ACADEMIC PERSPECTIVE

APJESS

Journal of Engineering and Smart Systems

Volume : 10

Issue : 3

Year : 2022

Volume 10 / Issue 3

Academic Platform Journal of Engineering and Smart Systems

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/72618

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:

ResearchArticles: 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 indi-vidually 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. 203103-1–203103-7, 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 inacceptable. 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 Paper
Academic Platform Journal of Engineering and Smart Systems (APJESS) 10(3), 115-123, 2022
ISSN: 2822-2385
Received: 02-Jul-2022 Accepted: 23-Jul-2022
https://doi.org/10.21541/apjess.1139862

Using Machine Learning Algorithms to Analyze Customer Churn in the Software as a Service (SaaS) Industry

*[1]Levent ÇALLI, [2]Sena KASIM

[1] Department of Management and Organization, Sakarya University, Sakarya, Turkey, lcalli@sakarya.edu.tr
[2] Department of Information Systems Engineering, Sakarya University, Sakarya, Turkey, senakasim@gmail.com

Abstract

Companies must retain their customers and maintain long-term relationships in industries with intense competition. Customer churn analysis is defined in the literature as identifying customers who may leave a company to take appropriate marketing precautions. While customer churn research is prevalent in B2C (Business to Customer) business models such as the telecoms and retail sectors, customer churn analysis in B2B (business to business) models is a relatively emerging topic. In this regard, the study carried out a customer churn analysis by considering an ERP (enterprise resource planning) company with a software as a service (SaaS) business model. Different machine learning algorithms analyzed ten features determined by selection methods and expert opinions. According to the analysis results, the random forest algorithm gave the best result. Additionally, it has been observed that the number of products and customer features has a relatively higher weight for the prediction of churner.

Keywords: Customer Churn, SaaS, Machine Learning, Random Forest, Data Mining

1. INTRODUCTION

In an era where the reduction of costs is an essential factor coupled with intense competitive pressure, organizations must fully maximize their existing customer; therefore, customer retention efforts are implemented to detect decreasing loyalty of the customers with churn analysis [1]. Dissatisfaction, high cost, poor quality, lack of features, privacy concerns, or many different features can be caused the loss of customers. Hence, identifying these features that change according to the industries that decrease loyalty of customers and making reasonable efforts by companies will contribute to a positive change in the situations of customers who are likely to churn [2]. Consequently, loyal customers make more purchases, pay a premium price, and acquire new customers through favorable word-of-mouth, which positively impacts the company's long-term reputation [3]. Any company that wants to survive in business cannot simply ignore the churner and loyalty concept.

Churn analysis has been carried out in different industries in the academic literature. The most intense studies focus on B2C business models such as telecommunication [2], [4]–[10], financial services [11]–[14], and retail [15]–[17] industries. However, few studies have been conducted considering the B2B business model. In B2B markets, fewer customers commonly make larger and more frequent purchases, for instance, compared with the telecommunication or supermarket industries. Hence, customers are more valuable for companies operating the B2B model, and customer retention is central to developing long relationships [18].

In this regard, this study aims to fill the literature gap with a customer churn approach considering the SaaS (software as a service) industry. This research analyzed customer churn using various machine learning algorithms in a cloud ERP company. Furthermore, feature selection techniques were used, and more effective ones for customer churn analysis were identified. The research findings are expected to have academic and practical benefits, particularly in the SaaS field, where few studies exist.

2. LITERATURE REVIEW

2.1. Churn Analysis

Today, our ability to store and analyze all types of data as a result of the information society due to the rapid development of information and communication technologies (ICT) in recent years has resulted in the acquisition of valuable knowledge through data mining

* Corresponding Author

methods [19]. Customer churn analyses using the data collected from the customer are often used today within the framework of the loyalty concept, which is vital for all businesses to achieve long-term relationships with their customers. In this way, companies can identify customers with the potential to leave them using various data mining and machine learning methods to effort the necessary marketing activities for this group called churners.

In the literature, companies using the B2C business model have commonly been analyzed using the customer churn approach. Telecom is the most widely researched industry in this field. For example; Huang et al. [4] indicate that different features and machine learning methods can effectively analyze customer churn. Ahn et al. [7] show that customer transaction and billing data are important factors for customer churn in mobile telecom services. Verbeke et al. [6] present that a small number of datasets can predict churner with high accuracy and Bhattacharyya and Dash [10] discuss the customer churn analysis in the telecom industry from a bibliometric perspective.

Table 1. Academic literature review on churn analysis in SaaS

Study	Methods	Feature Selection	Features and Data Size	Sector	Findings
[20]	• Logistic Regression • Random Forest • XGBoost	No	21 Features 8,256 Observations	No Information	Logistic Regression and XGboost performed relatively well with %72 and %75 AUC scores, respectively. The number of Type-A User Login is the feature with the highest weight.
[21]	• Long short-term memory (LSTM) • Convolutional Neural Network (CNN) • Support Vector Machine • Random Forest	No	5 Feature Categories No data size information	Advertising	Random Forest and Support Vector Machine performed best with 83% and %81 accuracies, respectively. The LSTM and CNN, as the deep learning methods, performed poorly due to a lack of data. As platform usage data, the customer's minutes spent on the platform and the number of active users carries the most weight.
[22]	• Logistic Regression • Support Vector Machine • Decision Tree • Random Forest	Chi Square Anova	23 Features 1788 Observations	Inventory Management	Random Forest performed best with 92% accuracy The number of transactions is the feature with the highest weight.
[23]	• Logistic Regression • Random Forest	No	43 Feature under 4 Categories 8869 Observations Random oversampling and undersampling methods were used.	Cloud-based business phone system and call center	Successful results with little data could not be obtained for both algorithms considering precision and recall ratios. The features with the highest weight are the number of users, number of integrations and call quality.

