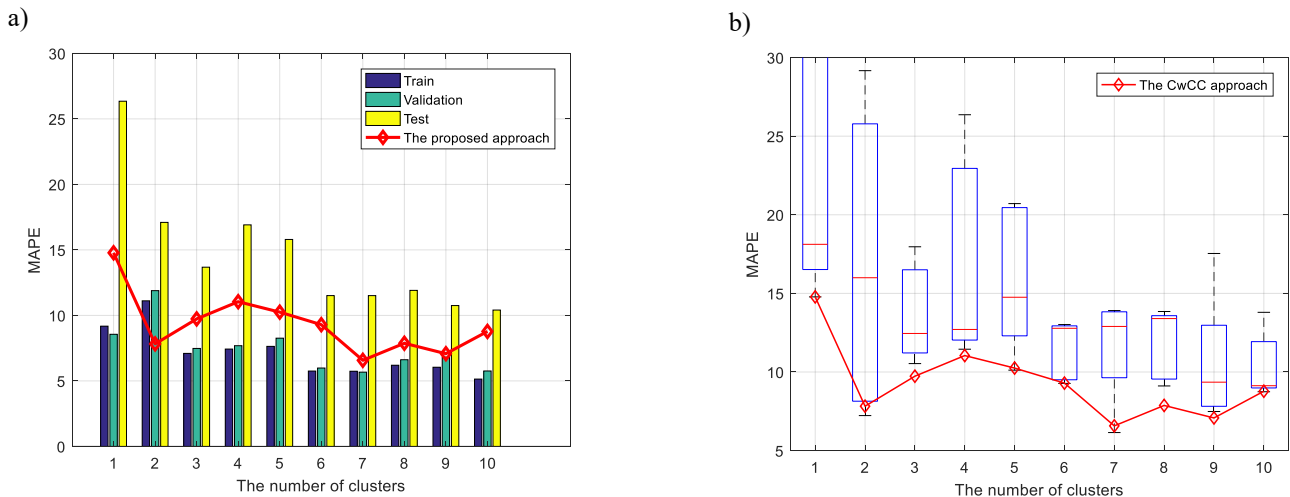


Figure 13. a) MAPE performances of the CwCC approach using MLP method vs. CwSC approach using MLP method. **b)** Boxplot of average performance of CwSC approach vs. the performance CwCC using MLP models.

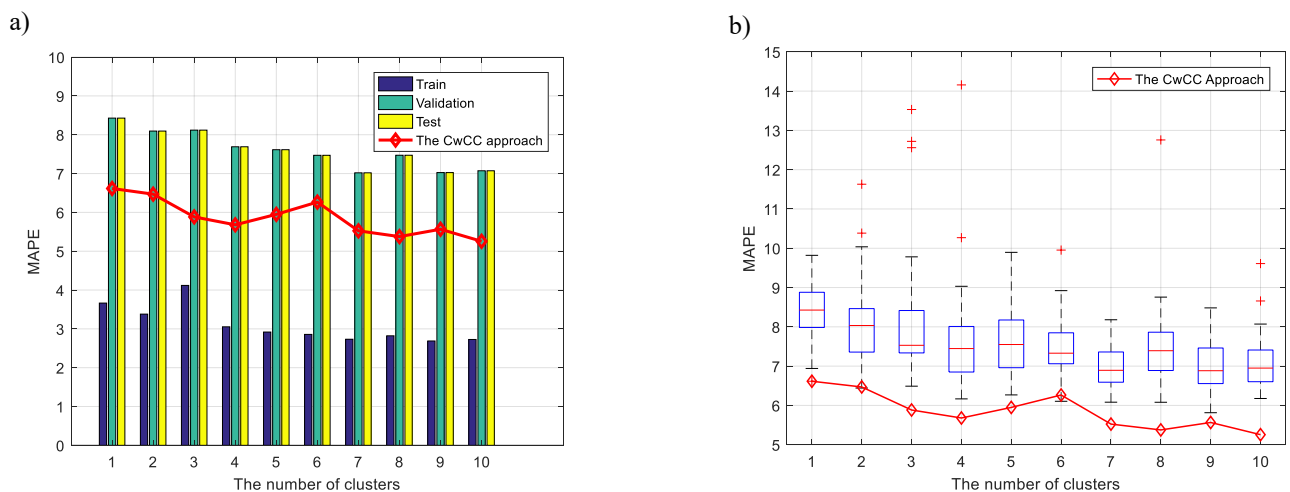


Figure 13 a) MAPE performance of CwCC approach using LSTM method vs. CwSC approach. **b)** Boxplot of average performance of CwSC approach vs. the performance CwCC using LSTM method.

4.4. Performances based LSTM Method

Our results indicate that as the number of clusters increases, there may be performance increases (although not monotonically) of the CwSC approach (average) and of the CwCC approach in terms of MAPE on train, validation, and test datasets (Fig. 14a).

Figure 14b depicts the effectiveness of the CWCC approach using LSTM models. The performance of CwCC is better than or equal to CwSC in 9/10 cases. And in the remaining case, 1/10, the performance of CwCC is close to the best performance of CwSC. Again, we note that the best LSTM configuration of CwSC cannot be known beforehand, thus the best performance of CwSC is not guaranteed. We observe that as the number of clusters increases there is not a monotonical increase in performance of CwCC approach using LSTM models, but nevertheless the performance is better than the case of non-aggregated forecasting (i.e., $k=1$).

Figure 14 compares the performances of CwCC approach using ARIMA, MLP, and LSTM models. The results indicate that LSTM is the best method, and it outperforms other methods, in all cases of different number of clusters.

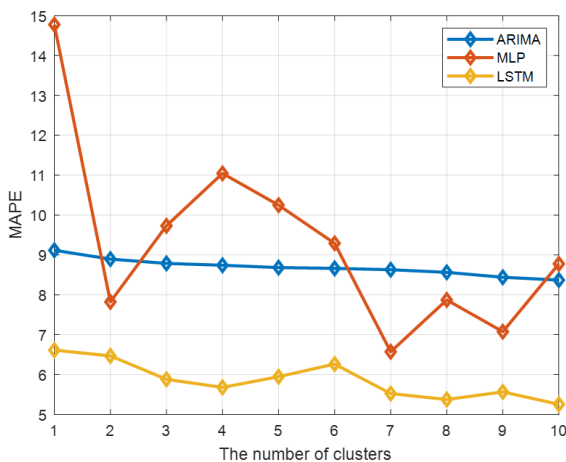


Figure 14. ARIMA, MLP, and LSTM performances vs. the number of clusters.

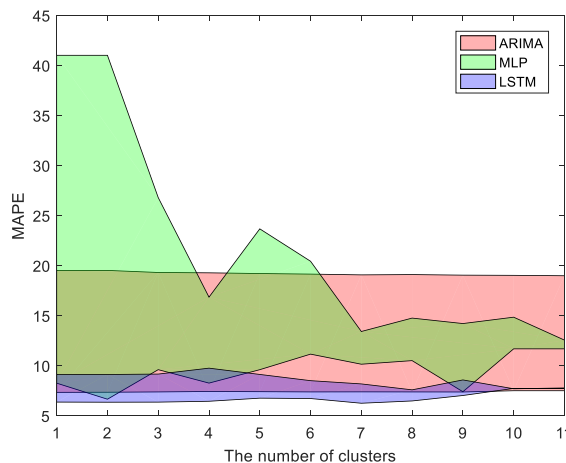


Figure 15. Mean plus standard deviations and mean minus standard deviations of ARIMA, MLP, and LSTM performances sketched as polygons vs. the number of clusters.

Moreover, Figure 15 compares the standard deviations of three methods, sketched as 2-dimensional polygons, where the upper line of each polygon is calculated as the mean plus standard deviation of the performances whilst the lower line is calculated as the mean minus standard deviation of the performances through different cluster numbers. It is observed that the performance characteristics of CwCC approach with ARIMA remain relatively stable. This might be due to the fact that ARIMA is a linear model and the sum of group forecasts would be numerically close to the aggregate-level forecast as a whole. The performance of MLP significantly improves with the number of clusters. Additionally, the standard deviation of the performances drops substantially increasing the certainty of the performance towards higher number of clusters. The performance of LSTM also increases with the number of clusters, and the standard deviation of the performances drops indicating the certainty towards higher number of clusters. As can be inferred from Figure 15, LSTM is a more reliable method to be applied in the CwCC approach than MLP especially when using a lower number of clusters. Moreover, as the number of clusters increases, the performance of LSTM statistically outperforms ARIMA and MLP in average, even surpassing their best cases. Thus, we conclude that LSTM is a statistically more reliable method to employ in a CwCC approach.

5. DISCUSSION

For clustering, we used 24-hour load profile characteristics. However, there is room for exploring novel model formulations to describe individual consumers, which can offer deeper understanding of how aggregate consumption patterns are formed. The proposed framework here can incorporate such novel formulations, as the literature in this field continues to grow.

Herein, we aimed to improve the forecast of aggregate-level individual power consumers. However, the aggregation phenomenon can appear also in other fields, such as wind forecasting, electric car availability. Thus, the proposed framework herein is likely to find application in other fields where the aggregated patterns may arise.

One drawback of our method is that it employs an ensemble of forecasters instead of a single one, which enhances the performance at the cost of increased complexity. The complexity of the method becomes a significant concern, especially when considering that demand response programs may be embedded into electronic devices rather than operating in a stand-alone computer. Additionally, in some cases the small amount of performance increase may not justify the level of complexity.

6. CONCLUSIONS AND RECOMMENDATIONS

We propose and evaluate the performance of the proposed CwCC approach for clustering-based aggregate-level short-term load forecasting using ARIMA, MLP and LSTM forecasting methods. We showed that the method is effective. Additionally, we provide clear and reproducible steps which can be useful for practitioners of this field.

We believe that the subject of clustering-based load forecasting deserves more research, as there is room for improvement. The proposed CwCC approach provides one step in this direction. In the future, CwCC approach can be upgraded by adding new components and can be tested by other forecasting methods, with more clusters. Nevertheless, CwCC approach is a practical method which can be readily used in real life to increase aggregate-level forecasting performance.

Author contributions: Concept – M.A, G.D; Data Collection &/or Processing – M.A, G.D; Literature Search – M.A, G.D; Writing – M.A, G.D.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study has received no financial support.

REFERENCES

- [1] G. Pau, M. Collotta, A. Ruano, and J. Qin, "Smart Home Energy Management," *Energies (Basel)*, vol. 10, no. 3, p. 382, Mar. 2017, doi: 10.3390/en10030382.
- [2] A. Iranpour Mobarakeh, R. Sadeghi, H. Saghafi esfahani, and M. Delshad, "Techno-economic energy management of micro-grid in the presence of distributed generation sources based on demand response programs," *International Journal of Electrical Power & Energy Systems*, vol. 141, p. 108169, Oct. 2022, doi: 10.1016/j.ijepes.2022.108169.
- [3] A. Shewale, A. Mokhadde, N. Funde, and N. D. Bokde, "A Survey of Efficient Demand-Side Management Techniques for the Residential Appliance Scheduling Problem in Smart Homes," *Energies (Basel)*, vol. 15, no. 8, p. 2863, Apr. 2022, doi: 10.3390/en15082863.
- [4] A. Kahraman, O. Bulut, E. Biyik, C. Guzelis, and G. Demirkiran, "Stochastic Microgrid Control Problems: Effects of Load Distribution and Planning Horizon," in *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, IEEE, Oct. 2019, pp. 1–6. doi: 10.1109/ASYU48272.2019.8946439.
- [5] F. Agner, "Creating Electrical Load Profiles Through Time Series Clustering," 2019.
- [6] S. Yilmaz, J. Chambers, and M. K. Patel, "Comparison of clustering approaches for domestic electricity load profile characterisation - Implications for demand side management," *Energy*, vol. 180, pp. 665–677, Aug. 2019, doi: 10.1016/j.energy.2019.05.124.
- [7] K. Zhou, S. Yang, and Z. Shao, "Household monthly electricity consumption pattern mining: A fuzzy clustering-based model and a case study," *J Clean Prod*, vol. 141, pp. 900–908, Jan. 2017, doi: 10.1016/j.jclepro.2016.09.165.
- [8] G. Le Ray and P. Pinson, "Online adaptive clustering algorithm for load profiling," *Sustainable Energy, Grids and Networks*, vol. 17, Mar. 2019, doi: 10.1016/j.segan.2018.100181.
- [9] S. Lin, F. Li, E. Tian, Y. Fu, and D. Li, "Clustering load profiles for demand response applications," *IEEE Trans Smart Grid*, vol. 10, no. 2, pp. 1599–1607, Mar. 2019, doi: 10.1109/TSG.2017.2773573.
- [10] E. Mele, C. Elias, and A. Ktena, "Electricity use profiling and forecasting at microgrid level," 2018.
- [11] M. Alhussein, K. Aurangzeb, and S. I. Haider, "Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting," *IEEE Access*, vol. 8, pp. 180544–180557, 2020, doi: 10.1109/ACCESS.2020.3028281.
- [12] Y. Yang, W. Li, T. A. Gulliver, and S. Li, "Bayesian Deep Learning-Based Probabilistic Load Forecasting in Smart Grids," *IEEE Trans Industr Inform*, vol. 16, no. 7, pp. 4703–4713, Jul. 2020, doi: 10.1109/TII.2019.2942353.
- [13] C. Alzate and M. Sinn, "Improved electricity load forecasting via kernel spectral clustering of smart meters," *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 943–948, 2013, doi: 10.1109/ICDM.2013.144.
- [14] T. K. Wijaya, M. Vasirani, S. Humeau, and K. Aberer, "Cluster-based aggregate forecasting for residential electricity demand using smart meter data," in *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, Institute of Electrical and Electronics Engineers Inc., Dec. 2015, pp. 879–887. doi: 10.1109/BigData.2015.7363836.
- [15] A. Shahzadeh, A. Khosravi, and S. Nahavandi, "Improving load forecast accuracy by clustering consumers using smart meter data," in *2015 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jul. 2015, pp. 1–7. doi: 10.1109/IJCNN.2015.7280393.
- [16] S. Bandyopadhyay, T. Ganu, H. Khadilkar, and V. Arya, "Individual and aggregate electrical load forecasting: One for all and all for one," in *e-Energy 2015 - Proceedings of the 2015 ACM 6th International Conference on Future Energy Systems*, Association for Computing Machinery, Inc, Jul. 2015, pp. 121–130. doi: 10.1145/2768510.2768539.
- [17] F. Fahiman, S. M. Erfani, S. Rajasegarar, M. Palaniswami, and C. Leckie, "Improving load forecasting based on deep learning and K-shape clustering," in *Proceedings of the International Joint Conference on Neural Networks*, Institute of Electrical and Electronics Engineers Inc., Jun. 2017, pp. 4134–4141. doi: 10.1109/IJCNN.2017.7966378.
- [18] T. Jarabek, P. Laurinec, and M. Lucka, "Energy load forecast using S2S deep neural networks with k-Shape clustering," in *2017 IEEE 14th International Scientific Conference on Informatics, IEEE*, Nov. 2017, pp. 140–145. doi: 10.1109/INFORMATICS.2017.8327236.
- [19] A. Cini, S. Lukovic, and C. Alippi, "Cluster-based Aggregate Load Forecasting with Deep Neural Networks," in *2020 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jul. 2020, pp. 1–8. doi: 10.1109/IJCNN48605.2020.9207503.

- [20] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," *IEEE Trans Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019, doi: 10.1109/TSG.2018.2818167.
- [21] 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.
- [22] Australian Government, "Smart Grid Smart City (SGSC). Customer trial data," <https://data.gov.au/dataset/ds-dga-4e21dea3-9b87-4610-94c7-15a8a77907ef/details>, May 20, 2022.
- [23] O. Motlagh, A. Berry, and L. O'Neil, "Clustering of residential electricity customers using load time series," *Appl Energy*, vol. 237, pp. 11–24, Mar. 2019, doi: 10.1016/j.apenergy.2018.12.063.

Artificial Neural Networks-Based Route Selection Model for Multimodal Freight Transport Network During Global Pandemic

¹Yaşanur KAYIKCI, ²Elif CESUR

¹Sheffield Business School, Sheffield Hallam University, Sheffield, United Kingdom, yananur.kayikci@gmail.com 

²Department of Industrial Engineering, İstanbul Medeniyet University, İstanbul, Türkiye, elif.karakaya@medeniyet.edu.tr 

Abstract

The global pandemic caused major disruptions in all supply chains. Road transport has been particularly affected by the challenges posed by the COVID-19 pandemic. The selection of an efficient and effective route in multimodal freight transport networks is a crucial part of transport planning to combat the challenges and sustain supply chain continuity in the face of the global pandemic. This study introduces a novel optimal route selection model based on integrated fuzzy logic approach and artificial neural networks. The proposed model attempts to identify the optimal route from a range of feasible route options by measuring the performance of each route according to transport variables including, time, cost, and reliability. This model provides a systematic method for route selection, enabling transportation planners to make smart decisions. A case study is conducted to exhibit the proposed model's applicability to real pandemic conditions. According to the findings of the study, the proposed model can accurately and effectively identify the best route and provides transportation planners with a viable option to increase the efficiency of multimodal transport networks. In conclusion, by proposing an innovative and efficient strategy for route selection in complex transport systems, our research significantly advances the field of transportation management.

Keywords: Multimodal Transportation; Freight Transportation; Route Selection; Fuzzy Logic; Artificial Neural Network

1. INTRODUCTION

The global coronavirus outbreak, also known as COVID-19, has had an unprecedented impact on world trade, causing disruptions to global supply chains and resulting in severe financial losses and operational shutdowns in various industries [1-2]. As businesses face the rapidly evolving challenges arising from the pandemic, companies are becoming proactive in their decision-making processes to ensure business continuity and create flexible business models [3-4].

International transport networks, especially road transport, are seriously affected by this event. The COVID-19 pandemic has posed significant difficulties for transportation companies, particularly in how they manage customer relationships. Mitreğa and Choi's multi-method study explores how small and medium transportation companies navigate customer relationships during this pandemic [5]. Additionally, Vrabac et al. have proposed a discrete time SEIR model for analyzing the spread of the virus in transportation networks [6].