The customer churn study conducted for the banking industry shows that customers who use more banking services have become more loyal. Customers who use less than three services are the group that needs attention [13]. In the study conducted by Keramati et al. [12] in the banking industry, variables such as the number of mobile, internet, and telephone bank transactions and demographic variables such as age, gender, and educational differences are influential on customer churn prediction.

While customer churn analysis-related studies conducted in the field of B2C are common in the literature, there are few studies on companies using the B2B business model, where the SaaS companies are one of them. The following section discusses the concept of SaaS and research done in this area.

2.2. Software-as-a-Service (SaaS)

SaaS is a business model that offers cloud-based services to clients as an alternative to standalone software that requires installation, maintenance, IT infrastructure, and support services (backup, upgrade, security), especially for B2B [24]. According to Fortune Business Insights (2022)'s report [25] and Jones [27], the global SaaS marketplace is expected to grow from $130 billion in 2021 to $716 billion in 2028, with Microsoft, Salesforce, Adobe, SAP, and Oracle accounting for 51% of the market. Since the market's growth potential will bring more competition, it is vital to determine the factors clients consider in their SaaS preferences to retain and gain new customers. Allen [28] states that a SaaS firm's 3% and 8% monthly customer churn is average. In this respect, monitoring customer churn rates through data mining approaches and intervention to prevent loss are essential for this highly competitive market.

In the academic literature, churn analysis studies tend to focus on the telecommunications, financial services, and retail sectors as B2C business models, while there are just a few studies in the SaaS industry as a B2B business model. These studies are shown in Table 1.

Ge et al. [20] performed churn analysis using logistic regression, random forests, and XGBoost algorithms in their study conducted with the data of a SaaS company, considering 8256 observations and 21 features. They observed that Logistic regression and XGBoost algorithms made relatively good predictions for churner. Additionally, it has been discovered that online usage behaviors, such as login and project numbers, strongly predict customer churn.

In another study conducted in the scope of SaaS, Rautio [21] performed churn analysis with different machine learning methods, considering a company operating in the advertising industry. The research's features mainly focused on client business data, such as spending, platform usage, previous customer service interactions, and service-related customer feedback. In the study, where the support vector machine algorithm gave the best results, it was determined that platform usage metrics feature relatively more weight in prediction. The study also found that deep learning methods did not perform well in predicting churner.

Amornvetchayakul and Phumchusri [22] used four machine learning algorithms in their study with 1778 samples for churn analysis, considering a SaaS company that provides inventory management services (inventory levels, purchase orders, delivery, etc.) for SMEs (Small and Medium Enterprises). The number of transactions in the current month and the number of transactions in the previous month features were discovered as having considerably higher weights than other features in their research, which considered 23 features. The prediction of the random forest algorithm has stated that it has higher accuracy than the decision tree, logistic regression, and support vector machine algorithms.

Lastly, a study focused on churn analysis of a firm in the cloud-based business phone system and call center industry revealed that neither the random forest approach nor logistic regression could achieve sufficient accuracy rates due to strongly overfitting [23]. The number of users, number of integrations, and call quality features were shown to be more critical in predicting churners with the random forest algorithm when random oversampling and undersampling approaches were utilized.

3. METHOD

This research focused on a software company based in Germany and Turkey that specialized mainly in the ERP (enterprise resource planning) industry. A total of 1951 observations were analyzed. Initially, the weight of sixteen features was evaluated, and then the prediction phase that considered churn and non-churn customers was performed using various machine learning methods. The feature selection stage used the Chi-square, information gain, gain ratio, and Gini index methods. Decision Tree, Logistic Regression, Naive Bayes, K-NN, Random Forrest, and Neural Networks algorithms were utilized as classification methods. This process is shown in Figure 1.

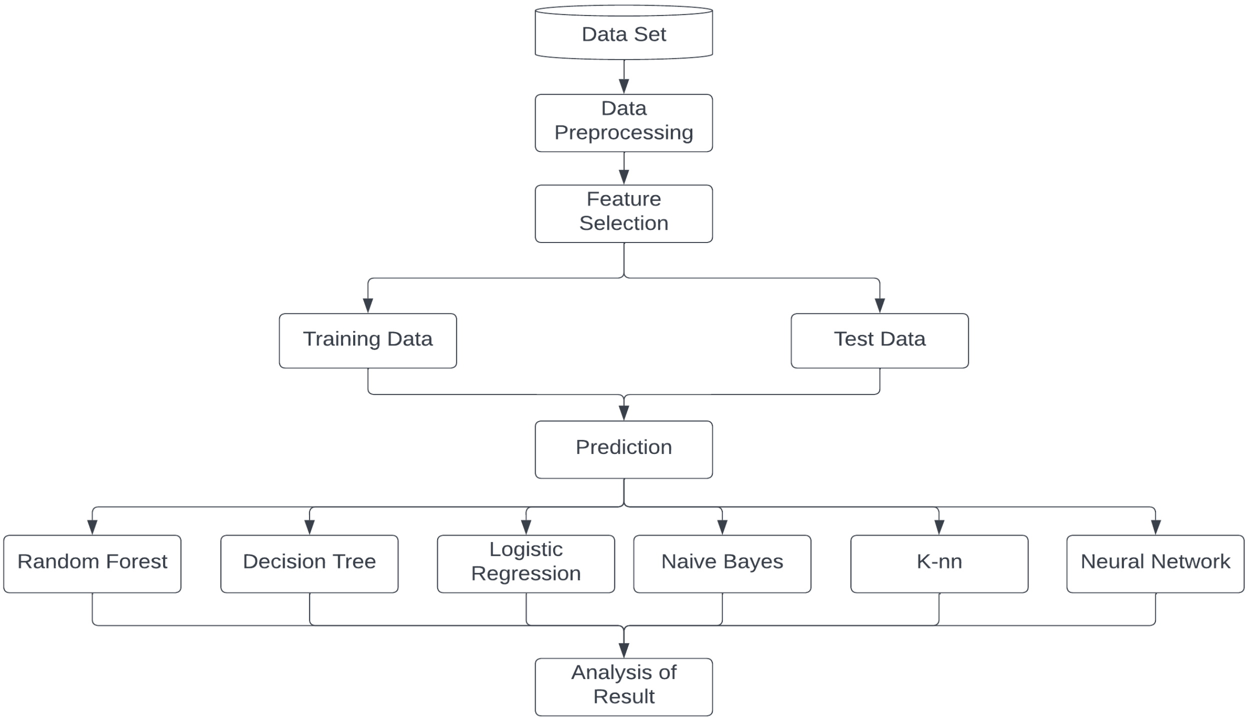


Figure 1. Research process

3.1. Feature Selection

In data mining, feature selection is the preferred technique to reduce dataset size to achieve more efficient analysis and adapt the dataset better to match the preferred analysis method [29]. This study used the chi-square, information gain, gain ratio, and Gini index methods to determine the features to be considered in the churn analysis.

3.1.1. Chi-Square

The Chi-square test, a non-parametric method, is used to examine whether there is a relationship between two categorical variables [30]. The formula for the Chi-square is as follows.

$$x_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

In formula (1), c represents the degrees of freedom, O the observed values, and E the expected value [31].

3.1.2. Information Gain and Gain Ratio

Information gain is used to identify the best features that provide the most information about a class and uses the idea of entropy which is defined as a measure of purity or the degree of uncertainty of a random variable [32]. The information gain calculates the entropy difference before and after the division and determines the purity of the in-class elements. The entropy is calculated using the following formula (2) [33];

$$H(S) = -\sum_{i=1}^{N} p_i \, log_2 \, p_i \qquad (2)$$

S*: Set of all examples in the dataset*
N*: Number of distinct class values*
p$_i$*: Event probability*

The information gain is calculated with the formula (3) shown below [33];

$$Information\ Gain(A,S) = H(S) - \sum_{j=1}^{v} \frac{|S_j|}{|S|} . H(S_j) = H(S) - H(A,S) \qquad (3)$$

H(S)*: Entropy of the whole dataset S*
|S$_j$**|***: Number of the instance with j value of an attribute A*
|S|*: Total number of instances in dataset S*
v*: Set of distinct values of an attribute A*
H(S$_j$**)***: Entropy of subset of instances for attribute A*
H (A,S)*: Entropy of an attribute A*

The gain ratio method is estimated by dividing the information gain value by the entropy value to get accurate results due to the asymmetrical nature of the information gain results [34].

3.1.3. Gini Index

As an impurity splitting method, Gini Index is appropriate for binary and continuous numeric values for calculating feature weight. Assume that S is the collection of s samples with m different classes (C_i, i=1,…m). S can be divided into m subsets based on class differences (S_i, i=1,…m). If S_i is the sample set for class C_i, and S_i is the number of samples in S_i, then the Gini index of S is calculated with the following formula (4) [35];

$$Gini\ Index\ (S) = 1 - \sum_{i=1}^{m} P_i^2 \qquad (4)$$

In this formula, P_i refers to the probability that each given sample belongs to C_i and is measured with s_i/s. When Gini Index has a minimum value of 0, it indicates that all members of the set fall under the same class, indicating that it can gather the most valuable information [35]. IBM SPSS and Orange Data Mining software were utilized for feature selection algorithms [36], [37].

Table 2 shows the ranking of features according to the four feature selection methods. Ten key features were selected for churn analysis within the scope of analysis results and expert opinion.

3.2. Predictive Analysis Algorithms

Churn analysis was performed with Decision Tree, Logistic Regression, Naive Bayes, K-NN, Random Forest, and Neural Networks algorithms which are common approaches in the literature for this study. Each algorithm shortly explains in this section.

Table 2. Feature weight ranking by selection methods

Ranking	Chi-square	Gini Index	Information Gain	Gain Ratio
1	Number of Invoices	The number of customers	The number of customers	Number of Cash Register Connections
2	Number of Offers	Number of Invoices	Number of Invoices	The number of customers
3	Number of Support	Number of products	Number of Offers	Custom Report Usage
4	The number of customers	Number of Offers	Number of Support	Mail Connection
5	Number of products	Number of Support	Number of products	Number of Offers
6	Number of Cash Register Connections	Number of Users	Number of Users	Number of Users
7	Cargo Usage	Cargo Usage	Number of Cash Register Connections	Number of Invoices
8	Number of Payment Documents	Number of Cash Register Connections	Cargo Usage	Number of Support
9	Custom Report Usage	Custom Report Usage	Custom Report Usage	Cargo Usage
10	Mail Connection	Number of Payment Documents	Number of Payment Documents	Number of Payment Documents
11	Number of Cash Register Receipts	Number of Orders	Mail Connection	Number of products
12	Number of Users	Number of Cash Register Receipts	Number of Orders	Number of Cash Register Receipts
13	Number of Orders	Mail Connection	Number of Cash Register Receipts	Number of Production Orders
14	Customer Group	Customer Group	Customer Group	Number of Orders
15	Number of Production Orders	Number of Production Orders	Number of Production Orders	Customer Group
16	Number of Marketplaces	Number of Marketplaces	Number of Marketplaces	Number of Marketplaces

3.2.1. Decision Tree

The decision tree is one of the most popular data categorization techniques in literature, which is based on the information gain theory explained in the previous section. Nodes and branches create the decision tree structure. Nodes represent tests on certain features, while branches indicate test results.