Border restrictions and controls imposed by countries, and local traffic restrictions applied by municipalities at city/province entrances and exits lead to long waiting times on roads or at borders, delays in shipment due to labor shortages in warehouses, terminals, and ports, shortages of trucks, truck drivers, vehicles and equipment, and delays in cargo collection and delivery, cargo handling, causing serious impacts on transport networks and operational challenges [7]. This situation forces companies to use advanced logistics infrastructure and technologies that will protect them from vulnerabilities as much as possible, ensure their competitiveness, increase their efficiency, reduce logistics costs, and utilize multimodal transportation solutions that will reduce their dependence on road transport in international freight transportation operations [2, 8].

Multimodal transport refers to carrying transport goods placed in transport units (e.g., portable containers, trailers, semi-trailers, and similar cases) using various transport means (e.g., RoRo vessels) taken from a place (origin port/departure terminal) in one country to a specified place (destination port/delivery terminal) in another country through at least two different transport modes among land, sea, inland waterways, and air under a single transport

contract or bill of lading [9-10]. This enables multimodal transport to provide an international transport network that is focused on productive activities, efficient operations, and sustainability, while including multiple transportation modes, unlike road transport, which is a single mode. In addition, multimodal transport is seen to have greater resilience and elasticity than unimodal transportation types in the event of any epidemic, disaster, or similar situation that may occur in the transport network [8, 10]. Different combinations of transport modes are possible in the organization of multimodal transport, such as rail-road, sea-road, inland waterway-railway, and sea-railway. During the transfer process between transport modes, the transport units cannot be changed, and the loads within them cannot be handled [11].

The COVID-19 pandemic has significantly impacted multimodal transportation systems as well, leading to various challenges. One such challenge is developing appropriate evaluation criteria and assessment methods to measure the resilience of transportation systems [12]. Furthermore, it is crucial to understand the impact of COVID-19 on transport turnover, as COVID-19 has caused people to increase their consumption behaviors through freight transport [13].

The transportation chain includes three levels: pre-haulage, main-haulage, and end-haulage, which are connected to transfer centers that enable changes in transportation modes [14]. These transfer centers can be sea terminals, rail terminals, road terminals, or multimodal transport centers. While road transport is used for the pre-haulage and end-haulage segments, which cover short distances within a country or region, the main-haulage segment, which covers distances greater than 300 km, often crosses different countries and even continents and is carried out by other transportation modes, especially railways and/or sea/inland waterway transportation [14-16]. Road transport is mainly used for operations in the pre-haulage and end-haulage segments, requiring much shorter transport times than the rest of the transport process [8].

Multimodal transport is organized by a multimodal transport provider (MTP) that provides door-to-door or terminal-to-terminal services and consists of a single rail or sea/inland waterway cargo operator or a consortium of multiple rail and sea cargo operators [8, 16]. Based on a cargo transport contract called the multimodal transport document, multimodal transport provides uninterrupted delivery of cargo from the seller's door or origin terminal to the buyer's door or destination terminal [17].

When considering the combination of transportation types in transport networks, there are many alternative routes with different performances. For MTPs, the most important challenge they face is determining which alternative route to choose, especially in the event of supply chain disruptions caused by the global pandemic.

Several methods have been proposed to address the route selection problem in transportation planning. Hybrid multi-criteria decision-making (MCDM) models [18], deep learning methods [19-20], ensembling methods [21], and

reinforcement learning methods [20] are some of the recent methods proposed. These approaches provide a promising direction for future research to enhance the efficiency and effectiveness of route selection in transportation systems.

In this study, a route selection model integration of fuzzy logic and an artificial neural network (ANN) approach is proposed to identify the most suitable route among the alternative multimodal freight transportation routes in a transport network. The integrated fuzzy logic approach and ANNs were chosen as a solution methodology to construct a route selection model in a transport network that includes multimodal transportation, allowing practitioners to respond instantaneously to changing business and environmental conditions and make rapid decisions about selecting the best route from the alternatives provided. In particular, MTPs can predict which routes may be more elastic and resilient based on the results of this study and offer these routes as a service to their customers. The integrated fuzzy logic and ANN model considers only sea and rail transportation, which includes operations in the main transportation segment from terminal/port to terminal/port, while road transport is excluded from the model.

The remainder of the paper is structured as follows. Firstly, a comprehensive literature review is presented, and its relevance to the research topic is provided. Then, the methodology including fuzzy logic and ANNs, used for the proposed model in the article is demonstrated. In the following section, the proposed model is applied with a case study, and its analysis is performed. Then, the results are obtained, and discussions are presented. Finally, the article ends with a conclusion section.

2. LITERATURE REVIEW

Transport networks should be designed flexibly to easily adapt to changing business and environmental conditions, as well as unexpected events such as natural disasters (earthquakes, tsunamis, floods, droughts, etc.) and various pandemics (COVID-19, SARS, MERS and H1N1) to enhance resilience [22]. Providing different multimodal freight transportation routes for the main transportation segment in the transportation chain to customers engaged in international trade is a critical decision for MTPs [9-10]. Planning a multimodal freight route allows carriers to maximize profitability, minimize shipping costs across all service lines and also negotiate a more favorable price with MTPs using appropriate pricing strategies [9, 23]. Factors such as transport cost, on-time service quality, transportation risk, scheduled transportation types and delivery times, and transportation economies of scale all exert their influence on the decision-making process. Balancing these diverse objectives adds complexity to an already intricate task. [24]. In addition, capacity constraints of transfer centers can create problems when switching from one mode of transportation to another [25-26].

Route selection is a crucial problem in transportation planning, and various methods have been proposed in the literature to tackle this issue. Qu and Chen proposed a hybrid MCDM model that combines the Fuzzy Analytic Hierarchy Process (AHP) and ANN theory [18]. They obtained a more

effective route selection process by considering multiple criteria simultaneously. On the other hand, Campigotto et al. proposed a personalized favorite route recommendation algorithm that utilized Bayesian learning techniques. This approach considered users' preferences and past behavior to provide a customized route recommendation [27]. Abduljabbar et al. proposed the use of artificial intelligence (AI) methods as a smart solution to handle complex transportation systems that cannot be managed by traditional methods. They proposed the use of deep learning methods to learn and optimize the routing decisions in real-time [28]. Additionally, Abedalla et al. proposed a weighted average ensembling method of Convolutional Neural Network and Gradient-Boosted Decision Trees for route selection. The results showed promising accuracy rates for the proposed method [21]. Chen and Chang addressed the route selection problem using a semi-supervised learning method [29]. This approach utilized prior training data to identify the best route while considering the trade-off between accuracy and computation efficiency. Finally, Koohathongsumrit and Chankham proposed a route selection model in multimodal supply chains by integrating fuzzy risk assessment model (FRAM), best-worst method (BWM), and measurement of alternatives and ranking according to the compromise solution (MARCOS) [30].

Although many existing studies have already modeled the multimodal transportation route selection problems [24, 31], there is currently no study that includes an approach to route selection problems for multimodal transportation services within the framework of the literature review. Moreover, there has been a dearth of studies that propose a solution methodology by combining the fuzzy logic approach and Artificial Neural Networks (ANNs). This innovative fusion remains unexplored in the existing literature. Therefore, this study will represent a first in this sense. ANNs are practically employed in solving learning, generalization, specification, identification, classification, association, and optimization problems. With this method, information obtained from the samples is recorded in networks, essentially networks are trained with the data, over time the networks become better at making decisions, recognizing patterns, and producing reliable results in similar situations [32] and also during the COVID-19 pandemic [33]. The reason for using this integrated methodology is its applicability to redirecting alternative multimodal transportation routes in response to changes or fluctuations in fare rates, contracts and weather conditions in freight cases. The utilization of fuzzy logic theory facilitates a comprehensive comparison of alternative routes based on multiple targets. This approach ensures the maintenance of transport service quality at the desired level, while simultaneously minimizing transport time and cost. By employing fuzzy logic, the model selects the most suitable route that optimally meets the specified criteria. Route selection also takes into account the scheduling of different transport modes and capacities. For each identified route, the proposed model employs data on time, cost, and dependability of transport as input parameters. The route performance is calculated as the output parameter following the use of the fuzzy logic technique. Then, the data set is used as training data for ANNs together with the obtained route performance values. Subsequently, the performance predictions are made based on this training data by applying

ANNs to a data set that includes much larger multimodal transportation routes. Practitioners can make informed decisions regarding the selection of a route by evaluating the calculated performance. A greater level of route efficiency suggests that MTPs are more likely to provide this particular multimodal transport path to their clients, and the senders.

3. RESEARCH METHODOLOGY

In this study, a model is proposed that integrates the fuzzy logic method and ANN approach. In the model, firstly, the performance of the routes is calculated using fuzzy logic algorithms based on the dataset of existing routes on the transport network. Then, together with the obtained performance values, this dataset is used as the training data for the ANN approach to train the transport network and calculate the performance values of the dataset containing alternative routes. The flowchart showing the proposed model is given in Figure 1.

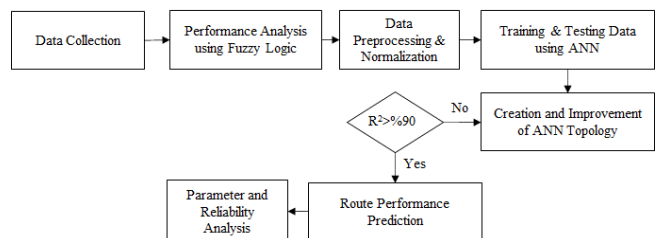


Figure 1. Model flowchart integrated with fuzzy logic approach and ANNs.

3.1. Fuzzy Logic Approach

Recently, an increasing number of research studies have used the fuzzy logic approach to apply the definition of human behavior in social environments, customer requirements in dynamic markets, logistics, and various other fields. As a well-known approach, the fuzzy logic method is attracting growing attention among both academicians and practitioners alike, as it can work with both numerical and non-numerical characters and generate classes to make better decisions [34].

The challenges presented by the COVID-19 pandemic have necessitated the development of new models to evaluate transportation-related risks. One such model, proposed by Tang et al., integrates an extended type-2 fuzzy model with a Bayesian network to evaluate COVID-19 medical waste transportation risk [35]. To effectively manage public transportation during the pandemic, Devenci et al. proposed a fuzzy Einstein model to evaluate and rank prioritization [36]. Özden & Celik sought to identify critical service quality priorities in cargo transportation before and during the COVID-19 outbreak using the Fuzzy Importance-Performance-Impact Analysis method [37].

The fuzzy logic approach is used in this work to create a rule-driven model to predict route performance by considering some basic characteristics of routes on the transport network, including transport cost, transport time, and transport reliability.

Within the literature, numerous fuzzy models are categorized into three primary classifications: 1) the Mamdani model, also referred to as the Max-Min method [38], 2) the Takagi-Sugeno model [39], and 3) Kosko's model [34]. For the purposes of this study, the widely recognized Mamdani model was adopted, wherein the following rules were repeatedly applied:

- Firstly, fuzzification is performed, where all input variables are transformed into fuzzy variables.
- Then, the output detection step applies the pre-defined fuzzy rules to the input values.
- The output accumulation stage, which totals up all fuzzy rules to finish the output, is crucial.
- Finally, defuzzification ties the fuzzy output with its original form.

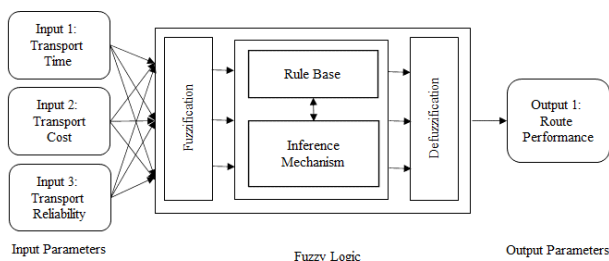


Figure 2. Fundamentals of Fuzzy Logic.

The fundamental principles of fuzzy logic are illustrated in Figure 2. The fuzzy set theory employs membership functions $\mu_a(a)$, $\in [0,1]$ which assigns a degree of membership to each object on a given scale [40]. Various fuzzy membership functions have been used, including triangular, Gaussian, and trapezoidal membership functions [11]. Figure 3 shows the triangular fuzzy numbers $\mu_a(a)$ and trapezoidal fuzzy numbers $\mu_b(b)$. For this study, triangular membership functions are adopted as they are the most widely used membership functions [41]. A triangular fuzzy number a can be defined by a triplet $(a1, a2, a3)$; $a1 \leq a2 \leq a3$, and the membership function $\mu_a(a)$ is presented in Equation (1):

$$\mu_a(a) = \begin{cases} (a - a1)/(a2 - a1) & a1 \leq a \leq a2 \\ (a - a3)/(a3 - a2) & a2 \leq a \leq a3 \\ 0 & a < a1 \text{ or } a > a3 \\ 1 & a = a2 \end{cases} \quad (1)$$

The fuzzy approach is a method that employs pre-established rules composed of a series of fuzzy "if-then" statements, gathered in the form of linguistic expressions through expert opinions obtained through questionnaires or Delphi sessions. Each rule is understood as an "implication" and comprises an "antecedent" (the "if" part) and a "consequent" (the "then" part). The fuzzy rule's general format is presented in Equation (2):

$$\text{If it is } A, \text{ Then is } B \quad (2)$$

where A and B are, respectively, linguistic concepts in the language universe that are specified by sets of fuzzy concepts.

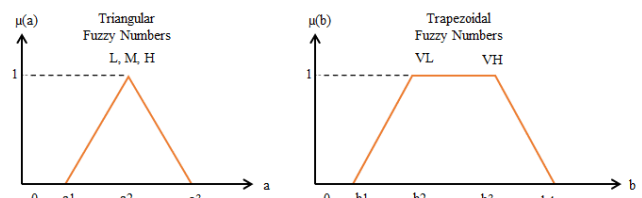


Figure 3. Triangular and trapezoidal fuzzy numbers.

3.2. Artificial Neural Network

The ANNs have been developed as generalizations of mathematical models of biological nervous systems. An ANN is a network consisting of collections of very simple processors (neurons) each of which has (a small amount of) local memory, and which operates only on their local data and receives inputs via connections, or links, that are typically unidirectional [42]. Each unit in the network has a rule for computing its output signal by collecting incoming signals and calculating an output signal that is then sent to other units in the network. The rule for computing the output is known as an activation function [43]. The structure of a neural network consists of three layers. The first layer is the input layer that directly interacts with the external world. The second layer is the hidden layer where the computation is performed according to the provided function. The last layer is the output layer where we obtain the output. Information in neural networks is stored as synaptic weights between neurons. The network propagates the input data layer by layer until the output data is generated. If the networks, such as multi-layer perceptrons, with a backpropagation algorithm, have outputs different from the desired output, an error is calculated. Then the error propagates back through the network, and the synaptic weights are adjusted as the error propagates [44]. Generalization is the unique ability that makes ANNs such a powerful tool. In general, neural networks (represented as mathematical models) are a collection of simple computing units connected by a connectivity system. Figure 4 illustrates the fundamental principles of ANN.

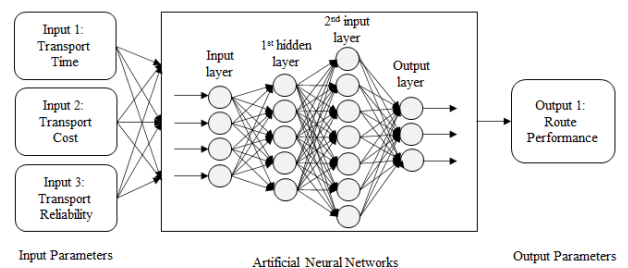


Figure 4. Basic Principles of ANN.

4. CASE STUDY

4.1. Case Presentation and Data Collection

An international multimodal freight transport company in Istanbul specializes in offering comprehensive services for sea and rail transportation using a variety of transport means (e.g., RoRo ships, RoLa trains, ISUs) to provide frequent services from Turkey to Europe, particularly using semi-trailers as transport units for daily shipments. During a global pandemic, road transport between Turkey and Germany was

seriously affected. In particular, long queues built up at border crossings and continued throughout the border closure. On the other hand, there was an increase in demand for multimodal services between Istanbul and Hamburg. This chaotic situation led to taking some measures. The company decided to offer a new service route that is suitable for potential customers (freight forwarders) between the origin-destination (O-D) points. The proposed new service route entails operating from a loading terminal in Istanbul to a unloading terminal in Hamburg. These terminals, which can be either port terminals or inland rail terminals, serve as O-D nodes respectively. Additionally, there may exist multiple transfer terminals, known as multimodal centers,

where modal shifts occur. These centers enable the transition between transportation modes, whether it involves changing within the same modes (e.g., sea-sea, road-road) or between diverse modes (e.g., road-rail, rail-sea), along the route between the O-D points. The new service route should be innovative enough to reduce transport time and cost, increase transport reliability, and provide high transport route performance. In making the decision, the company assigned three expert interns who searched for all possible routes between the O-D points and identified 17 various routes where any given transportation unit (cargo) is transferred once or twice (modal shift).

Table 1. The list of alternative routes

#	OD	Node (n1)	Leg (l1)	Node (n2)	Leg (l2)	Node (n3)	Leg (l3)	Node (n4)	Time (day)	Cost (€)	Reliability (%)
1	Pendik-Kiel	Pendik	Sea	Trieste	Rail	Kiel	n/n	n/n	4	3298	90
2	Pendik-Hamburg	Pendik	Sea	Trieste	Rail	Vienna	Rail	Hamburg	6	2958	86,7
3	Pendik-Hamburg	Pendik	Sea	Trieste	Rail	Linz	Rail	Hamburg	6	3078	86,7
4	Istanbul-Hamburg	Istanbul	Rail	Köln	Rail	Hamburg	n/n	n/n	6	2750	75
5	Pendik-Hamburg	Pendik	Sea	Trieste	Rail	Kiel	Rail	Hamburg	6	3448	85,0
6	Pendik-Hamburg	Pendik	Sea	Trieste	Rail	Köln	Rail	Hamburg	6	3298	83,3
7	Pendik-Hamburg	Pendik	Sea	Trieste	Rail	Salzburg	Rail	Hamburg	7	2898	86,7
8	Pendik-Hamburg	Pendik	Sea	Trieste	Rail	Duisburg	Rail	Hamburg	7	3398	80,0
9	Pendik-Hamburg	Pendik	Sea	Trieste	Rail	Milano	Rail	Hamburg	7	2561	86,7
10	Istanbul-Kiel	Istanbul	Sea	Trieste	Rail	Kiel	n/n	n/n	5	3213	87,5
11	Istanbul-Hamburg	Istanbul	Sea	Trieste	Rail	Linz	Rail	Hamburg	6	2993	85,0
12	Istanbul-Hamburg	Istanbul	Sea	Trieste	Rail	Salzburg	Rail	Hamburg	6	2813	85,0
13	Istanbul-Hamburg	Istanbul	Sea	Trieste	Rail	Vienna	Rail	Hamburg	6	3113	85,0
14	Istanbul-Hamburg	Istanbul	Sea	Trieste	Rail	Duisburg	Rail	Hamburg	6	3313	78,3
15	Istanbul-Hamburg	Istanbul	Sea	Trieste	Rail	Ludwigs-hafen	Rail	Hamburg	6	2913	85,0
16	Istanbul-Hamburg	Istanbul	Sea	Trieste	Rail	Milano	Rail	Hamburg	6	2476	85,0
17	Istanbul-Hamburg	Istanbul	Sea	Trieste	Rail	Budapest	Rail	Hamburg	7	2613	78,3

In this paper, the dataset encompasses sea and rail cargo alternative paths operated by an MTP between different terminals or ports. These routes may involve multiple steps and nodes, including transfer stations, within the network. The design of the overall multimodal freight route can incorporate both sea and rail transport modes or utilize either of them individually. Table 1 presents a compilation of 17 existing multimodal routes from Istanbul to Hamburg, specifying the presence of one or two transfers. Among these routes, #1 from Pendik to Kiel, #10 from Istanbul to Kiel, and #4 from Istanbul to Hamburg involve one transfer each. The transfer terminals, such as Trieste and Cologne, facilitate the shift between transport modes, be it from sea to rail or vice versa, or from rail to rail. The remaining routes involve two transfers.