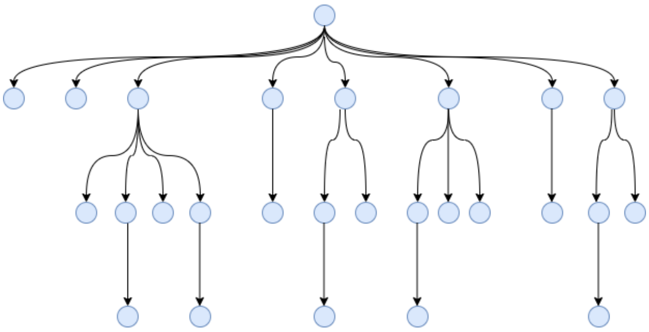


Figure 2. Decision Tree model

Although the decision tree technique has several alternative approaches, the most well-known C5 method splits the sample using the feature with the maximum information gain. It then repeats this process until the subset can no longer be divided [17]. An example decision tree model is shown in Figure 2.

3.2.2. Logistic Regression

Logistic regression is a commonly used statistical model for predicting event probability. In the Logistic regression model seen in detail in formula 5, the dependent variable y is a binary in the formula used to determine whether an event occurred.

$$prob(y=1) = \frac{e^{\beta_0+\sum_{k=1}^{K}\beta_k x_k}}{1-e^{\beta_0+\sum_{k=1}^{K}\beta_k x_k}} \quad (5)$$

The independent inputs are x_1, x_2,....,x_k. The maximum likelihood approach can estimate β_1, $\beta_{2,....,}$ β_k as the regression coefficients based on the available training data [4].

3.2.3. Naïve Bayes

The Naive Bayes algorithm is a well-known classification method used in the machine learning literature. It has attracted much interest due to its ease of use and good performance [38]. The formula (6) is shown below.

$$P(c|x) = \frac{P(c|x)P(c)}{P(x)} \quad (6)$$

Bayes theorem estimates the posterior probability of class given predictor, P(c|x), from P(c) (the prior probability of a class), P(x) (the prior probability of predictor), and P(x|c) (represents the likelihood, which is the probability of predictor class given) [39].

3.2.4. K-NN (The K-Nearest Neighbors Algorithm)

The K-Nearest Neighbors algorithm is an easy-to-implement, simple with few hyperparameters required, supervised learning algorithm that produces classifications or predictions for clustering of a single data point using proximity techniques such as Euclid, Manhattan, Minkowski, and Hamming [40].

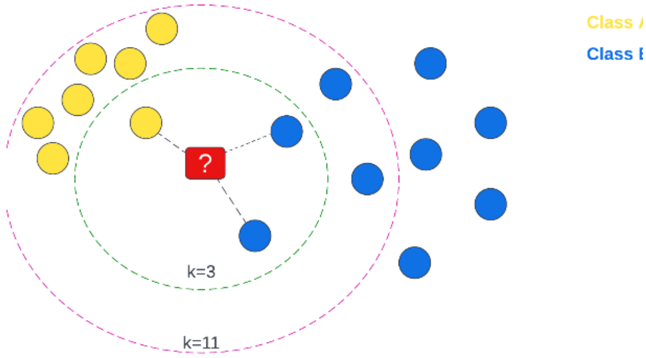


Figure 3. The K-Nearest Neighbors algorithm

K-NN uses most k nearest neighbors for a new data point whose class is being looked for to assign [41]. This situation is seen in Figure 3. For instance, the red shape containing an unclassified data point may belong to either class A or B according to the alternative k values.

3.2.5. Random Forest

The random forest algorithm is applied to a wide range of prediction problems as a well-known approach in the literature due to its capacity to handle small sample sizes and as well as high-dimensional feature spaces, with a few parameters for tuning [42]. The random forest comprises many independent decision trees that act as an ensemble method, and each tree in the random forest generates a class prediction. The model's prediction is based on the class with the most votes, as seen in Figure 4 [43].

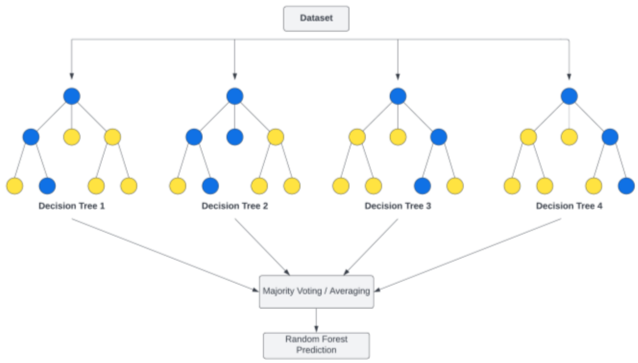


Figure 4. Random Forest

3.2.6. Neural Networks

Neural networks are a method of explaining cognitive, decision-making, and other intelligent control behaviors by using the way the human brain operates as a kind of data processing and analysis [17]. A classic neural network consists of three layers: an input layer, a hidden layer, and an output layer, all connected by neurons, as seen in Figure 5.

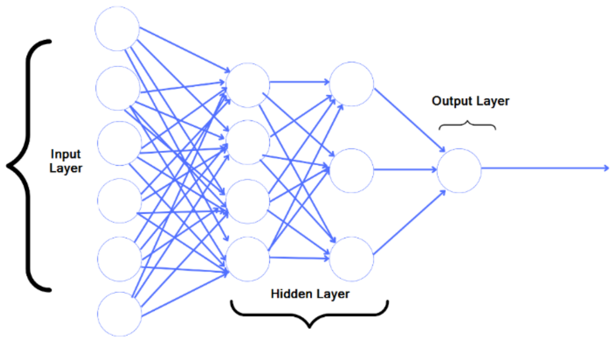


Figure 5. Neural Networks

A key advantage of utilizing neural networks for data modeling is that they can fit complicated nonlinear models without needing to be specified in advance, unlike other nonlinear estimation methods [44].

4. RESULTS

As a result of the feature selection analysis stage and considering the opinions of industry experts, it was determined that ten features would be appropriate for churn analysis. Table 3 shows these features and their descriptions.

Table 3. Selected features

Features	Description
Number of products	The number of products registered in the customer's account.
Number of Customers	The total number of customers and suppliers listed on the customer's account.
Number of Offers	The number of offers that the customer creates in the account.
Number of Orders	The number of orders created by the customer.
Number of Invoices	The number of invoices created by the customer.
Cargo Usage	The number of cargo companies the customer has used in the account.
Number of Users	The number of users the customer has in the ERP software that can access the system at the same time.
Custom Report Usage	The reports are specially made for the customer, excluding the standard reports in the ERP software.
Number of Cash Register Receipts	The number of cash register receipts created by the customer in the ERP software.
Email Connection	Email connection status used for the proposal side of the customer's ERP software.

The churn status of customers was considered in this study based on their active usage of the software. 836 customers in the data set are coded as churn (0), and 1115 are active customers (1). In this regard, it is thought that the dependent variable has a balanced distribution.

Correlation analysis, an essential part of descriptive statistics, was done in Python to reveal the relationships between variables. Figure 6 shows the correlation of features with each other and the customer churn as the dependent variable. A strong positive correlation is shown by dark colors, whereas light tones indicate a weak positive correlation.

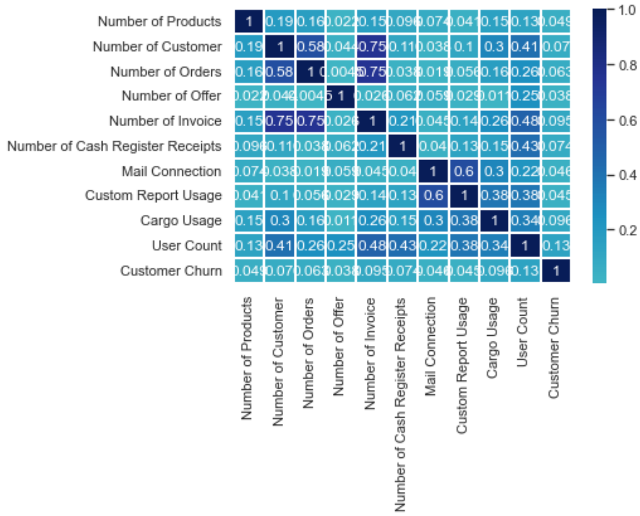


Figure 6. Correlation heatmap

The normalization process was conducted in the next step by assigning values between 0 and 1 due to the features' value variations. This method allows using a single scale while keeping the distinctions in the value ranges and avoiding information loss of each feature.

Table 4. Number of training and test data

	Training Dataset	Test Dataset
%75-%25	1463	488
%70-%30	1365	586

In the next stage, the dataset was divided into test and training data. Then, the model was tested with different algorithms. This study used two different ratios of training data, 70% and 75%. The number of customers that create the training and test data sets is shown in Table 4.

Table 5. Churn Analysis results

Predictive Algorithms	%75-%25 Accuracy	%70-%30 Accuracy
Decision Tree	74.59%	73.54%
Logistic Regression	57.58%	57.84%
Naive Bayes	47.95%	48.12%
K-NN	65.36%	74.23%
Random Forest	78.27%	77.47%
Neural Network	57.99%	58.19%

As seen in Table 5, the churn analysis results with the training and test data in both ratios show that the random forest algorithm gives the relatively best results with higher accuracy rates. The results also show that the decision tree algorithm produces better results than other algorithms. Additionally, the K-NN method provides much better accuracy when the number of training data sets is decreased. The weights of the features used in the predictive algorithms for the random forest and decision tree are shown in Table 6.

Table 6. Features weight

Decision Tree		Random Forest	
Feature	**Weight**	**Feature**	**Weight**
Number of Customer	45,0%	Number of Customer	42,4%
Number of Products	29,6%	Number of Product	33,3%
Number of Invoice	15,4%	Number of Invoice	9,4%
Number of Orders	4,0%	Number of Orders	5,4%
Number of Users	2,2%	Number of Offer	3,1%
Cargo Usage	1,7%	Cargo Usage	2,8%
Number of Offer	1,3%	Number of Users	2,8%
Custom Report Usage	0,5%	Number of Cash Register Receipts	0,3%
Mail connection	0,2%	Mail Connection	0,3%
Number of Cash Register Receipts	0,0001%	Custom Report Usage	0,3%

In this respect, it is seen that the number of customers and products has a high weight in both algorithms predicting the customer's churn status. Mail connection, custom report usage, number of cash register receipts, cargo usage, number of the offer, number of users, and number of orders have low weight, while the number of invoice feature has a relatively moderate weight in both algorithms.

5. CONCLUSION

Customer churn is a severe challenge to any business, and one way to deal with it is to predict which customers are most likely to leave the company and then target those segments with marketing efforts to encourage them to be loyal [45]. Customer loyalty is valuable since acquiring new customers is more costly than keeping existing customers, and a loyal customer serves as an honorary lawyer for the company, becoming a positive reference for potential customers and, as a result, generating more profit [19]. While academic studies on customer churn analysis are popular in a few industries, especially in B2C fields such as telecommunications, grocery retail, or bank, research on software as a service (SaaS) as a B2B industry is very limited. Thus, the study aimed to fill a gap in the academic literature by focusing on a cloud ERP business.