Since performance continuity cannot be easily evaluated by only considering some parameters, the achievement of performance usability for all multimodal transportation systems is a significant concern. However, accurate performance usability values for all systems cannot be guaranteed by using simple equations. In addition, it is quite difficult to express the correlation between control parameters and performance usability with a mathematical function. Therefore, performance usability can be formulated using fuzzy logic rules [45].

4.2. Application of Fuzzy Logic

In this study, the fuzzy logic model receives input parameters, it calculates the corresponding output parameters to facilitate the route selection decision-making process. The input parameters are described in such a way:

Transport time: This parameter signifies the total number of travel days required for transportation between O-D points.

Transport cost: This parameter encompasses the overall costs associated with rail and/or sea transportation, including expenses for port/terminal operations, loading/unloading and customs procedures.

Transport reliability: This parameter reflects the percentage of transportation services that reach their destinations within the specified time (on-time arrival). Notably, railway services often face challenges in completing their journeys without delays due to congestion in the railway network and limited operational flexibility. Leading railways employ digital technologies to optimize tariffs and control the network, thereby increasing reliability even with increased train frequency, aiming to overcome operational challenges.

The output parameter, route performance, holds significant importance in ensuring the timely provision and evaluation of multimodal services.

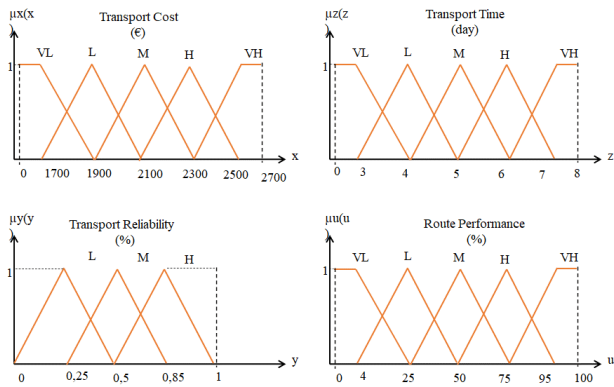


Figure 5. The fuzzy sets for output and input variables

The fuzzy set of input and output parameters is given in Figure 5, where membership functions ($\mu_x(x)$, $\mu_z(z)$, $\mu_y(y)$, $\mu_u(u)$) are created by determining five linguistic variables for transport cost, transport time, and route performance, namely; Very Low (VL), Low (L), Medium (M), High (H), and Very High (VH). However, the transport reliability parameter is characterized by three linguistic variables: Low (L), Medium (M), and High (H). The determination of membership functions and the maximum/minimum inference range is based on expert opinions. For instance, the transport cost ranges from €2300 to €3800, while the transport time spans from 3 to 8 days with an increment of 1. On the other hand, both transport reliability and route performance range from 0 to 100, divided into 25 intervals. As a result of the maximum/minimum inference, the membership functions exhibit overlapping pairs.

For this research, 75 fuzzy "if-then" rules are generated based on expert opinions. An example of such a rule #34 can be given as follows:

IF Transport Cost = Medium (M), and
 Transport Time = High (H), and
 Transport Reliability = Low (L)
 THEN Route Performance = Very Low (VL)

The Appendix contains a comprehensive inventory of fuzzy "if-then" rules, in which cost, reliability, and time form the "if part," while performance constitutes the "then part" of the fuzzy rules.

4.3. Artificial Neural Network Topology

The most important factor in predicting route performance with ANNs is to construct the neural network architecture correctly. First, we divide our data into training and test data to make predictions. The current data set is divided into 75% training data and 25% test data. ANNs try to extract models by observing existing patterns in the training data. However, encountering two problems during this process is very common: 1) Overfitting or 2) Underfitting. In this case, the

model will not work correctly and the error rate between actual and predicted values will be high.

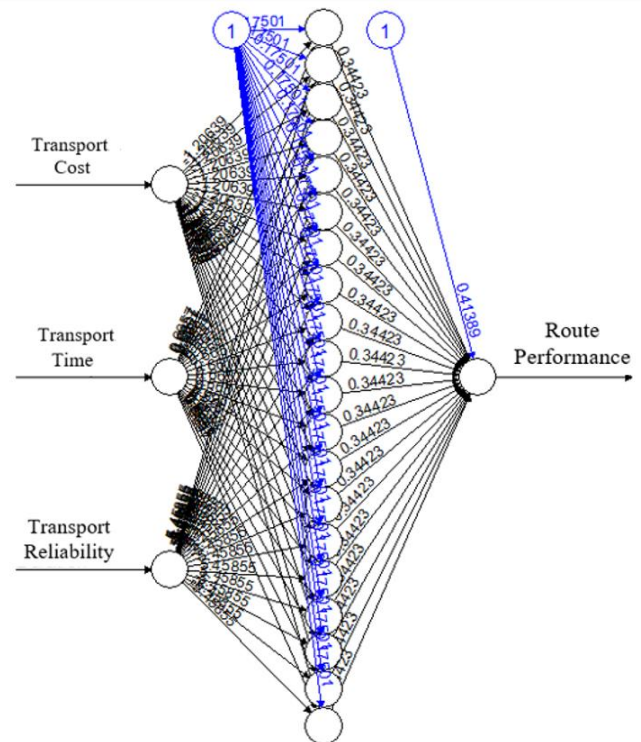


Figure 6. Structure of ANN.

In the scope of the case study, performance predictions were made for a data set containing 273 different route alternatives based on the performance values obtained from the Fuzzy model, which is the first step of the study. Figure 6 shows the ANN design. It can be seen that there are three evaluation criteria (input) and one outcome parameter (output). When the neural network topology is examined in detail, 20 hidden layers are used, and "logistic sigmoid" is chosen as the activation function. The logistic sigmoid with an S-shaped structure is a good classifier with the best-known neural network activation function and prevents some generated values from being lost as it produces values between [0,1]. A very small error rate of "0.00005" was determined as the threshold value. A fixed number of "0.5" was assigned as the initial input weight. The "feed-forward" method was used as the ANN algorithm. The sum of squared errors (SSE) was selected as the error function parameter. Since the activation function for the linear output parameter was determined as the logistic sigmoid, the value "false" was assigned.

5. RESULT AND DISCUSSION

In the first step of the study, the performance results calculated using the fuzzy logic approach are given in Figure 7. Three out of the 17 multimodal cargo routes accomplished higher performance according to the results: Pendik-Kiel with 89%, Istanbul-Hamburg with 83%, and Istanbul-Kiel with 82% route performance, respectively. Semi-trailers between Istanbul and Kiel made one transfer while cargos between Istanbul and Hamburg made two transfers. Therefore, it is possible to say that the number of transfers also has an impact on route performance. Delays and cancellations may occur for various reasons (such as

infrastructure deficiencies, capacity problems, etc.) However, container transportation, which standardizes dimensions and properties, can decrease multimodal transfer times, and raise cost and fuel efficiency. In addition, automatic transfer management systems such as ISU, Megawing, Flexiwaggon, ModaLohr and so on can facilitate transfer operations and optimize efficiency on platforms, which can also increase route performance. MTP could determine one of the top three routes with the highest performance and if there is any interruption in the chosen route, the others can be considered as alternative routes. As a result of this case study, a novel multimodal cargo transportation route between Istanbul and Kiel has been selected as a cargo route, and the additional two routes between Istanbul and Hamburg will be considered as alternative route opportunities.

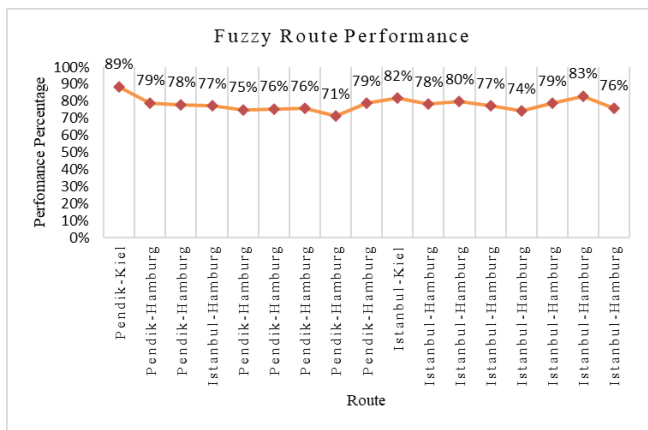


Figure 7. Evaluation of the performance of different route options

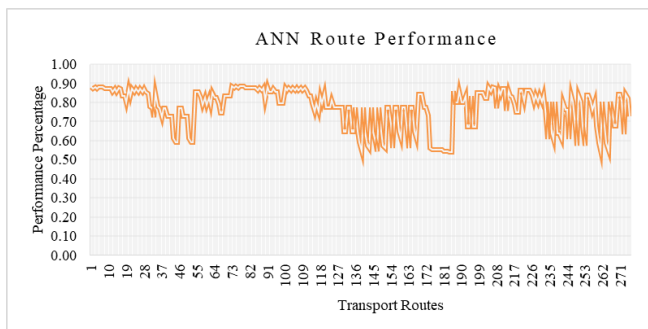


Figure 8. Route performance using ANN.



Figure 9. Clustering analysis using ANNs.

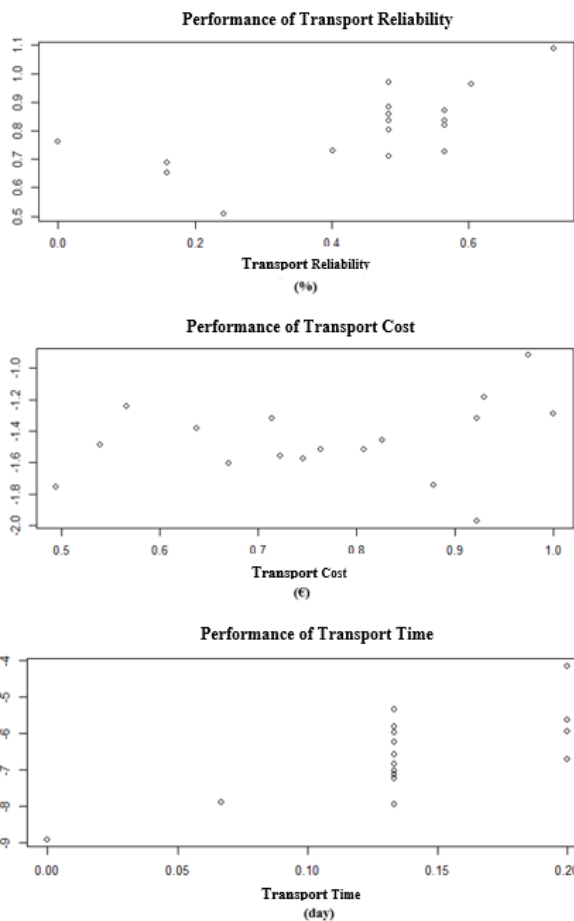


Figure 10. Parameter analysis as a result of ANN.

As a second stage of the case study, the results obtained from fuzzy logic were used to train ANNs to solve the pattern between route safety, transport time, and cost parameters. In this stage, the ANN method is one of the best methods that can be used when there is not enough information about the relationship between input values, and when the positive and negative effects of these values are considered to predict the output value. The method can make predictions about route performance on the test data set using the patterns it obtained in its training data. In this study, the success of the ANN model is shown to have a correlation coefficient of 96%, indicating that the model performs well. Figure 8 illustrates that the route performances in the new data set for which predictions are being requested exhibit a significant amount of fluctuation.

In Figure 9, a clustering was performed to provide more detailed information about the predictions made, and it was observed that the common feature exhibited by these alternatives was that the travel time was very long. This directly affected the performance of the route and caused it to be at very low levels.

The relationship between the evaluation criteria: transport cost, route safety, transport time, and route performance, are summarized in the following graphs in the parameter analysis conducted through the ANN method. When we look at the relationship between transport cost and performance, we can see that the effect of cost fluctuates around the 0-axis. It is known that cost is an important parameter in determining route performance, but it is not a very

significant input in distinguishing. It has only shown an effect between 1.0-2.0. The main determinant is seen to be the effect of transport time. Transport time has become the most distinctive parameter (4.0-9.0) and decreases in route performance are observed when each day increases. The effect of route reliability has only been effective in the range of 0.5 - 1.1. In Figure 10, it can be seen how much each input value affects the route performance result.

6. CONCLUSION

The COVID-19 epidemic has had a significant impact on transportation infrastructure across the world. It is obvious that the pandemic has accelerated current trends and generated new problems for the sector, even though the full amount of the impact is still unknown. Transportation networks will need to be redesigned as the globe adjusts to the pandemic to make sure they can meet modifying needs in society. Especially due to the lockdowns, travel restrictions, and social distancing measures implemented by the government, finding, and evaluating alternative routes has become critically important. At this point, this study proposes a novel approach for selecting the most suitable route among alternative multimodal freight transport routes using an ANN integrated with a fuzzy logic model. By considering transport time, cost, and reliability data, the proposed model predicts which routes are more elastic and resistant to vulnerabilities. As a result of the proposed ANN integrated with a fuzzy logic model, it was determined that the Pendik and Kiel route is the most suitable route. When looking at the characteristics of this route, even though transportation cost is high, it is seen that it is selected as the best route since transportation reliability and transportation time values are low. Considering the transportation under COVID-19, the alternative route of Pendik-Kiel is more reliable in terms of COVID-19 transmission due to the short distance, and it is a fast route against lockdown problems, which is very significant.

This allows MTPs to offer more reliable and efficient services to their customers. The model was trained using a dataset that includes route performance estimates as output parameters and performance predictions were made for a larger dataset of multimodal transportation routes. Overall, this approach can improve the decision-making process in the selection of optimal routes for multimodal freight transportation services.

Future research can expand on the proposed route selection model by incorporating additional parameters like road, construction, accident situations transportation volume, and distance, transportation risk. Furthermore, including road transport in the model can aid in determining the location of inland terminals for sea and rail transportation based on O-D flows analyzed through data analytics. The performance of the model can also be compared to other machine learning algorithms such as random forest and support vector machine methods. These potential advancements can further improve the accuracy and efficiency of the route selection process in multimodal freight transportation services.

ACKNOWLEDGEMENT

The first author, Dr. Yasanur Kayikci, gratefully acknowledges the support provided by the Scientific and Technological Research Council of Turkey under grant number 116C048 for this research.

Author contributions: Concept – Y.K., E.C.; Data Collection & Processing – Y.K, E.C.; Literature Search – Y.K., E.C.; Writing – Y.K., E.C.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: Scientific and Technological Research Council of Turkey, Grant/Award Number: 116C048.

REFERENCES

- [1] T. M. Choi, "Innovative 'Bring-Service-Near-Your-Home' operations under Corona-Virus (COVID-19/SARS-CoV-2) outbreak: Can logistics become the Messiah?," *Transp. Res. E Logist. Transp. Rev.*, vol. 140, 2020, doi: 10.1016/j.tre.2020.101961.
- [2] R. Kiani Mavi, N. Kiani Mavi, D. Olaru, S. Biermann, and S. Chi, "Innovations in freight transport: a systematic literature evaluation and COVID implications," *Int. J. Logist. Manag.*, vol. 33, no. 4, 2022, doi: 10.1108/IJLM-07-2021-0360.
- [3] D. Ivanov, "Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case," *Transp. Res. E Logist. Transp. Rev.*, vol. 136, 2020, doi: 10.1016/j.tre.2020.101922.
- [4] D. Ivanov, "Viable supply chain model: integrating agility, resilience and sustainability perspectives—lessons from and thinking beyond the COVID-19 pandemic," *Ann. Oper. Res.*, vol. 319, no. 1, 2022, doi: 10.1007/s10479-020-03640-6.
- [5] M. Mitreğa and T. M. Choi, "How small-and-medium transportation companies handle asymmetric customer relationships under COVID-19 pandemic: A multi-method study," *Transp. Res. E Logist. Transp. Rev.*, vol. 148, Apr. 2021, doi: 10.1016/j.tre.2021.102249.
- [6] D. Vrabac, M. Shang, B. Butler, J. Pham, R. Stern, and P. E. Pare, "Capturing the Effects of Transportation on the Spread of COVID-19 with a Multi-Networked SEIR Model," *IEEE Control. Syst. Lett.*, vol. 6, pp. 103–108, 2022, doi: 10.1109/LCSYS.2021.3050954.
- [7] A. Bombelli, "Integrators' global networks: A topology analysis with insights into the effect of the COVID-19 pandemic," *J. Transp. Geogr.*, vol. 87, 2020, doi: 10.1016/j.jtrangeo.2020.102815.
- [8] A. Mutlu, Y. Kayikci, and B. Çatay, "Planning Multimodal Freight Transport Operations: a Literature Review," *22nd ISL 2017 Ljubljana*, no. July, 2017.