According to the research findings, it is seen that the number of customers and the number of products are the most important features in customer churn analysis. The number of invoices and orders features also shows a relatively higher weight than others. Thus, research findings reveal that features of the fundamental business process have a meaningful relationship with customer churn and are factors that need more attention for managers. In this sense, the results show some parallelism with Amornvetchayakul and Phumchusri [22] and Sergue [23]. For example, Amornvetchayakul and Phumchusri [22] indicate that the importance of the number of transactions in the inventory management sector is relatively higher than in others. Additionally, Sergue [23] highlights that the number of users in the Cloud-based business phone system and call center industry is relatively higher weight than other features in predicting churner. In terms of prediction accuracy, the random forest algorithm achieves the best result in this study, which is similar to the results observed by Rautio [21] and Amornvetchayakul and Phumchusri [22].

As a result, the churn analysis processes of B2B business models, which have less data for many businesses compared to B2C business models, undoubtedly involve many difficulties. In this study, the customer churn analysis process was carried out for a cloud ERP company and contributed to the academic literature and the practical field. Although it is an important gap that there are very few studies in this literature, it is thought that new studies will emerge in the light of the findings of this study. Using different features and testing different predictive algorithms will enrich the academic literature and be guiding for managers.

Author Contributions: Concept – L.C., S.K.; Data Collection &/or Processing – S.K; Literature Search – L.C., S.K.; Writing – L.C.
Conflict of Interest: This study was produced from the master's thesis entitled "Customer Churn Analysis with Data Mining Methods: Software Industry," conducted at Sakarya University, Institute of Natural Sciences, with the supervisor Levent ÇALLI.
Financial Disclosure: The authors declared that this study has received no financial support.

REFERENCES

[1] N. Glady, B. Baesens, and C. Croux, "Modeling churn using customer lifetime value," Eur. J. Oper. Res., vol. 197, no. 1, pp. 402–411, 2009, doi: 10.1016/j.ejor.2008.06.027.

[2] A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo, and S. Anwar, "Customer churn prediction in telecommunication industry using data certainty," J. Bus. Res., vol. 94, no. February 2018, pp. 290–301, 2019, doi: 10.1016/j.jbusres.2018.03.003.

[3] J. Ganesh, M. J. Arnold, and K. E. Reynolds, "Understanding the customer base of service providers: An examination of the differences between switchers and stayers," J. Mark., vol. 64, no. 3, pp. 65–87, 2000, doi: 10.1509/jmkg.64.3.65.18028.

[4] B. Huang, M. T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications," Expert Syst. Appl., vol. 39, no. 1, pp. 1414–1425, 2012, doi: 10.1016/j.eswa.2011.08.024.

[5] K. Kim, C. H. Jun, and J. Lee, "Improved churn prediction in telecommunication industry by analyzing a large network," Expert Syst. Appl., vol. 41, no. 15, pp. 6575–6584, 2014, doi: 10.1016/j.eswa.2014.05.014.

[6] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," Eur. J. Oper. Res., vol. 218, no. 1, pp. 211–229, 2012, doi: 10.1016/j.ejor.2011.09.031.

[7] J. H. Ahn, S. P. Han, and Y. S. Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry," Telecomm. Policy, vol. 30, no. 10–11, pp. 552–568, 2006, doi: 10.1016/j.telpol.2006.09.006.

[8] C. F. Tsai and Y. H. Lu, "Customer churn prediction by hybrid neural networks," Expert Syst. Appl., vol. 36, no. 10, pp. 12547–12553, 2009, doi: 10.1016/j.eswa.2009.05.032.

[9] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," Simul. Model. Pract. Theory, vol. 55, pp. 1–9, 2015, doi: 10.1016/j.simpat.2015.03.003.

[10] J. Bhattacharyya and M. K. Dash, "What Do We Know About Customer Churn Behaviour in the Telecommunication Industry? A Bibliometric Analysis of Research Trends, 1985–2019," FIIB Bus. Rev., 2021, doi: 10.1177/23197145211062687.

[11] Y. Xie, X. Li, E. W. T. Ngai, and W. Ying, "Customer churn prediction using improved balanced random forests," Expert Syst. Appl., vol. 36, no. 3 PART 1, pp. 5445–5449, 2009, doi: 10.1016/j.eswa.2008.06.121.

[12] A. Keramati, H. Ghaneei, and S. M. Mirmohammadi, "Developing a prediction model for customer churn from electronic banking services using data mining," Financ. Innov., vol. 2, no. 1, 2016, doi: 10.1186/s40854-016-0029-6.

[13] A. Bilal Zoric, "Predicting Customer Churn in Banking Industry using Neural Networks," Interdiscip. Descr. Complex Syst., vol. 14, no. 2, pp. 116–124, 2016, doi: 10.7906/indecs.14.2.1.

[14] J. Burez and D. Van den Poel, "Handling class imbalance in customer churn prediction," Expert Syst. Appl., vol. 36, no. 3 PART 1, pp. 4626–4636, 2009, doi: 10.1016/j.eswa.2008.05.027.

[15] M. A. de la Llave Montiel and F. López, "Spatial models for online retail churn: Evidence from an online grocery delivery service in Madrid," Pap. Reg. Sci., vol. 99, no. 6, pp. 1643–1665, 2020, doi: 10.1111/pirs.12552.

[16] W. Buckinx and D. Van Den Poel, "Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting," Eur. J. Oper. Res., vol. 164, no. 1, pp. 252–268, 2005, doi: 10.1016/j.ejor.2003.12.010.