- [9] Y. Kayıkcı, A. Mutlu, and B. Çatay, "A Decision Support System and its Application for International Sea-Rail Multimodal Freight Transportation Operations," *Uludağ Uni. J. Faculty Eng.*, vol. 23, no. 4, pp. 99–116, Dec. 2018, doi: 10.17482/uumfd.455356.
- [10] Y. Kayıkcı, "Analysis of Cost Allocation Methods in International Sea-Rail Multimodal Freight Transportation," *J. Yaşar Uni.*, vol. 15, no. 57, 2020, doi: 10.19168/jyasar.568692.
- [11] Y. Kayıkcı and V. Stix, "Causal mechanism in transport collaboration," *Expert Syst Appl.*, vol. 41, no. 4, 2014, doi: 10.1016/j.eswa.2013.08.053.
- [12] A. Liu, Z. Li, W. L. Shang, and W. Ochieng, "Performance evaluation model of transportation infrastructure: Perspective of COVID-19," *Transp. Res. Part. A Policy Pract.*, vol. 170, 2023, doi: 10.1016/j.tra.2023.103605.
- [13] S. J. Ho, W. Xing, W. Wu, and C. C. Lee, "The impact of COVID-19 on freight transport: Evidence from China," *MethodsX*, vol. 8, Jan. 2021, doi: 10.1016/j.mex.2020.101200.
- [14] Y. Kayıkcı and B. Çatay, "Revenue-based slot allocation and pricing framework for multimodal transport networks," 22nd ISL 2017 Ljubljana, no. July, 2017.
- [15] L. A. Tavasszy, "Modal Shift Target for Freight Transport Above 300 km: An Assessment," 2011. https://www.acea.auto/files/SAG_17.pdf
- [16] Ulaştırma Denizcilik ve Haberleşme Bakanlığı, *Türkiye Kombine Taşımacılık Strateji Belgesi*. 2014.
- [17] UN, "United Nations Convention On International Multimodal Transport Of Goods," *Unif. Law Rev.*, vol. os-8, no. 2, 1980, doi: 10.1093/ulr/os-8.2.157.
- [18] L. Qu and Y. Chen, "A hybrid MCDM method for route selection of multimodal transportation network," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008. doi: 10.1007/978-3-540-87732-5_42.
- [19] Z. Lv, Y. Li, H. Feng, and H. Lv, "Deep Learning for Security in Digital Twins of Cooperative Intelligent Transportation Systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, 2022, doi: 10.1109/TITS.2021.3113779.
- [20] Z. Zhao and Y. Liang, "A deep inverse reinforcement learning approach to route choice modeling with context-dependent rewards," *Transp. Res. Part C Emerg. Technol.*, vol. 149, 2023, doi: 10.1016/j.trc.2023.104079.
- [21] A. Abedalla et al., "MTRecS-DLT: Multi-Modal Transport Recommender System using Deep Learning and Tree Models," in *2019 6th International Conference on Social Networks Analysis, Management and Security, SNAMS 2019*, 2019. doi: 10.1109/SNAMS.2019.8931864.
- [22] A. Liu, Z. Li, W.-L. Shang and W. Ochieng, "Performance Evaluation Model of Transportation Infrastructure: Perspective of COVID-19," *Transp. Res. Part A Policy Pract.*, vol. 170, 1036, 2023, doi: 10.1016/j.tra.2023.103605
- [23] C. Barnhart and H. D. Ratliff, "Modeling Intermodal Routing," *J. Bus. Logist.*, vol. 14, no. 1, 1993.
- [24] T. S. Chang, "Best routes selection in international intermodal networks," *Comput. Oper. Res.*, vol. 35, no. 9, 2008, doi: 10.1016/j.cor.2006.12.025.
- [25] H. Min, "International intermodal choices via chance-constrained goal programming," *Transp. Res. Part A Policy Pract.*, vol. 25, no. 6, 1991, doi: 10.1016/0191-2607(91)90013-G.
- [26] J. Woxenius, "Generic framework for transport network designs: Applications and treatment in intermodal freight transport literature," *Transp. Rev.*, vol. 27, no. 6, 2007, doi: 10.1080/01441640701358796.
- [27] P. Campigotto, C. Rudloff, M. Leodolter, and D. Bauer, "Personalized and Situation-Aware Multimodal Route Recommendations: The FAVOUR Algorithm," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, 2017, doi: 10.1109/TITS.2016.2565643.
- [28] R. Abduljabbar, H. Dia, S. Liyanage, and S. A. Bagloee, "Applications of artificial intelligence in transport: An overview," *Sustainability*, vol. 11, no. 1. 2019. doi: 10.3390/su11010189.
- [29] J. I.-Z. Chen and J.-T. Chang, "Route Choice Behaviour Modeling using IoT Integrated Artificial Intelligence," *J. Artif. Intell. Capsul. Netw.*, vol. 2, no. 4, 2021, doi: 10.36548/jaicn.2020.4.006.
- [30] N., Koohathongsumrit and W. Chankham, "Route Selection in Multimodal Supply Chains: A Fuzzy Risk Assessment Model-BWM-MARCOS Framework," *Appl. Soft Comput.*, vol. 137, 110167, 2023, doi: 10.1016/j.asoc.2023.110167
- [31] A. Kengpol, W. Meethom, and M. Tuominen, "The development of a decision support system in multimodal transportation routing within Greater Mekong sub-region countries," *Int. J. Prod. Econ.*, 2012. doi: 10.1016/j.ijpe.2011.02.024.
- [32] H. Altınkaya and M. Yılmaz, "Estimation of the Daily Production Levels of a Run-of-River Hydropower Plant Using the Artificial Neural Network," *Acad. Platf. J. Eng. and Smart Syst.*, vol. 11, no. 2, pp. 62-72, 2023, doi:10.21541/apjess.1223119
- [33] E. Yıldız, M. M. Kelek , F. O. Hocaoğlu and Y. Oğuz , "Forecasting The Impact of Vaccination on Daily Cases in Turkey for Covid-19," *Acad. Platf. J. Eng. and Smart Syst.*, vol. 11, no. 1, pp. 19-26, 2023, doi:10.21541/apjess.1137177
- [34] E. Sabeur and G. Denis, "Human behavior and social network simulation: Fuzzy sets/logic and agents-based approach," in *Agent Directed Simulation Symposium, ADS 2007 - Proceedings of the 2007 Spring Simulation Multiconference, SpringSim 2007*, 2007.

- [35] J. Tang, X. Liu, and W. Wang, "COVID-19 medical waste transportation risk evaluation integrating type-2 fuzzy total interpretive structural modeling and Bayesian network," *Expert. Syst. Appl.*, vol. 213, Mar. 2023, doi: 10.1016/j.eswa.2022.118885.
- [36] M. Deveci, D. Pamucar, I. Gokasar, D. Delen, and L. Martínez, "A fuzzy Einstein-based decision support system for public transportation management at times of pandemic," *Knowl. Based Syst.*, vol. 252, Sep. 2022, doi: 10.1016/j.knsys.2022.109414.
- [37] A. T. Özden and E. Celik, "Analyzing the service quality priorities in cargo transportation before and during the Covid-19 outbreak," *Transp. Policy (Oxf)*, vol. 108, pp. 34–46, Jul. 2021, doi: 10.1016/j.tranpol.2021.04.025.
- [38] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *Int. J. Man. Mach. Stud.*, vol. 7, no. 1, 1975, doi: 10.1016/S0020-7373(75)80002-2.
- [39] T. Takagi and M. Sugeno, "Fuzzy Identification of Systems and Its Applications to Modeling and Control," *IEEE Trans Syst Man Cybern*, vol. SMC-15, no. 1, 1985, doi: 10.1109/TSMC.1985.6313399.
- [40] Saaty T, *The Analytic Hierarchy Process*, vol. 45, no. 2–3. New York: McGraw-Hill., 1980. doi: 10.1016/0377-2217(90)90209-t.
- [41] M. Becher, "Simultaneous capacity and price control based on fuzzy controllers," *Int. J. Prod. Econ.*, vol. 121, no. 2, 2009, doi: 10.1016/j.ijpe.2006.09.014.
- [42] P. H. Sydenham and R. Thorn, *Handbook of Measuring System Design*. Wiley, 2005.
- [43] Z. H. Zhou, "Rule extraction: Using neural networks or for neural networks?," *J. Comput. Sci. Technol.*, vol. 19, no. 2, 2004, doi: 10.1007/BF02944803.
- [44] Y. Chen and J. Li, "Recurrent Neural Networks algorithms and applications," in *Proceedings - 2021 2nd International Conference on Big Data and Artificial Intelligence and Software Engineering, ICBASE 2021*, 2021. doi: 10.1109/ICBASE53849.2021.00015.
- [45] A. Mahmoudi, H. Shavandi, and K. Nouhi, "Analysing Price, Quality and Lead Time Decisions with the Hybrid Solution Method of Fuzzy Logic and Genetic Algorithm," *J. Optim. Ind. Eng.*, vol. 10, no. 1–9, 2012, [Online]. Available: www.SID.ir

APPENDIX


Table 2. Fuzzy rule based on the experts' opinions


Rule Number #	<i>If part</i>			<i>Then part</i>	Rule Number #	<i>If part</i>			<i>Then part</i>
	Cost (€)	Reliability (%)	Time (day)	Performance (%)		Cost (€)	Reliability (%)	Time (day)	Performance (%)
1	VL	L	VL	M	39	M	M	H	L
2	VL	L	L	M	40	M	M	VH	L
3	VL	L	M	L	41	M	H	VL	VH
4	VL	L	H	L	42	M	H	L	VH
5	VL	L	VH	VL	43	M	H	M	H
6	VL	M	VL	VH	44	M	H	H	M
7	VL	M	L	H	45	M	H	VH	L
8	VL	M	M	M	46	H	L	VL	L
9	VL	M	H	L	47	H	L	L	L
10	VL	M	VH	L	48	H	L	M	VL
11	VL	H	VL	VH	49	H	L	H	VL
12	VL	H	L	H	50	H	L	VH	VL
13	VL	H	M	H	51	H	M	VL	M
14	VL	H	H	H	52	H	M	L	M
15	VL	H	VH	M	53	H	M	M	L
16	L	L	VL	M	54	H	M	H	VL
17	L	L	L	M	55	H	M	VH	VL
18	L	L	M	L	56	H	H	VL	H
19	L	L	H	L	57	H	H	L	M
20	L	L	VH	L	58	H	H	M	M
21	L	M	VL	H	59	H	H	H	L
22	L	M	L	H	60	H	H	VH	VL
23	L	M	M	M	61	VH	L	VL	L
24	L	M	H	M	62	VH	L	L	L
25	L	M	VH	L	63	VH	L	M	VL
26	L	H	VL	VH	64	VH	L	H	VL
27	L	H	L	H	65	VH	L	VH	VL
28	L	H	M	H	66	VH	M	VL	L
29	L	H	H	H	67	VH	M	L	L
30	L	H	VH	M	68	VH	M	M	L
31	M	L	VL	M	69	VH	M	H	VL
32	M	L	L	L	70	VH	M	VH	VL
33	M	L	M	L	71	VH	H	VL	M
34	M	L	H	VL	72	VH	H	L	L
35	M	L	VH	VL	73	VH	H	M	L
36	M	M	VL	M	74	VH	H	H	VL
37	M	M	L	M	75	VH	H	VH	VL
38	M	M	M	M					

Legend: Very Low (VL), Low (L), Medium (M), High (H), Very High (VH)

An Explorative Analysis of Tweets Sentiments for Investment Decision in Stock Markets

¹Emine ATEŞ, ^{*2}Aysun GÜRAN

¹Department of Computer Engineering, Dogus University, İstanbul, Türkiye, emine.ates.34@gmail.com 

²Department of Computer Engineering, Dogus University, İstanbul, Türkiye, adogrusoz@dogus.edu.tr 

Abstract

Nowadays lots of researches report that positive or negative social media posts have significant effects on stock market returns. In particular, as well as long-term effects, it may be important to identify the short-term effects of public opinions on the stock market. Unfortunately, the number of studies conducted in this context for Turkey is very limited. Therefore, this study investigates the relationship between Turkish tweets and the trend of Bist30 index returns for short-term analysis. Correlation analysis is a statistical method used to measure the strength and direction of relationship between two variables. The existence of a short-term correlation between sentiments of financial tweets and the stock market returns may be an indicator that social media can be used as a resource to predict the sudden changes of stock markets returns. In this study, the correlation analysis steps are conducted on the polarity scores and changes in Bist30 index returns with all the necessary statistical tests. The contribution of our study is that analyzes are made on a daily, weekly and monthly basis. The experimental results show that there are significant positive correlations between the sentiment polarity values and changes in Bist30 index returns for even short-term analysis. As a result, this study points out a useful pathway for the future researches to show that social media posts may convey useful information for financial markets even for short term analysis.

Keywords: Correlation analysis, sentiment analysis, Turkish tweets, Bist30

1. INTRODUCTION

In our age where a customer-oriented approach prevails, financial enterprises have started to give more importance to the use of social media. Not only financial enterprises but also individual investors try to reach positive or negative general opinions about companies by following the news shared on social media. They update their asset portfolios based on this information and/or may turn to new investment channels. Liu et al. [1] state that people invested in stocks that are familiar to them. One of the reasons that feeds the familiarity of stocks is the news sources shared about financial markets [2]. Social media enables developments to spread much faster than many sources and allows more ideas to be collected at the same time [3]. Sprenger et al. [4] state that individual investors who follow Twitter are more profitable.

Nowadays, when the use of social media is so widespread, the number of studies examining whether positive or negative comments shared about financial businesses create a dynamism in financial markets is increasing. Mao et al. [5] investigate whether the daily number of tweets can be used to predict the S&P 500 stock market index over a period of

4 months. Sprenger et al. [6] report a relationship between stock-related tweets sentiments and stock market returns on a daily basis. Valle-Cruze et al. [7] analyzes the behavior of stock markets based on tweets during the COVID-19 pandemic and reports that the markets reacted 0 to 10 days after the tweets were shared. Karabulut [8] states that Gross National Happiness index of Facebook is efficient to estimate US daily returns. The number of studies conducted in this context for Turkey is very limited [2, 9, 10, 11]. These studies did not evaluate the analyzed period by dividing it into shorter periods. Based on such studies, the main motivation in this study is to investigate whether there is a correlation between the sentiments of tweets and the trend of the Bist30 index returns on a daily, weekly and monthly basis. Thus, it will be possible to investigate whether there are significant short-term correlation results between social media posts and the trend of stock markets. Additionally, it can be seen whether social media platforms provide useful information for short term analysis. Therefore, our aim is to investigate the relationship between the Turkish tweets related to Borsa Istanbul 30 (Bist30) companies and the trend of the Bist30 index returns on a daily, weekly and monthly basis.

In order to determine the existence of the relationship between Bist30 index returns and tweets, the tweets are labeled as positive, negative or neutral after passing through sentiment analysis stages. For this process, dictionary-based or machine learning-based approaches can be used. In dictionary-based approaches, textual data is labeled using domain-independent sentiment dictionaries such as Afinn [12], OpinionMiner [13], and Sent WordNet [14]. Although these dictionaries speed up the labeling process, they may be insufficient because they cannot carry the variation of emotions in special areas or different cultures. Loughran and McDonalds [15] state that the results of the analysis are affected negatively, if general-purpose sentiment dictionaries are used to classify financial text data. Unfortunately, there is no open-source Turkish emotion dictionary created specifically for the finance field. For all these reasons, to create the training dataset, the opinions of financial investors are taken as a basis in the labeling phase of social media posts. During the period between 07.05.2018 and 30.04.2019, when our study was started, 57933 tweets shared on Twitter regarding the stocks listed in the Bist30 index are labeled as positive, negative and neutral by financial investors manually. By using these labeled tweets as a training dataset, a test dataset consisting of two hundreds thousand tweets from the same time period are assigned to three classes as positive, negative and neutral by different machine learning algorithms (Support Vector Machines (SVM), Logistic Regression (LR), Naïve Bayes (NB), Decision Trees (DT), K-Nearest Neighborhood (KNN) classifiers and Majority Voting (MO) ensemble learner). Actually, we conducted these stages in our previous study [11]. This study differs from our previous study [11] in that the analyzes are carried out daily, weekly and monthly. We conducted the experiments for daily, weekly and monthly time periods in the time interval (07.05.2018 - 30.04.2019) in which tweets are examined. Hence, we can see whether social media platforms provide useful information for short term analysis. By using the number of tweets under the positive and negative categories, we calculate sentiment polarity scores of the tweets for daily, weekly and monthly basis. After that, the correlation analysis steps are applied on the polarity scores and the trend of Bist30 index returns for daily, weekly and monthly basis analysis with all the necessary statistical tests. The experimental results show that there are significant correlations between the sentiment polarity values and the changes of Bist30 index returns. We think this study points out a useful pathway for the future researches to show that social media posts may also convey useful information for short-term analysis.

We present our study in this paper with six more sections, which include related studies, dataset and methods, results, and conclusion.

2. RELATED WORKS

Researches that analyze whether there is a relationship between the sentiment values of tweets and stock market indexes' returns make use of methods such as sentiment analysis, correlation analysis and Granger causality analysis.

Sentiment analysis aims to evaluate the feelings and thoughts of communities on specific or general subjects. The studies

most dominantly use machine learning approaches to extract positive, negative or neutral opinions of communities. Kaukaras et al. [16] apply sentiments analysis on financial tweets with seven different machine learning algorithms. De Oliveira Carosia et al. [17] utilize artificial neural network architecture to perform sentiment analysis in financial news in Brazilian Portuguese and state that investment strategies based on sentiment analysis can bring profitability for investors. Utilizing financial related reviews on Eastmoney.com, Qui et al. [18] proposes a modified sentiment index depend on the number of positive, negative and neutral reviews that are extracted by Baidu AI Cloud sentiment analysis model. Yue et al. [19] presents a survey paper about sentiment analysis in social media. Hamraoui and Boubaker [20] investigate the correlations between tweets and the Tunisian financial market considering over a 12-month period and show that tweets can be utilized for price volatility. By using dictionary-based software, Bollen et al. [21], divided approximately 9.8 million tweets shared in a 10-month period into six sentiment categories (Calm, Alert, Sure, Vital, Kind, and Happy). They stated that there was a Granger causality between the tweets and the Dow Jones Index (DJIA) return values. Similarly, Mittal and Goel [22] examined the relationship between DJIA values and 476 million Twitter posts divided into 4 different sentiment categories with a dictionary-based approach and stated that there was a Granger causality between DJIA index returns and public moods captured on tweets. The paper reported that two public moods 'calm' and 'happy' could be used to predict future DJIA movements.