[17] X. Hu, Y. Yang, L. Chen, and S. Zhu, "Research on a Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network," 2020 IEEE 5th Int. Conf. Cloud Comput. Big Data Anal. ICCCBDA 2020, pp. 129–132, 2020, doi: 10.1109/ICCCBDA49378.2020.9095611.

[18] B. Janssens, M. Bogaert, A. Bagué, and D. Van den Poel, "B2Boost: instance-dependent profit-driven modelling of B2B churn," Ann. Oper. Res., 2022, doi: 10.1007/s10479-022-04631-5.

[19] W. Verbeke, D. Martens, C. Mues, and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," Expert Syst. Appl., vol. 38, no. 3, pp. 2354–2364, 2011, doi: 10.1016/j.eswa.2010.08.023.

[20] Y. Ge, S. He, J. Xiong, and D. E. Brown, "Customer churn analysis for a software-as-a-service company," in 2017 Systems and Information Engineering Design Symposium, SIEDS 2017, 2017, pp. 106–111, doi: 10.1109/SIEDS.2017.7937698.

[21] A. Rautio, "Churn rediction in SaaS using Machine Learning," 2019.

[22] P. Amornvetchayakul and N. Phumchusri, "Customer Churn Prediction for a Software-as-a-Service Inventory Management Software Company: A Case Study in Thailand," in 2020 IEEE 7th International Conference on Industrial Engineering and Applications, ICIEA 2020, 2020, pp. 514–518, doi: 10.1109/ICIEA49774.2020.9102099.

[23] M. Sergue, "Customer Churn Analysis and Prediction using Machine Learning for a B2B SaaS company," 2020, [Online]. Available: www.kth.se/sci.

[24] D. Ma, "The Business Model of 'Software-As-A-Service,'" in IEEE International Conference on Services Computing (SCC 2007), 2007, no. July, pp. 701–702, doi: 10.1109/SCC.2007.118.

[25] Fortunebusinessinsights, "The software as a service market Size," 2022. https://www.fortunebusinessinsights.com/software-as-a-service-saas-market-102222.

[26] E. Jones, "Cloud Market Share – a Look at the Cloud Ecosystem in 2022," KINSTA BLOG, 2022. https://kinsta.com/blog/cloud-market-share/#:~:text=The SaaS market is dominated,impressive annual growth of 34%25.

[27] E. Jones, "Cloud Market Share – a Look at the Cloud Ecosystem in 2022," KINSTA BLOG, 2022. .

[28] K. Allen, "Churn Rate vs Retention Rate: How to Calculate These SaaS KPIs," woopra.com, 2022. https://www.woopra.com/blog/churn-rate-vs-retention-rate.

[29] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," 2015 38th Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2015 - Proc., no. May, pp. 1200–1205, 2015, doi: 10.1109/MIPRO.2015.7160458.

[30] A. Field, Discovering Statistics Using IBM SPSS Statistics. SAGE Publications Ltd, 2018.

[31] sampath kumar Gajawada, "Chi-Square Test for Feature Selection in Machine learning,"

https://towardsdatascience.com/, 2019. https://towardsdatascience.com/chi-square-test-for-feature-selection-in-machine-learning-206b1f0b8223.

[32] N. Tyagi, "What is Information Gain and Gini Index in Decision Trees?," https://www.analyticssteps.com/, 2021. https://www.analyticssteps.com/blogs/what-gini-index-and-information-gain-decision-trees.

[33] U. Krčadinac, "Classification – Decision Trees," 2015. http://ai.fon.bg.ac.rs/wp-content/uploads/2015/04/Classification-Decision-Trees-2015.pdf.

[34] S. K. Trivedi, "A study on credit scoring modeling with different feature selection and machine learning approaches," Technol. Soc., vol. 63, no. September, p. 101413, 2020, doi: 10.1016/j.techsoc.2020.101413.

[35] W. Shang, H. Huang, H. Zhu, Y. Lin, Y. Qu, and Z. Wang, "A novel feature selection algorithm for text categorization," Expert Syst. Appl., vol. 33, no. 1, pp. 1–5, 2007, doi: 10.1016/j.eswa.2006.04.001.

[36] J. Demšar et al., "Orange: Data Mining Toolbox in Python," J. Mach. Learn. Res., vol. 14, pp. 2349–2353, 2013, [Online]. Available: http://jmlr.org/papers/v14/demsar13a.html.

[37] IBM Corp., "IBM SPSS Statistics for Windows, Version 26.0," 2019. 2019.

[38] V. V. Saradhi and G. K. Palshikar, "Employee churn prediction," Expert Syst. Appl., vol. 38, no. 3, pp. 1999–2006, 2011, doi: 10.1016/j.eswa.2010.07.134.

[39] S. Sayad, "An Introduction to Data Science," 2022. saedsayad.com/data_mining_map.htm.

[40] IBM, "What is the k-nearest neighbors algorithm?," 2022. https://www.ibm.com/topics/knn.

[41] O. Kramer, Dimensionality Reduction with Unsupervised Nearest Neighbors, vol. 51. 2013.

[42] G. Biau and E. Scornet, "A random forest guided tour," Test, vol. 25, no. 2, pp. 197–227, 2016, doi: 10.1007/s11749-016-0481-7.

[43] T. Yiu, "Understanding Random Forest," https://towardsdatascience.com/, 2019. https://towardsdatascience.com/understanding-random-forest-58381e0602d2.

[44] D. J. Livingstone, D. T. Manallack, and I. V. Tetko, "Data modelling with neural networks: Advantages and limitations," J. Comput. Aided. Mol. Des., vol. 11, no. 2, pp. 135–142, 1997, doi: 10.1023/A:1008074223811.