Eliaçık and Erdoğan [9] labeled 2408 Twitter posts, covering a 6-month period, with positive and negative labels, and proposed an approach that demonstrates the credibility of the person who shared the post in order to calculate the sentiment polarity values of tweets. As a result, they found a significant correlation between the proposed sentiment polarity value and the weekly value changes of the Bist100 index. In addition, they reported that the Pearson coefficient increased when the extraordinary events occurred during the collection period of the data set.

Yıldırım and Yüksel [10] classified 500 Twitter posts collected about a company in the telecommunications sector and traded on Borsa Istanbul with positive and negative tags using machine learning methods. They performed the daily correlation analysis between the sentiment polarity values obtained from tweets classified in their study and the firm's stock values and reported that there was a moderate and negative relationship between the two samples.

Ateş and Güran [11] applied the correlation and Granger causality analysis methods for Turkish financial tweets. They utilize different text representations (tokens, n-grams, Doc2Vec combined vectors etc.) and apply six different machine learning algorithms to assign the Turkish financial-related tweets into positive, negative and neutral classes. Based on the sentiment analysis of Turkish financial-related tweets, the results of correlations and Granger causality analyses steps for two different time periods indicated a causality from the stock returns to tweet sentiments for a long term and significant correlations for both short and long terms.

Ranco et al. [23] labeled 100000 stock related tweets with positive, negative and neutral labels manually. By using this training dataset, they classified 1.5 million tweets with support vector machine algorithm. Based on the calculated sentiment polarity values they stated a low Pearson correlation and Granger causality results between sentiments and stocks.

After labeling 3216 posts shared on Twitter as positive, negative and neutral, Pagolu et al. [24] classified a test data set consisting of 250 thousand tweets with N-gram and Word2Vec methods. After obtaining the DJIA index changes for a period of one year, they designed a system to transform the relationship between tweets and index value changes into a classification problem by using the number of tweets under positive and negative categories. As a result, they determined that the emotional changes in tweets were effective on the index value changes.

Deng et al. [25], with a dictionary-based software, labeled approximately 18 million tweets collected from the StockTwit platform in a 4-year period as positive and negative. Afterwards, they analyzed whether there was Granger causality between the daily and hourly value changes of the DJIA index and the sentiment polarity values obtained from tweets. As a result, they stated that there is no Granger causality when the daily changes between the two-time series are taken as basis, but negative tweets in hourly analysis are effective in the causality relationship.

Zhao [26] conducted the correlation and Granger causality analysis between the Singapore Straits Times Index (STI) values, which consist of the first 30 stocks with the highest market value, and the happiness indexes obtained from Twitter posts. This happiness index is calculated by determining the word consisting of 10 thousand emotion indicators such as love, happiness, smile and fun in approximately 50 million tweets shared by Twitter users by natural language processing methods. In the end, they stated that there is a positive correlation between the Twitter happiness index and the STI Index return values covering the 4-year period, which is also supported by the Granger causality analysis.

3. DATASET AND SENTIMENT ANALYSIS PROCESS

In this study, two hundred thousand tweets posted between 07.05.2018 - 30.04.2019 are used as a corpus. By labelling 57933 tweets manually, the study applied sentiment analysis steps on two hundred thousand test tweets and labeled those tweets as positive, negative and neutral by using LR, SVM, NB, DT, KNN and MV classifiers. Before classifiers are applied, words starting with '@', '\$' characters, expressions containing URL addresses and punctuation marks in tweets are cleaned. Tokens such as '%+', '+%', '-%', '%-', ':)', ':(' are not removed because they are assumed to reflect positive/negative moods. In all experiments where the performance values of different classifiers are compared, 80% of the training dataset is used for training and 20% for testing. In addition, 10-fold cross-correction method is applied. Python programming language is used during the experiments. The linear kernel function is chosen for the

SVM algorithm. In the KNN method, the Euclidean distance is taken as a basis, taking into account the 3 nearest neighbors. The Gini index is used for DT classifier. In MV ensemble learning, majority voting strategy is used. Classification results of 57933 tweets vectorized with the use of 50 thousand features with high tf-idf values are as indicated in Table 1. At this stage, apart from word-based 1-grams, the effects of 2-grams and 3-grams are also examined.

Table 1. Classification results of 57933 tweets

	1-Gram	2-Gram	3-Gram
LR	%63,41 (+/- 0,28)	%63,60 (+/- 0,29)	%63,47 (+/- 0,24)
SVM	%62,53 (+/- 0,48)	%62,32 (+/- 0,54)	%62,32 (+/- 0,47)
NB	%59,55 (+/- 0,32)	%61,58 (+/- 0,43)	%61,40 (+/- 0,56)
KA	%58,01 (+/- 0,30)	%57,60 (+/- 0,36)	%57,47 (+/- 0,41)
KNN	%56,16 (+/- 0,45)	%55,43 (+/- 0,45)	%55,34 (+/- 0,49)
MV	%63,03 (+/- 0,24)	%63,39 (+/- 0,22)	%63,29 (+/- 0,20)

As it can be seen from Table 1, the best performance value belongs to the LR algorithm with 63.60% accuracy value. The best performance result has been obtained with 2-grams. Many experiments are carried out with the majority voting method, which is the ensemble learning method, and different combinations and individual classifiers are used. The best performance results are obtained by combining LR, SVM and NB algorithms (63.39%). However, this result is still lower than LR classifier. After this step, two hundred thousand tweets are classified with LR algorithm based on 2-grams and the number of positive and negative tweets are detected. Figure 1-3 show the daily, weekly and monthly distribution of the classified tweets that are posted between 07.05.2018 - 30.04.2019 for short-term analysis.

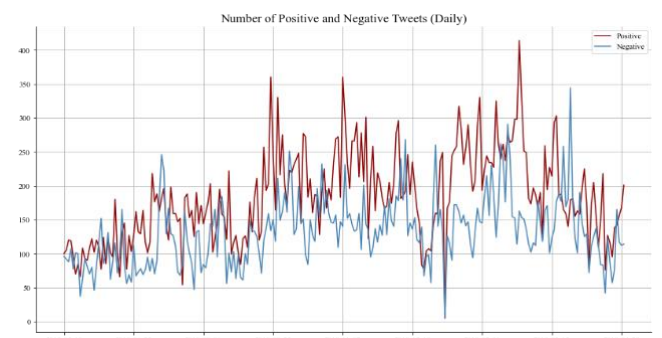


Figure 1. Distribution of the number of tweets per day

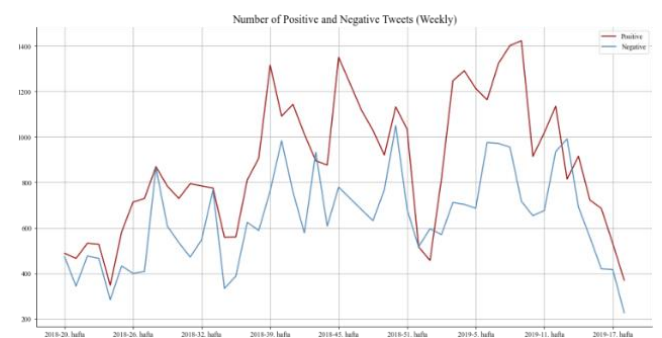


Figure 2. Distribution of the number of tweets per week

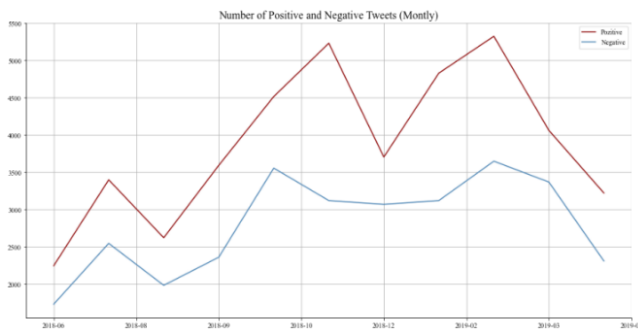


Figure 3. Distribution of the number of tweets per month

Detecting the number of positive and negative tweets per day, per week and per month, sentiment polarity values are calculated for each time period. We use the following equations for calculating the sentiment polarity values on a daily, weekly and monthly basis.

$$S_d = \frac{p_{td}}{p_{td} + n_{td}} - \frac{p_{td-1}}{p_{td-1} + n_{td-1}} \quad (1)$$

$$S_w = \frac{p_{tw}}{p_{tw} + n_{tw}} - \frac{p_{tw-1}}{p_{tw-1} + n_{tw-1}} \quad (2)$$

$$S_m = \frac{p_{tm}}{p_{tm} + n_{tm}} - \frac{p_{tm-1}}{p_{tm-1} + n_{tm-1}} \quad (3)$$

where p_{ti} and n_{ti} indicate the number of positive and negative tweets ($i = d, w, m$). After this step, we extract the daily, weekly and monthly movements of the closing prices of Bist30 with the following equations (Eq 4-5-6):

$$R_d = \frac{P_d - P_{d-1}}{P_{d-1}} \quad (4)$$

$$R_w = \frac{P_w - P_{w-1}}{P_{w-1}} \quad (5)$$

$$R_m = \frac{P_m - P_{m-1}}{P_{m-1}} \quad (6)$$

where P_d , P_w and P_m are the closing prices of the stock at day d , week w and month m respectively.

Then correlation analysis steps are performed on the sentiment polarity values and the changes of Bist30 index returns.

4. CORRELATION ANALYSIS

Pearson Correlation is a statistical analysis used to depict the strength and direction of relationship between two variables. In correlation analysis, the degree of correlation is shown with the correlation coefficient (r) that lies between -1 and $+1$. Positive one indicates a perfect positive linear relationship between the variables; whereas negative one points out a perfect negative association between variables and a value of zero coefficient indicates no correlation. The absolute value of r indicates the strength of the existing relationship. Although the strength may vary by discipline the following general guidelines can be used to interpret the correlation coefficient [27].

$0.1 < r < 0.3$	small / weak correlation
$0.3 < r < 0.5$	medium / moderate correlation
$0.5 < r $	large / strong correlation

Many statistical methods make assumptions about normality. In general, graphical and numerical methods can be used to assess the normality of the data [28, 29]. Skewness, kurtosis, Shapiro–Wilk test, Kolmogorov–Smirnov test, histogram and normal Q–Q plot are the most popular methods to evaluate the normal distribution.

Skewness refers to the deviation from symmetry and kurtosis reflects the degree of the sharpness / peakedness of a distribution [18]. In order to give “excess” kurtosis SPSS subtracts 3 from the kurtosis. By using skewness and excess kurtosis, the z -test can be used for testing the normality. In order to calculate the z -scores, we need to divide the skew values or excess kurtosis by their standard errors. If sample size is less than 50 and the z -values fall between $+1.96$ and -1.96 for either skewness or kurtosis, with an alpha level 0.05, then the data is said to be normally distributed [30]. If sample size is between $50 \leq n < 300$ and z -values fall between -3.29 and $+3.29$ for $\alpha = 0.05$, the data is assumed to be normally distributed [31]. For sample size greater than 300, the histograms and the absolute values of skewness and kurtosis should be examined for normality of the data (the absolute skewness value that is less than or equal to 2 or the absolute kurtosis (excess) that is less than or equal to 4 could be the limit values for normality) [32].

Pearson and Spearman statistical tests are the most commonly used tests to assess whether the data is normally distributed. If sample size is less than fifty, the Shapiro–Wilk test is preferred; otherwise, Kolmogorov–Smirnov test is used [31].

Histograms are used to show the distribution of a dataset. For the normally distributed dataset, the histogram graphs are nearly bell-shaped and symmetric about the mean [32,33].

In statistics, a normal Q–Q plot shows whether a dataset comes from Gaussian distribution. If the dataset is normally distributed, data points are seen as a straight line on the graph.

In our study, the suitability of sentiment polarity measurements and the stock value changes to normal distribution were assessed with the tests specified (Skewness& Kurtosis z -values, Kolmogorov-Smirnov or Shaphiro-Wilk tests, the Histograms and Normal Q-Q plots). We use IBM SPSS [34] software to perform the required steps.

5. NORMALITY TEST AND CORRELATION ANALYSIS RESULTS

The aim of this section is to perform the correlation analysis between the sentiment polarity values (S_d, S_w, S_m) and daily, weekly and monthly movements of the closing prices of Bist30 (R_d, R_w, R_m). Before starting the correlation analysis, it is tested whether the sentiment polarity values (S_d, S_w, S_m) and (R_d, R_w, R_m) values fit the normal distribution. For normality test, we analyze different numerical and visual methods: skewness and kurtosis z -values, the Kolmogorov-Smirnov/Shapiro-Wilk tests p -values, histogram and normal Q-Q plots. These numerical and visual outputs will indicate whether our analyzed variables are approximately normally

distributed.

5.1. Normality Test Results

In this section, in order to test the normality of R_i and S_i values ($i=d, w, m$), first of all we look at the skewness and kurtosis z-values, then the p-values of Kolmogorov-Smirnov/Shapiro-Wilk tests and finally histogram and normal Q-Q plots.

5.1.1. Daily Analysis Results

Using SPSS to test for normality of daily R_d and S_d values, first of all, we focus on skewness and excess kurtosis and divide the skew values and excess kurtosis by their standard errors to get the z-scores. According to Table 2, skewness and excess kurtosis z-scores for S_d are $(|-0.040/0.156|=0.25641, |-0.238/0.312|=0.76282)$ respectively. Since we have 242 trading dates and the absolute values of these numbers should not exceed 3.29 with an alpha level 0.05, it can be said that S_d values are approximately normally distributed. Similarly, the skewness and kurtosis z-scores for R_d are $(|-0.398/0.156|= 2.55128, |0.966/0.312|= 3.096154)$ respectively. Again, since the absolute values of these numbers are not exceed 3.29, we can accept the null hypothesis and say that R_d values are approximately normally distributed.

Table 2. Skewness and Kurtosis values for daily analysis

		Statistic	Std. Error
S_d	Skewness	-.040	.156
	Kurtosis	-.238	.312
R_d	Skewness	-.398	.156
	Kurtosis	.966	.312

Table 2 presents the results of Kolmogorov-Smirnov and the Shapiro-Wilk tests at a significance level of 0.05. As stated before the number of trading days considered between 07.05.2018 and 30.04.2019 was 242. Because of this, we focus on the Kolmogorov-Smirnov test as our numerical means of assessing normality. Since p values are greater than 0.05, Table 3 shows that daily S_d and R_d values are normally distributed.

Table 3. Tests of normality on a daily basis

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
S_d	.041	242	.200*	.996	242	.719
R_d	.047	242	.200*	.987	242	.028

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

As a next step, we look at the graphical figures (Figure 4 and Figure 5) for daily S_d and R_d values. The histogram diagrams show us that the S_d and R_d are approximately normally distributed.

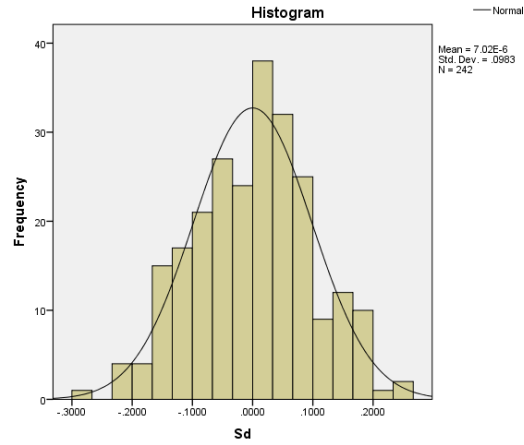


Figure 4. Histogram diagram of daily S_d

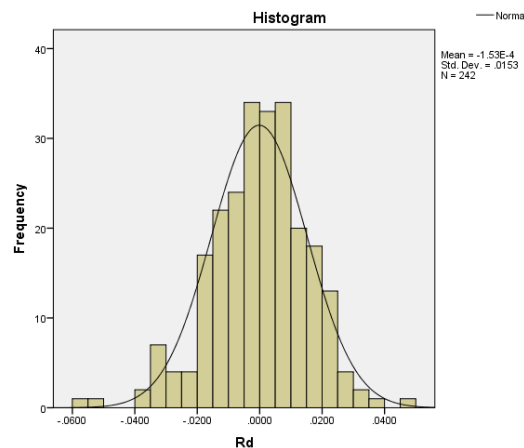


Figure 5. Histogram diagram of daily R_d

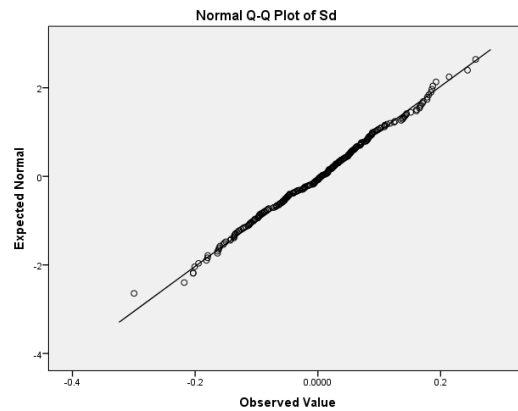


Figure 6. Q-Q plot for diagram of daily S_d

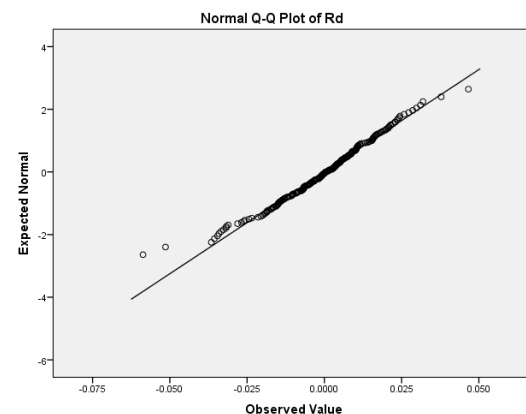


Figure 7. Q-Q plot for diagram of daily R_d

Then we look at the normal Q-Q diagrams and since the dots are along the lines, these figures indicate that S_d and R_d are approximately normally distributed.

All the results indicate that S_d and R_d are approximately normally distributed.

5.1.2. Weekly Analysis Results

Table 4 presents normality test results for weekly analysis. It is seen that skewness and excess kurtosis z-scores for weekly S_w are $(|-0.522/0.337|=1.5489614, |0.315/0.662|=0.475831)$ respectively. Since we have 50 trading weeks and the absolute values of these numbers should not exceed 1.96 with an alpha level 0.05, it can be said that S_w values come from a normal distribution.

Table 4 also indicates that the skewness and excess kurtosis z-scores for weekly R_w are $(|-0.371/0.337|=1.10089, |0.541/0.662|=0.817221)$ respectively. Having 50 trading weeks and since the absolute values of these numbers are not exceed 3.29 with an alpha level 0.05, we can accept the null hypothesis and say that R_w values are also approximately normally distributed.

Table 4. Skewness and kurtosis values for weekly analysis

		Statistic	Std. Error
S_w	Skewness	-.522	.337
	Kurtosis	.315	.662
R_w	Skewness	-.371	.337
	Kurtosis	.541	.662

There are several methods for normality test such as Kolmogorov-Smirnov normality test and Shapiro-Wilk’s test. Since it is more convenient to use Shapiro-Wilk’s test for small sample sizes, we focus on Shapiro-Wilk test statistics. The null hypothesis is determined as follows: “ H_0 : The data is normally distributed”. It will be rejected if the corresponding p-value is below 0.05. In SPSS output, the p-value is labeled as “Sig.” Since the Sig. values (0.485 and 0.837) are greater than 0.05 for weekly S_w and R_w respectively, we keep the null hypothesis. That’s why it can be said that S_w and R_w are approximately normally distributed.

Table 5. Tests of normality on a weekly basis

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
S_w	.114	50	.124	.978	50	.485
R_w	.078	50	.200*	.987	50	.837

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

The graphical figures for S_w and R_w values can be examined with Figure 8-11. The histogram diagrams (Figure 8 and Figure 9) appear to support that the S_w and R_w are approximately normally distributed. As it is seen the diagrams are quiet symmetrical. There are big peaks and small tails on the sides.

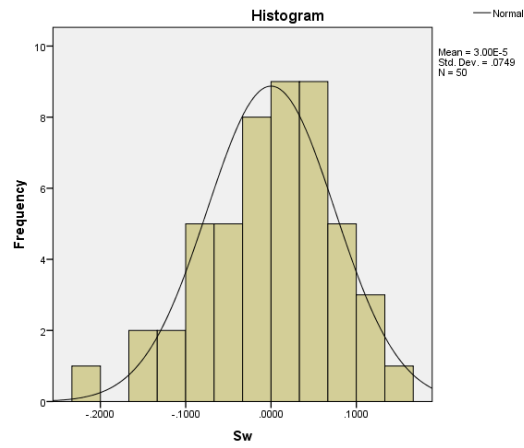


Figure 8. Histogram diagram of weekly S_w

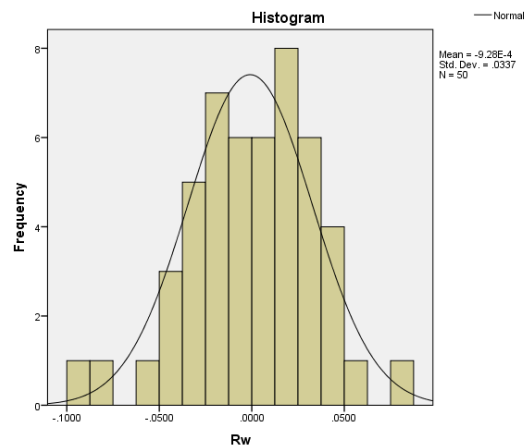


Figure 9. Histogram diagram of weekly R_w

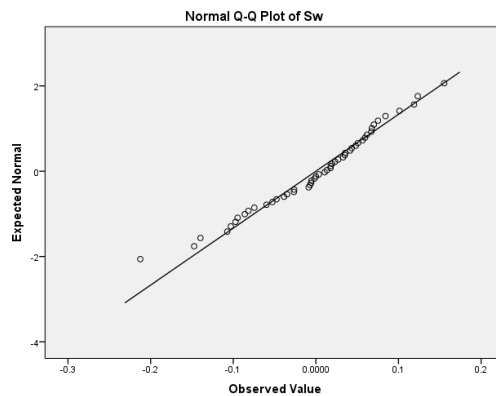


Figure 10. Q-Q plot for diagram of weekly S_w

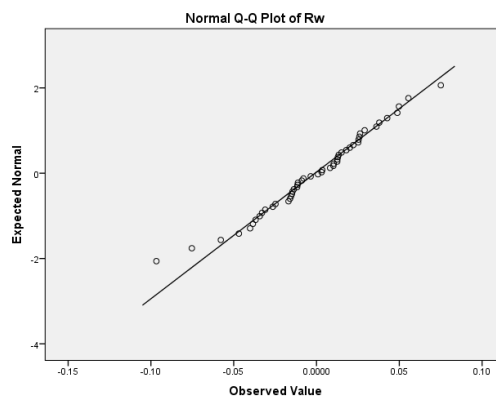


Figure 11. Q-Q plot for diagram of weekly R_w

Then we look at the normal Q-Q diagrams and since the dots are along the lines, Figure 10 and Figure 11 indicate that weekly S_w and R_w are approximately normally distributed.

All the results indicate that S_w and R_w are approximately normally distributed.

5.1.3. Monthly Basis Results

Table 6 shows normality test results for monthly analysis. Skewness and excess kurtosis z-scores for monthly S_m are $(|-0.425/0.661|=0.642965, |-0.839/1.279|=0.6559812)$ respectively. Since we have 11 trading months and the absolute values of these numbers should not exceed 1.96 with an alpha level 0.05, it can be said that monthly S_m values are approximately normally distributed. The skewness and excess kurtosis z-scores for monthly R_m are $(|-0.371/0.337|=1.10089, |0.541/0.662|=0.817221)$ respectively. Since we have 11 trading months and the absolute values of these numbers are not exceed 3.29 with an alpha level 0.05, we can accept the null hypothesis and say that monthly R_m values are also approximately come from a normal distribution.

Table 6. Skewness and kurtosis values for monthly analysis

		Statistic	Std. Error
S_m	Skewness	-.425	.661
	Kurtosis	-.839	1.279
R_m	Skewness	1.171	.661
	Kurtosis	2.005	1.279

It is more appropriate to use Shapiro-Wilk’s test for small sample sizes, because of this reason we focus on Shapiro-Wilk test statistics for monthly analysis. “ H_0 : The data is normally distributed” is the null hypothesis for this test of normality and it should be rejected if the Sig. Value (p-value) is less than 0.05. As it is seen from Table 7, the Sig. values (0.600 and 0.279) are greater than 0.05 for S_m and R_m respectively. This means we should keep the null hypothesis and assume that S_m and R_m are approximately normally distributed.

Table 7. Tests of normality on a monthly basis

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
S_m	.186	11	.200*	.947	11	.600
R_m	.172	11	.200*	.915	11	.279

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Figure 12 and 13 depict the histogram diagrams of S_m and R_m .

Based on the smallest-sized sample histogram diagrams don’t point out a symmetrical bell shape.

Figure 14 and 15 show Q-Q diagrams and since the dots are along the lines, they point out that monthly S_m and R_m are approximately normally distributed.

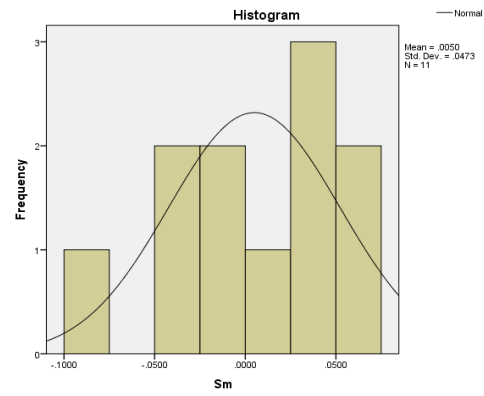


Figure 12. Histogram diagram of monthly S_m

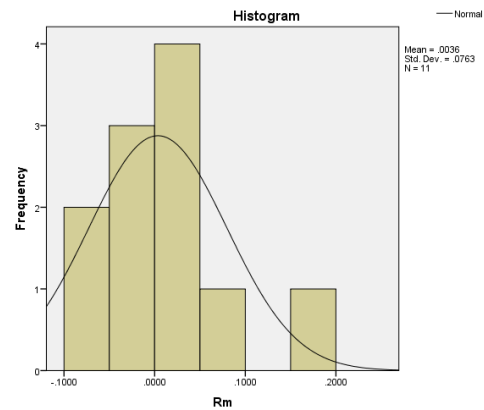


Figure 13. Histogram diagram of monthly R_m

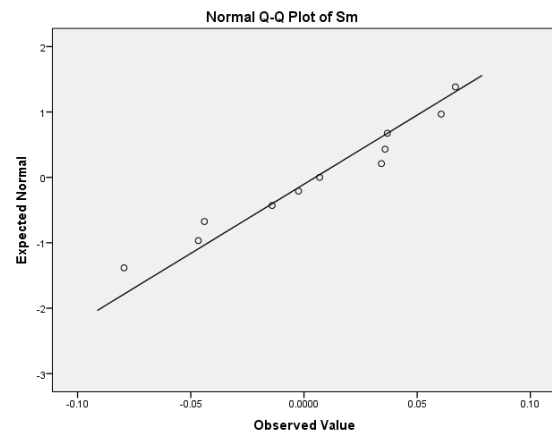


Figure 14. Q-Q plot for diagram of monthly S_m

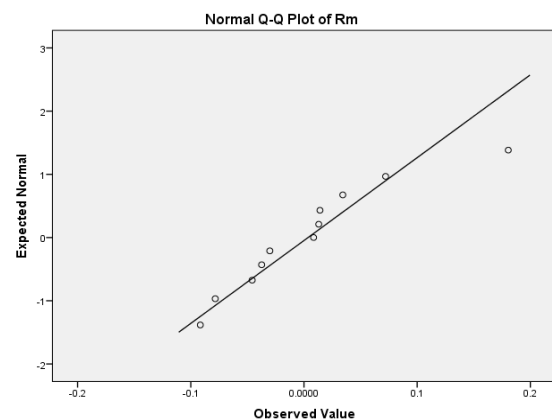


Figure 15. Q-Q plot for diagram of monthly R_m

Although the histogram diagram the belongs to the smallest-sized sample doesn't reflect a symmetrical bell shape, we could conclude that based on the other examined methods S_m and R_m satisfy the normality assumptions.

5.2. Correlation Results

We conduct various visual and statistical normality tests on our variables to investigate the assumption of normality. The results show that daily, weekly and monthly variables seem to support the normality assumptions. This section gives Pearson correlation results for daily, weekly and monthly analysis.

Table 8. Daily correlations between S_d and R_d

		S_d	R_d
S_d	Pearson Correlation	1	.509**
	Sig. (2-tailed)		.000
	N	242	242
R_d	Pearson Correlation	.509**	1
	Sig. (2-tailed)	.000	
	N	242	242

** . Correlation is significant at the 0.01 level (2-tailed).

Table 8 depicts the correlation results among S_d and R_d values. According to Table 8, it is seen that S_d and R_d have a statistically significant linear relationship ($r=.509$, $p < .001$). When the direction of the relationship is analyzed, it is seen that S_d and R_d tend to increase together. Since $|r| > .5$ we can say that the association is approximately strong.

Table 9. Weekly correlations between S_w and R_w

		S_w	R_w
S_w	Pearson Correlation	1	.602**
	Sig. (2-tailed)		.000
	N	50	50
R_w	Pearson Correlation	.602**	1
	Sig. (2-tailed)	.000	
	N	50	50

** . Correlation is significant at the 0.01 level (2-tailed).

Table 9 indicates the weekly correlation results between S_w and R_w values. As it is seen the linear relationship between S_w and R_w is also statistically significant ($r=.602$, $p < .001$). There is a positive association and the magnitude of the association is quite strong ($|r| > .5$).

Table 10. Monthly correlations between S_m and R_m

		S_m	R_m
S_m	Pearson Correlation	1	.707*
	Sig. (2-tailed)		.015
	N	11	11
R_m	Pearson Correlation	.707*	1
	Sig. (2-tailed)	.015	
	N	11	11

* . Correlation is significant at the 0.05 level (2-tailed).

Table 10 presents the monthly correlation results among S_m and R_m values. S_m and R_m have a significant linear relationship. Since ($|r| = 0.707 > 0.5$) it can be said that the association is strong. Due to the positive correlation

coefficient, it can be said that S_m and R_m moves together in the same direction.

6. CONCLUSION

Nowadays instead of making the analyzes suggested by classical theories, people can make their individual financial investment decisions by taking certain shortcuts. Positive and negative comments shared on social media have become one of those shortcuts that are taken into consideration when making investment decisions. With the effect of those comments, investors can revise their portfolios. In literature there are lots of researches that investigate the relations between social media posts and stock markets. Twitter is one of the most important social media platforms. Researches indicate that there is a correlation between tweets sentiments and the changes in stock prices for both long-term and short-term analysis. Unfortunately, most of the studies are related with analysis of social media platforms in English language. There are very limited number of studies related with Turkish language, especially for short-term analysis. Hence, with this motivation, the main aim of this study is to investigate the correlation between sentiments of tweets related to the Bist30 index and the index movement on daily, weekly and monthly basis.

We conduct sentiment analysis stages and correlation analysis steps to investigate whether there is a relation between tweet sentiments and returns of Bist30 index returns for short-term analysis. The analyzes are carried out for the time interval of 07.05.2018 - 30.04.2019. Before starting the correlation analysis, the normal distribution suitability of all variables are tested by using different graphical and numerical methods. After that the results of the correlation analysis between the sentiments of tweets and the change of index returns are presented. The presented results indicate that there are significant positive correlations between sentiments of tweets and the change of index returns. With these results, it has been seen that social media posts are likely to affect financial markets. The findings obtained as a result of our study are in parallel with many studies in the literature [6, 7, 8, 9, 10, 11, 22, 24, 26]. This study, which is based on short-term analysis, also supports that social media resources can be effective to predict the trend of stock markets.

For future studies, Granger causality analysis steps for short-term analysis will be studied and also we will investigate whether tweets can be an indicator that can be used to predict the returns of stock market indices. We think the present study points out a useful direction of future research to improve the understanding of the effects of social media on financial markets.

Author contributions: Emine Ateş: Data collection, Software, Literature review; Aysun Güran: Literature review, Investigation, Software, Writing-Review & Editing, Analysis of results.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study has received no financial support.


REFERENCES


- [1] Liu, L. X., Ann E. S., Yong Z., The long-run role of the media: Evidence from initial public offerings, *Management Science*, 60 (8), 1945-1964, 2014.
- [2] Atan, S., Çinar, Y. Borsa İstanbul'da finansal haberler ile piyasa değeri ilişkisinin metin madenciliği ve duygu (sentiment) analizi ile incelenmesi, *Ankara Üniversitesi Sbf Dergisi*, 74 (1), 1-34, 2019.
- [3] Teti, E., Dallochio, M., Aniasi, A. The relationship between twitter and stock prices. Evidence from the US technology industry, *Technological Forecasting and Social Change*, 149 (119747), 1-9, 2019.
- [4] Sprenger T. O., Tumasjan A., Sandner P. G., Welpe, I. M., Tweets and trades: The information content of stock microblogs, *European Financial Management*, 20 (5), 926-957, 2014.
- [5] Mao Y, Wei W, Wang B, Liu B, Correlating S&P 500 stocks with Twitter data. In: *Proceedings of the 1st ACM international workshop on hot topics on interdisciplinary social networks research*, 69–72, 2012.
- [6] Sprenger T.O., Sandner P.G., Tumasjan A., Welpe I.M., News or noise? Using Twitter to identify and understand company-specific news flow, *J Bus Finance Account* 41(7–8), 791–830, 2014.
- [7] Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A., & Sandoval-Almazán, R., Does twitter affect stock market decisions? financial sentiment analysis during pandemics: A comparative study of the h1n1 and the covid-19 periods. *Cognitive computation*, 14(1), 372-387, 2022.
- [8] Karabulut, Y., Can Facebook predict stock market activity? AFA 2013 San Diego Meetings Paper, Available at SSRN: <https://ssrn.com/abstract=2017099>, 2013.
- [9] Eliaçık A. B., Erdogan N., Mikro Bloglardaki Finans Toplulukları için Kullanıcı Ağırlandırmış Duygu Analizi Yöntemi, *Ulusal Yazılım Mühendisliği Sempozyumu*, İzmir, 782-793, 2015.
- [10] Yıldırım M., Yüksel C. A., Sosyal Medya ile Hisse Senedi Fiyatının Günlük Hareket Yönü Arasındaki İlişkinin İncelenmesi: Duygu Analizi Uygulaması. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, 22, 33-44, 2017.
- [11] Ates, E., Guran, A. (2021). Pearson correlation and Granger causality analysis of Twitter sentiments and the daily changes in Bist30 index returns. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 36(3), 1688-1701.
- [12] Nielsen F., A new anew: evaluation of a word list for sentiment analysis in microblogs, In *The ESQ2011 Workshop on Making Sense of Microposts*, Heraklion, Crete, 93-98, 2011.
- [13] T. Wilson, J. Wiebe, P. Hofmann, Recognizing contextual polarity in phrase-level sentiment analysis, in: *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing—HLT*, 347–354, 2005.
- [14] Baccianella, S., Esuli, A., Sebastiani, F., Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining, In *LREC'10*, Valetta, 2200-2204, 2010.
- [15] Loughran, T. I. M., McDonald, B., When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35-65, 2011.
- [16] Koukaras, P., Nousi, C., Tjortjis, C., Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning, In *Telecom MDPI*, 3(2), 358-378, 2022.
- [17] De Oliveira Carosia, A. E., Coelho, G. P., da Silva, A. E. A., Investment strategies applied to the Brazilian stock market: a methodology based on sentiment analysis with deep learning. *Expert Systems with Applications*, 184, 115470, 2021.
- [18] Qiu, Y., Song, Z., & Chen, Z. (2022). Short-term stock trends prediction based on sentiment analysis and machine learning. *Soft Computing*, 26(5), 2209-2224, 2022.
- [19] Yue, L., Chen, W., Li, X., Zuo, W., Yin, M., A survey of sentiment analysis in social media. *Knowledge and Information Systems*, 60(2), 617-663, 2019.
- [20] Hamraoui, I., Boubaker, A., Impact of Twitter sentiment on stock price returns. *Social Network Analysis and Mining*, 12(1), 1-15, 2022.
- [21] Bollen J., Mao H., Zeng X., Twitter mood predicts the stock market, *Journal of computational science*, 2 (1), 1-8, 2011.
- [22] Mittal A., Goel A., Stock prediction using twitter sentiment analysis, *Working Paper Stanford University*, CS 229, 1-5, 2012.
- [23] Ranco G., Aleksovski D., Caldarelli G., Grcar M., Mozetic I., The effects of Twitter sentiment on stock price returns, *PloS one*, 10 (9), 1-21, 2015.
- [24] Pagolu V. S., Reddy K. N., Panda G., Majhi B, Sentiment analysis of Twitter data for predicting stock market movements, In *2016 international conference on signal processing, communication, power and embedded system (SCOPEs)*, Paralakhemundi-Hindistan, 1345-1350, 3-5 Ekim, 2016.
- [25] Deng S., Huang Z. J., Sinha A. P., Zhao H., The Interaction between Microblog Sentiment and Stock Return: An Empirical Examination, *MIS quarterly*, 42 (3), 895-918, 2018.
- [26] Zhao R., Quantifying the correlation and prediction of daily happiness sentiment and stock return: The Case of Singapore, *Physica A: Statistical Mechanics and its Applications*, 533, 1-9, 2019.
- [27] <https://libguides.library.kent.edu/spss/pearsoncorr> Access date: June 1, 2019.

- [28] Bland M. 4th ed. Oxford: Oxford University Press; 2015. *An Introduction to Medical Statistics*.
- [29] Campbell MJ, Machin D, Walters SJ. 4th ed. Chichester: John Wiley & Sons, Ltd; 2007. *Medical Statistics: A text book for the health sciences*.
- [18] Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of cardiac anaesthesia*, 22(1), 67.
- [30] Ghasemi A, Zahediasl S. Normality tests for statistical analysis: A guide for non-statisticians. *Int J Endocrinol Metab*. 2012; 10:486–9.
- [31] Kim HY. Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod*. 2013; 38:52–4.
- [32] Armitage P, Berry G. 2nd ed. London: Blackwell Scientific Publications; 1987. *Statistical Methods in Medical Research*.
- [33] Barton B, Peat J. 2nd ed. Sydney: Wiley Blackwell, BMJ Books; 2014. *Medical Statistics: A Guide to SPSS, Data Analysis and Clinical Appraisal*.
- [34] IBM SPSS software. <https://www.ibm.com/tr-tr/analytics/spss-statistics-software>. Access date: June 1, 2019.

Synchronization of Incommensurate Fractional-Order King Cobra Chaotic System

¹Haris CALGAN, ^{*2}Abdullah GOKYILDIRIM

¹Department of Electrical and Electronics Engineering, Balıkesir University, Balıkesir, Turkey, haris.calgan@balikesir.edu.tr 

²Department of Electrical and Electronics Engineering, Bandırma Onyedi Eylül University, Balıkesir, Turkey, agokyildirim@bandirma.edu.tr 

Abstract

In this study, the incommensurate fractional-order King Cobra (IFKC) chaotic system has been investigated. Through bifurcation diagrams and Lyapunov exponent spectra, it has been determined that the IFKC system exhibits rich dynamics. Subsequently, using the Proportional Tilt Integral Derivative (P-TID) control method, synchronization of two IFKC chaotic systems with different initial values has been achieved. Upon examination of the obtained simulation results, it has been demonstrated that the identified IFKC chaotic system and the P-TID controller can be effectively utilized for secure communication.

Keywords: Incommensurate Fractional-Order; Chaos; Synchronization; TID Control

1. INTRODUCTION

The foundations of fractional calculus were laid approximately 300 years ago. Fractional-order analysis offers new perspectives for observing, modeling, and controlling the nature around us [1]. For this reason, many systems in fields such as physics [2], engineering [3], mathematical biology [4], health [5], computer science [6], and more can be described with the help of fractional derivatives.

Chaos theory has been meticulously examined and studied by numerous researchers since Lorenz's work in 1963. Particularly, the study of chaos in fractional dynamic systems in recent years has become an interesting topic. Therefore, fractional-order analyses have been carried out on well-known systems such as Lorenz, Chua, Chen, Rössler, Rucklidge, investigating chaotic behaviors [7-11].

After demonstrating that the dynamic behaviors of chaotic systems can be further diversified through fractional analysis, many researchers have explored the control and synchronization of fractional chaotic systems. The synchronization of chaotic systems forms the basis of chaotic masking, a chaotic-based secure communication method. Therefore, in chaotic masking, which is one of the chaotic-based secure communication methods, synchronization and control play a crucial role. The fundamental aim of synchronization is to ensure that two chaotic systems exhibit the same dynamic behavior after a certain period of time, facilitated by a designed controller. In the literature, various classical methods have been

employed for controlling or synchronizing chaotic systems, such as fractional-order PID [12], sliding mode [13], or optimal controllers [14]. Additionally, for the synchronization of chaotic systems, methods like time-delay feedback [15], active [16], passive [17], and adaptive control methods [18], linear quadratic regulator [19] as well as Lyapunov's direct control method [20], have been used.

When examining synchronization studies in the literature, it is observed that many studies involving fractional-order chaotic systems have utilized the commensurate fractional-order method. However, it has been reported that more complex chaotic behaviors can be obtained through incommensurate fractional-order analysis [21]. In this study, an analysis of the previously unexplored incommensurate behavior of the fractional-order King Cobra chaotic system [22] has been conducted. Chaotic behaviors have been identified using bifurcation diagrams and Lyapunov spectra. Moreover, synchronization of two chaotic systems has been achieved using the P-TID controllers, which is not widely employed in the literature for chaos control and synchronization.

The organization of this paper is structured as follows: In Section 2, the dynamic analysis of IFKC chaotic system is conducted using bifurcation diagrams and Lyapunov spectra. In Section 3, two IFKC chaotic systems with different initial conditions are synchronized using the P-TID control method. Finally, conclusions are drawn in Section 4.

2. DYNAMIC ANALYSES OF THE IFKC SYSTEM

Fractional derivatives and integrals have gained significance in engineering and mathematics, proving invaluable for scientists and researchers engaged in practical, real-world applications. One widely recognized fractional operator is Caputo's fractional derivative, introduced by Caputo in 1967 and applied in this paper. In the context of both continuous-time and discrete-time systems, the utilization of Caputo's differential operator facilitates the establishment of initial conditions for initial-value problems. The Caputo's derivative with starting point 0, of order q is defined as below [23]:

$$D^q x = J^{m-q} x^m \tag{1}$$

Here, m represents the integer closest to q , with $m > q$, and J^a denotes the a^{th} order Riemann-Liouville integral operator, expressed as [24]

$$J^a y = \frac{1}{\Gamma(a)} \int_0^t (t-\tau)^{a-1} y(\tau) d\tau \tag{2}$$

where $\Gamma(\cdot)$ is the Euler's gamma function. Using the q^{th} order Caputo fractional derivative, the King Cobra system is defined as follows

$$\begin{cases} D^{q_1} x_1(t) = a(y-x) + byz^2 \\ D^{q_2} y_1(t) = cx + dxz^2 \\ D^{q_3} z_1(t) = hz + k|x_1| \end{cases} \tag{3}$$

where parameters $a, b, c, d, h,$ and k are set to 10, 1, 5, -1, -5, and -6. Based on the theorem that defines the requisite condition for the existence of a double-scroll attractor in fractional-order systems [25], the system (3) demonstrates chaotic behavior when the commensurate fractional-order q value exceeds 0.8849 [22]. Hereby, if (3) is considered as commensurate fractional-order while $q_1=q_2=q_3=0.95$, the 2D phase portraits of the system are shown in Figure 1.

As illustrated in Figure 1, the system (3) is verified as a chaotic attractor and y-z phase portrait looks like a face of King Cobra at angry. However, the study in reference [22] has investigated the King Cobra chaotic system only by commensurate fractional-order analysis. In this study, the incommensurate fractional-order analysis is employed to discover a new chaotic response in the King Cobra system which has not been previously studied for this particular system. The effects of the incommensurate orders on the King Cobra chaotic system are investigated. The primary objective is to add dynamic richness to the system by selecting distinct fractional orders for each state equation. Hereby, the system involves more parameters to be adjusted in order to identify a wider range of chaotic behaviors. The bifurcation diagram is evaluated firstly, when 'a' is set as bifurcation parameters and incommensurate fractional orders q_1, q_2, q_3 are selected as 0.96, 0.97, 0.98, respectively. The bifurcation diagram of the system (3) is plotted in Figure 2 setting the initial conditions $(x_0, y_0, z_0) = (0.1, 1, 0.1)$ and change the parameter $a \in [4, 8]$.

It can be seen that when a is between 5.45 and 7.05, the incommensurate fractional-order system is in chaotic state. Setting the system parameter $a=7$, the initial value $[x_0, y_0, z_0]=[0.1, 1, 0.1]$ and $q_1=0.96, q_2=0.97, q_3=0.98$ the Lyapunov exponent is obtained as drawn in Figure 3. Note that the extended Benettion-Wolf algorithm for incommensurate fractional-order systems is used to determine Lyapunov exponents [26]. As highlighted in Refs. [23] and [26], the result of Lyapunov spectra depends highly on the Gramm-Schmidt coefficient in the algorithm. In this study, the Gramm-Schmidt coefficient is chosen as 0.9 with an integration step of 0.01. The corresponding Lyapunov exponents are obtained as $L_1=0.44, L_2=-0.03, L_3=-13.21$. The system exhibits chaotic behavior in this case due to the presence of Lyapunov exponents as $(+, 0, -)$ [27].

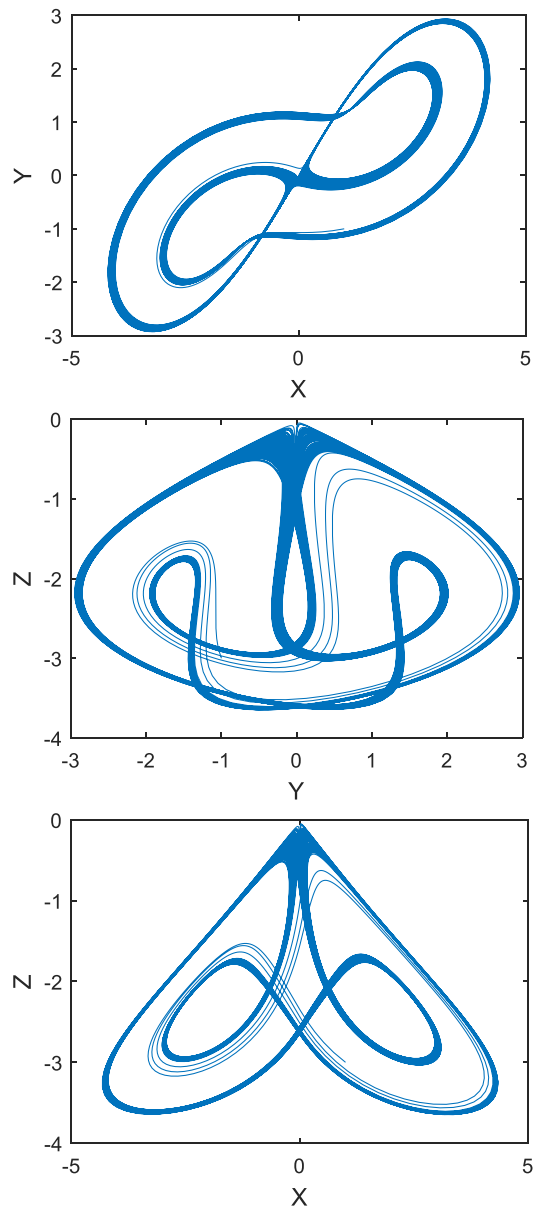


Figure 1. Phase planes of the system (3) when $q_1=q_2=q_3=0.95, a=6, b=1, c=5, d=-1, h=-5$ and $k=-6$.

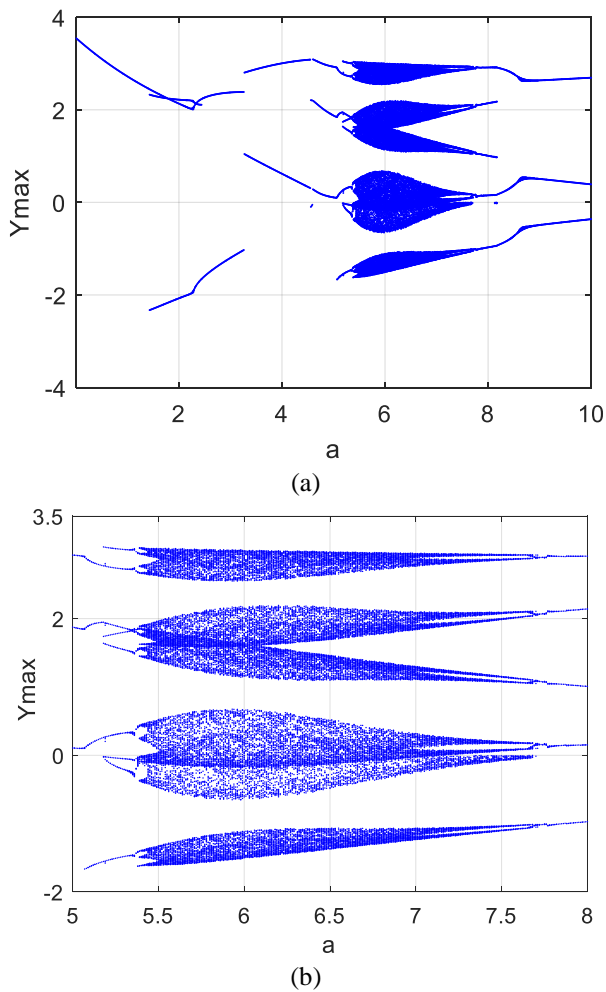


Figure 2. Bifurcation diagram of the system (3) when $q_1=0.96, q_2=0.97, q_3=0.98, b=1, c=5, d=-1, h=-5, k=-6$ and (a) $a \in [0,10]$, (b) $a \in [5,8]$.

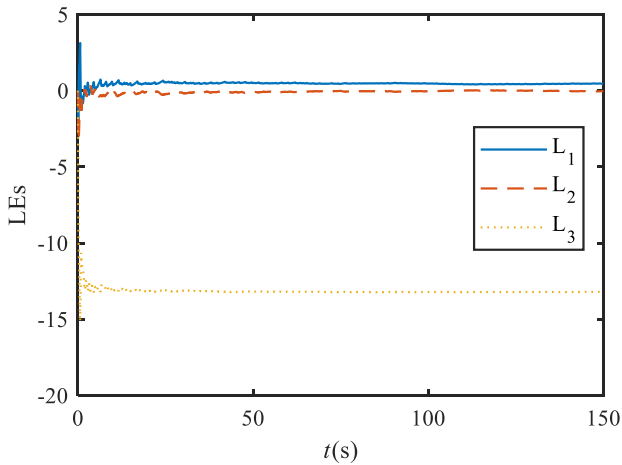


Figure 3. Lyapunov spectra of the system (3) when $q_1=0.96, q_2=0.97, q_3=0.98, a=7, b=1, c=5, d=-1, h=-5$ and $k=-6$.

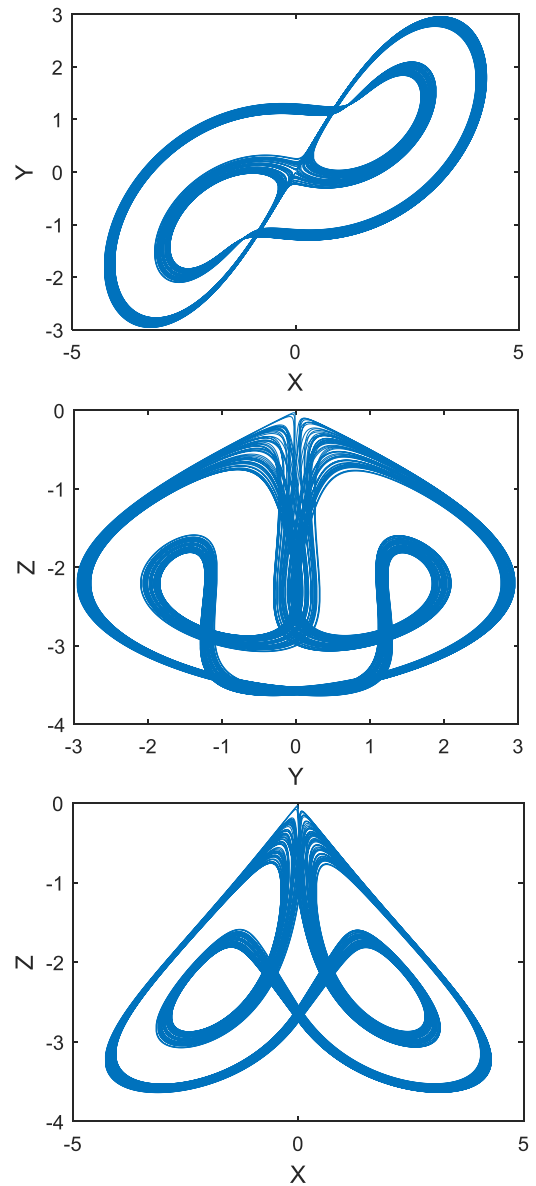


Figure 4. Phase planes of the system (3) when $q_1=0.96, q_2=0.97, q_3=0.98, a=7, b=1, c=5, d=-1, h=-5$ and $k=-6$.

Based on the incommensurate fractional-order analysis outcomes derived from bifurcation diagram and Lyapunov spectra, chaotic sequences are obtained by assigning the particular values to the system parameter $a=7$ and $q_1=0.96, q_2=0.97, q_3=0.98$. In Figure 4, phase portraits obtained by corresponding bifurcation diagram and Lyapunov spectra is provided. It is demonstrated through phase portraits that the Lyapunov exponent spectra are consistent with bifurcation diagram. The varying incommensurate fractional-order of equations has significant impact on the dynamic characteristics of the system. Consequently, more complex dynamic behaviors become observable when the system orders are incommensurate.

3. SYNCHRONIZATION

In chaos-based secure communication systems, the chaotic masking method is widely employed [20]. In this method, the signal generated by the chaotic system is added to the information signal and transmitted over a certain communication channel. On the receiver side, a second

chaotic system is also operated. The signal generated by secondary chaotic system operating on the receiver is subtracted from the incoming information signal to obtain the data. However, if the secondary chaotic system successfully synchronizes with the primary chaotic system, the information signal is obtained correctly. Therefore, chaotic synchronization has attracted the attention of many researchers and plays a crucial role, especially in chaotic masking methods. The purpose of synchronization is to ensure that the two chaotic systems in both the receiver and transmitter exhibit the same dynamic behavior once they are synchronized [28]. In this study, the synchronization of two distinct IFKC systems with different initial conditions has been realized using the P-TID controller.

3.1. P-TID Controller

The TID controller is one of the fractional-order control structures. In this structure, which combines fractional-order control with the classical PID controller, there is a tilted fraction known as the tilted factor. It is known that with the inclusion of the tilt component, the P-TID controller exhibits better tracking dynamics and is more effective against disturbances [29]. Therefore, the P-TID controller has been preferred in this study. The general function of the P-TID controller in the Laplace domain is provided in Eq. (4) [30].

$$G_C = k_p + k_i s^{-1/n} + k_i s^{-1} + k_d s \quad (4)$$

In the given equation, the controller parameters k_p, k_i, k_i, k_d are respectively the gains of the tilt, integral, and derivative components. In the tilt component given by $s^{-1/n}$, the coefficient n is a real positive number and is usually chosen between 2 and 3.

3.2. Design of error system

After defining the controller, two IFKC systems, each with distinct initial conditions, are established, wherein secondary system requires synchronization with the primary system. The primary and secondary systems are denoted with subscripts 1 and 2, respectively. The primary system is formulated as below:

$$\begin{cases} D^{q_1} x_1(t) = a(y_1 - x_1) + b y_1 z_1^2 \\ D^{q_2} y_1(t) = c x_1 + d x_1 z_1^2 \\ D^{q_3} z_1(t) = h z_1 + k |x_1| \end{cases} \quad (5)$$

Then, the secondary system is specified as follows:

$$\begin{cases} D^{q_1} x_2(t) = a(y_2 - x_2) + b y_2 z_2^2 + u_1 \\ D^{q_2} y_2(t) = c x_2 + d x_2 z_2^2 + u_2 \\ D^{q_3} z_2(t) = h z_2 + k |x_2| + u_3 \end{cases} \quad (6)$$

where u_1, u_2, u_3 are control signals to be designed. Synchronization errors are calculated by subtracting state responses of secondary system from primary system as

$e_1=x_1-x_2, e_2=y_1-y_2$ and $e_3=z_1-z_2$. Consequently, the error system takes the form as defined below:

$$\begin{cases} D^{q_1} e_1(t) = a(e_2 - e_1) + b(y_1 z_1^2 - y_2 z_2^2) - u_1 \\ D^{q_2} e_2(t) = c e_1 + d(x_1 z_1^2 - x_2 z_2^2) - u_2 \\ D^{q_3} e_3(t) = h e_3 + k(|x_1| - |x_2|) - u_3 \end{cases} \quad (7)$$

In this synchronization design, each state is synchronized by distinct P-TID controllers defined by u_1, u_2 and u_3 . The aim is to minimize e_1, e_2 and e_3 , since the controllers are activated so that both systems given in (5) and (6) are synchronized.

3.3. Numerical Simulations

As detailed in dynamic analysis, incommensurate fractional-orders q_1, q_2, q_3 are selected 0.96, 0.97 and 0.98, respectively. The controller parameters are determined by trial-error method and the best tracking dynamics are observed while k_p, k_i, k_i and k_d parameters are chosen as 1.5, 0.1, 0.7, 0.5 respectively. The tilted factor n is set to 0.15. Note that, each P-TID controller is equivalent but takes distinct error signals as input and controller parameters are determined by a trial-and-error method.

To demonstrate the synchronization of the system (7), numerical simulations are carried out using the MATLAB program [31]. Note that, despite Caputos's derivate is employed in dynamical analysis, the synchronization studies use the memory principle of Grünwald-Letnikow fractional-order solver definition code provided in [32] with a fixed step size $t_{step}=5e-3$. Prior to synchronization, the initial conditions for the primary and secondary systems are set as $x_1(0)=0.1, y_1(0)=1, z_1(0)=0.1$ and $x_2(0)=0, y_2(0)=-0.5, z_2(0)=0$, respectively. During the numerical simulations, the P-TID controllers are activated at $t(s)=20$, initiating the synchronization process. The synchronization error functions $e_x(t), e_y(t)$ and $e_z(t)$ are depicted in Figure 5.

As anticipated, the corresponding Figure 5 shows the successful control achieved by the designed P-TID controller over the synchronization of IFKC chaotic system. After the controllers are activated at $t(s)=20$, all of the error functions go to zero which yields effective synchronization as illustrated in Figure 6.

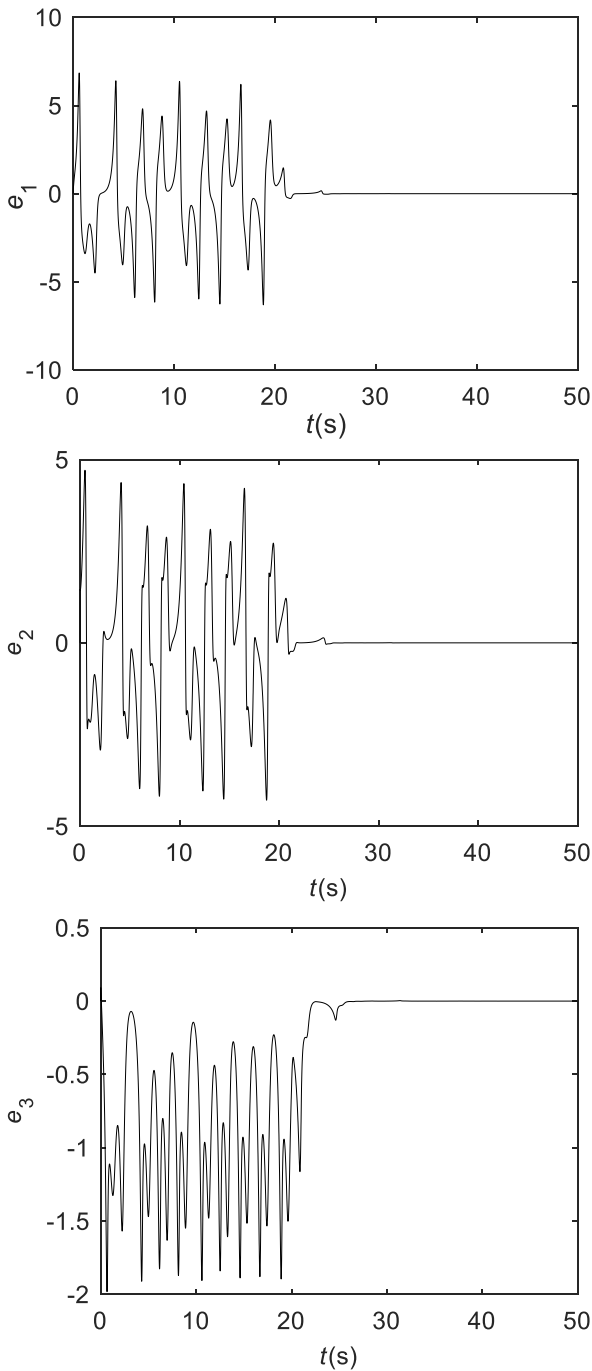


Figure 5. Error functions $e_x(t)$, $e_y(t)$ and $e_z(t)$ of P-TID based synchronization.

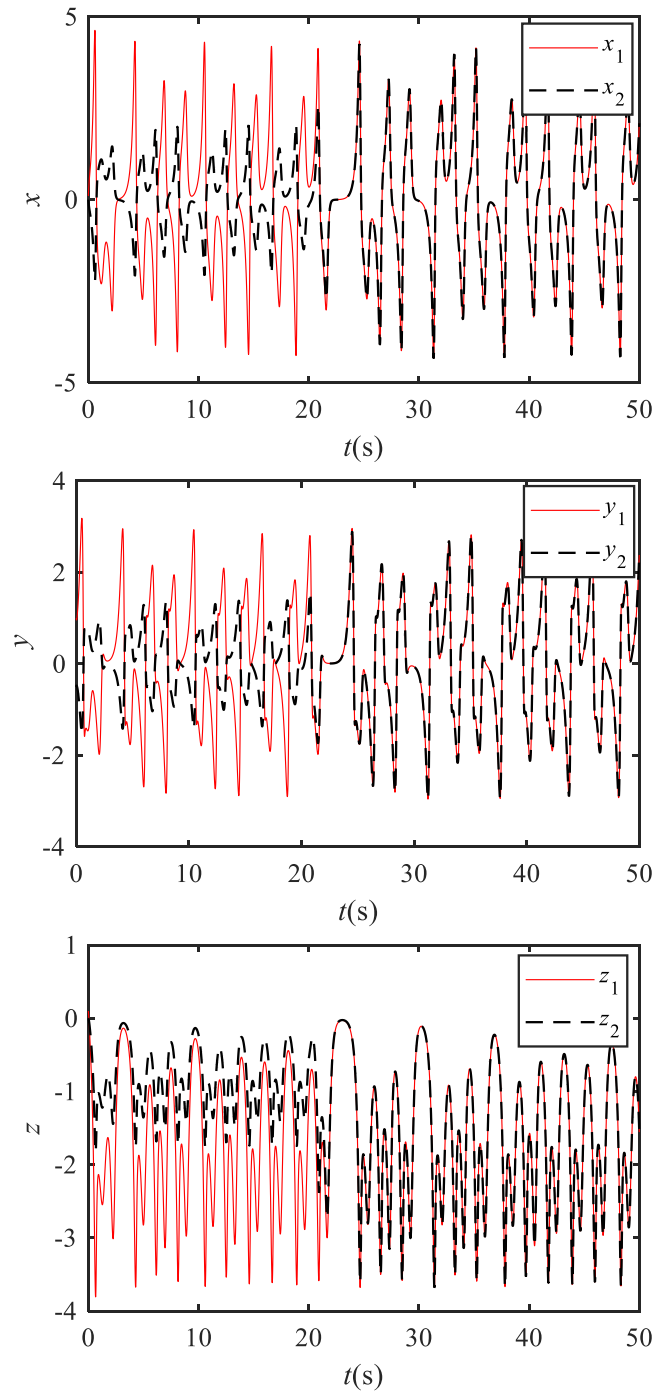


Figure 6. Time series of P-TID-based synchronization.

4. CONCLUSION

In this study, incommensurate fractional-order analysis of King Cobra system is realized. The increased dynamic diversity is contributed by the varying fractional-order values of the presented system. Points of chaos in the system due to parameter changes are identified through bifurcation diagram analyses and Lyapunov spectra. Due to these analyses, more complex chaotic behavior is observed when fractional-orders are chosen as $q_1=0.96$, $q_2=0.97$ and $q_3=0.98$. Later, two IFKC chaotic systems, each initialized with distinct initial conditions, are synchronized. In this context, the secondary system achieves synchronization with the primary system through the utilization of a P-TID controller. The obtained time series of P-TID-based synchronization demonstrate that the established system

can be employed for further secure communication studies. However, controller performance may be improved through future research using optimization methods.

Author contributions: Concept – H.C., A.G.; Data Collection & Processing – H.C., A.G.; Literature Search – H.C., A.G.; Writing – H.C., A.G.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study has received no financial support.

REFERENCES

- [1] M. Demirtas, E. Ilten, and H. Calgan, "Pareto-based multi-objective optimization for fractional order PI^λ speed control of induction motor by using Elman Neural Network," *Arab. J. Sci. Eng.*, vol. 44, no. 3, pp. 2165–2175, 2019.
- [2] Z. Wei, A. Akgul, U. E. Kocamaz, I. Moroz, and W. Zhang, "Control, electronic circuit application and fractional-order analysis of hidden chaotic attractors in the self-exciting homopolar disc dynamo," *Chaos, Solitons & Fractals*, vol. 111, pp. 157–168, 2018.
- [3] E. Ilten and M. Demirtas, "Fractional order super-twisting sliding mode observer for sensorless control of induction motor," *COMPEL - Int. J. Comput. Math. Electr. Electron. Eng.*, vol. 38, no. 2, pp. 878–892, Mar. 2019.
- [4] M. Odabaşı, "Exact analytical solutions of the fractional biological population model, fractional EW and modified EW equations," *An Int. J. Optim. Control Theor. Appl.*, vol. 11, no. 1, pp. 52–58, 2021.
- [5] A. A. Hamou, R. R. Q. Rasul, Z. Hammouch, and N. Özdemir, "Analysis and dynamics of a mathematical model to predict unreported cases of COVID-19 epidemic in Morocco," *Comput. Appl. Math.*, vol. 41, no. 6, p. 289, 2022.
- [6] D. Avcı and F. Soytürk, "Optimal control strategies for a computer network under virus threat," *J. Comput. Appl. Math.*, vol. 419, p. 114740, 2023.
- [7] I. Grigorenko and E. Grigorenko, "Chaotic dynamics of the fractional Lorenz system," *Phys. Rev. Lett.*, vol. 91, no. 3, p. 34101, 2003.
- [8] T. T. Hartley, C. F. Lorenzo, and H. K. Qammer, "Chaos in a fractional order Chua's system," *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.*, vol. 42, no. 8, pp. 485–490, 1995.
- [9] A. Gokyildirim, H. Calgan, and M. Demirtas, "Fractional-Order sliding mode control of a 4D memristive chaotic system," *J. Vib. Control*, p. 10775463231166188, 2023.
- [10] A. Akgul, C. Arslan, and B. Arıcıoğlu, "Design of an interface for random number generators based on integer and fractional order chaotic systems," *Chaos Theory Appl.*, vol. 1, no. 1, pp. 1–18, 2019.
- [11] A. Akgul, Y. Adiyaman, A. Gokyildirim, B. Arıcıoğlu, M. A. Pala, and M. E. Cimen, "Electronic circuit implementations of a fractional-order chaotic system and observing the escape from chaos," *J. Circuits, Syst. Comput.*, vol. 32, no. 05, p. 2350085, Mar. 2023.
- [12] O. Atan, M. Turk, and R. Tuntas, "Fractional order controller design for fractional order chaotic synchronization," *Int. J. Nat. Eng. Sci.*, vol. 7, no. 3, 2013.
- [13] G. Xu, S. Zhao, and Y. Cheng, "Chaotic synchronization based on improved global nonlinear integral sliding mode control☆," *Comput. Electr. Eng.*, vol. 96, p. 107497, 2021.
- [14] F. Motallebzadeh, M. R. J. Motlagh, and Z. R. Cherati, "Synchronization of different-order chaotic systems: Adaptive active vs. optimal control," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 17, no. 9, pp. 3643–3657, 2012.
- [15] C. Ge, C. Hua, and X. Guan, "Master-slave synchronization criteria of Lur'e systems with time-delay feedback control," *Appl. Math. Comput.*, vol. 244, pp. 895–902, 2014.
- [16] S. Çiçek, A. Ferikoglu, and I. Pehlivan, "A new 3D chaotic system: Dynamical analysis, electronic circuit design, active control synchronization and chaotic masking communication application," *Optik (Stuttg.)*, vol. 127, no. 8, pp. 4024–4030, 2016.
- [17] S. Keyong, B. Ruixuan, G. Wang, W. Qiutong, and Z. Yi, "Passive synchronization control for integer-order chaotic systems and fractional-order chaotic systems," in *2019 Chinese Control Conference (CCC)*, 2019, pp. 1115–1119.
- [18] Z.-A. S. A. Rahman, H. A. A. Al-Kashoash, S. M. Ramadhan, and Y. I. A. Al-Yasir, "Adaptive control synchronization of a novel memristive chaotic system for secure communication applications," *Inventions*, vol. 4, no. 2, p. 30, 2019.
- [19] P. Alexander, S. Emiroğlu, S. Kanagaraj, A. Akgul, and K. Rajagopal, "Infinite coexisting attractors in an autonomous hyperchaotic megastable oscillator and linear quadratic regulator-based control and synchronization," *Eur. Phys. J. B*, vol. 96, no. 1, p. 12, 2023.
- [20] A. Gokyildirim, U. E. Kocamaz, Y. Uyaroglu, and H. Calgan, "A novel five-term 3D chaotic system with cubic nonlinearity and its microcontroller-based secure communication implementation," *AEU - Int. J. Electron. Commun.*, vol. 160, p. 154497, Feb. 2023.
- [21] C. Ma, J. Mou, J. Liu, F. Yang, H. Yan, and X. Zhao, "Coexistence of multiple attractors for an incommensurate fractional-order chaotic system," *Eur. Phys. J. Plus*, vol. 135, pp. 1–21, 2020.
- [22] P. Muthukumar, P. Balasubramaniam, and K. Ratnavelu, "Synchronization and an application of a novel fractional order King Cobra chaotic system," *Chaos An Interdiscip. J. Nonlinear Sci.*, vol. 24, no. 3, 2014.

- [23] A. Gokyildirim, "Circuit realization of the fractional-order Sprott K chaotic system with standard components," *Fractal Fract.*, vol. 7, no. 6, p. 470, 2023.
- [24] K. Rajagopal et al., "Multistability and coexisting attractors in a new circulant chaotic system," *Int. J. Bifurc. Chaos*, vol. 29, no. 13, p. 1950174, 2019.
- [25] M. S. Tavazoei and M. Haeri, "A necessary condition for double scroll attractor existence in fractional-order systems," *Phys. Lett. A*, vol. 367, no. 1–2, pp. 102–113, 2007.
- [26] M.-F. Danca, "Matlab code for Lyapunov exponents of fractional-order systems, part ii: The noncommensurate case," *Int. J. Bifurc. Chaos*, vol. 31, no. 12, p. 2150187, 2021.
- [27] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Phys. D Nonlinear Phenom.*, vol. 16, no. 3, pp. 285–317, Jul. 1985.
- [28] S. Emiroglu, A. Akgül, Y. Adıyaman, T. E. Gümüş, Y. Uyaroglu, and M. A. Yalçın, "A new hyperchaotic system from T chaotic system: dynamical analysis, circuit implementation, control and synchronization," *Circuit World*, vol. 48, no. 2, pp. 265–277, 2022.
- [29] H. Calgan and M. Demirtas, "Design and implementation of fault tolerant fractional order controllers for the output power of self-excited induction generator," *Electr. Eng.*, vol. 103, no. 5, pp. 2373–2389, 2021.
- [30] T. Amieur, M. Bechouat, M. Sedraoui, S. Kahla, and H. Guessoum, "A new robust tilt-PID controller based upon an automatic selection of adjustable fractional weights for permanent magnet synchronous motor drive control," *Electr. Eng.*, pp. 1–18, 2021.
- [31] S. Matlab, "Matlab," MathWorks, Natick, MA, 2012.
- [32] H. Li, Y. Shen, Y. Han, J. Dong, and J. Li, "Determining Lyapunov exponents of fractional-order systems: A general method based on memory principle," *Chaos, Solitons & Fractals*, vol. 168, p. 113167, 2023.