[45] S. A. Neslin, S. Gupta, W. Kamakura, L. U. Junxiang, and C. H. Mason, "Defection detection: Measuring and understanding the predictive accuracy of customer churn models," J. Mark. Res., vol. 43, no. 2, pp. 204–211, 2006, doi: 10.1509/jmkr.43.2.204.

Research Paper
Academic Platform Journal of Engineering and Smart Systems (APJESS) 10(3), 124-139, 2022
ISSN: 2822-2385
Received: 31-Jan-2022 Accepted: 27-Jul-2022
https://doi.org/10.21541/apjess.1065912

Hierarchical Approaches to Solve Optimization Problems

*[1]Ferda Nur ARICI, [2]Ersin KAYA

[1]Department Of Computer Engineering, Necmettin Erbakan University, Konya, Turkey, fnarici@erbakan.edu.tr
[2] Department Of Computer Engineering, Konya Technical University, Konya, Turkey, ekaya@ktun.edu.tr

Abstract

Optimization is the operation of finding the most appropriate solution for a particular problem or set of problems. In the literature, there are many population-based optimization algorithms for solving optimization problems. Each of these algorithms has different characteristics. Although optimization algorithms give optimum results on some problems, they become insufficient to give optimum results as the problem gets harder and more complex. Many studies have been carried out to improve optimization algorithms to overcome these difficulties in recent years. In this study, six well-known population-based optimization algorithms (artificial algae algorithm - AAA, artificial bee colony algorithm - ABC, differential evolution algorithm - DE, genetic algorithm - GA, gravitational search algorithm - GSA, and particle swarm optimization - PSO) were used. Each of these algorithms has its own advantages and disadvantages. These population-based six algorithms were tested on CEC'17 test functions and their performances were examined and so the characteristics of the algorithms were determined. Based on these results, hierarchical approaches have been proposed in order to combine the advantages of algorithms and achieve better results. The hierarchical approach refers to the successful operation of algorithms. In this study, eight approaches were proposed, and performance evaluations of these structures were made on CEC'17 test functions. When the experimental results are examined, it is concluded that some hierarchical approaches can be applied, and some hierarchical approaches surpass the base states of the algorithms.

Keywords: Optimization; Population-based Algorithm; CEC'17; Hierarchical Approaches

1. INTRODUCTION

Optimization is the operation of searching and identifying the most appropriate solution for a specific problem or set of problems. Optimization is used in almost every field, from engineering to medical, from agriculture to economics. The aim of optimization is to use the resources to a minimum, and the earnings are maximum. For the present problem, it is very critical to determine and use the most appropriate optimization method in line with this purpose [1].

The algorithms that solve optimization problems are called optimization algorithms. These algorithms are examined under two headings: deterministic and stochastic. Deterministic algorithms are algorithms that always follow the same path when the same starting points are given. Due to the disadvantages of these algorithms, such as being problem-dependent, locating them at local points in the problem, having high computational costs for large and difficult problems, researchers have turned to stochastic algorithms for the solution of optimization problems [2]. Stochastic algorithms are relied on randomness. Because these algorithms are randomness, they do not always guarantee to find the optimum, but they give close to optimum quality results with an acceptable calculation cost [1].

Stochastic algorithms are examined under two headings as heuristic and meta-heuristic. Heuristic algorithms use trial and error to find reasonable solutions for complex problems within an acceptable period of time [3]. The metaheuristic is a superior strategy that is more general than heuristics, which can easily be applied to unlike optimization problems, aiming to combine basic heuristic methods that will enable a more comprehensive investigation of the solution space [4].

Metaheuristic algorithms are classified as nature-inspired or evolutionary. Nature-inspired algorithms are one of the most studied areas in recent years. These algorithms are developed by inspiring the life of living creatures in nature. An example of these algorithms is the artificial bee colony algorithm [5]. The artificial bee colony algorithm is an algorithm inspired

* Corresponding Author

by the life behavior of bees. Artificial algae algorithm [3], particle swarm optimization algorithm [6], ant colony algorithm [7] are examples of these algorithms. Evolutionary algorithms are developed rely on the theory of evolution in biology. Genetic algorithm [8] and differential evolution algorithms [9] are examples of evolutionary algorithms.

Population-based algorithms are also used in metaheuristic algorithms. In these algorithms, the solution set is called the population [2]. In this study, six famous population-based optimization algorithms (artificial algae algorithm - AAA, artificial bee colony algorithm - ABC, differential evolution algorithm - DE, genetic algorithm - GA, gravitational search algorithm -GSA, and particle swarm optimization - PSO) were used. Each of these algorithms has its own parameters. Changing these parameters creates differences on the local and global search abilities of the algorithm. For example, one of the control parameters of the AAA is the energy loss parameter. If this parameter is low, the local search ability of the algorithm increases, and there is a risk of getting stuck to local minimums. If the energy loss parameter is high, the algorithm's global search capability increases, and thus the rate of convergence to the global best decreases [3]. Determining the most suitable values of the parameters in algorithms is an important problem.

It is necessary to well know the characteristics of the algorithms used to propose a hierarchical approach. For this purpose, a previous study was conducted to determine the characteristics of the algorithms used in this study and to compare them with each other. These six algorithms used in the previous study were tested on CEC'17 test functions [10]. Then the performances of these algorithms were examined, and their characteristics were determined. Based on these results, hierarchical approaches have been proposed in order to obtain better results than the individual results of the algorithms. In this context, eight approaches have been proposed. The performance evaluations of these structures were also made on CEC’17 test functions. When the results were examined, it was concluded that some hierarchical approaches were applicable.

The organization of the article is as follows: in the second section, related works are given. In section 3, the algorithms used in the study are introduced. In section 4, the proposed hierarchical approaches are described. In section 5, experimental setup and experimental results are given. In section 6, the results of this study were evaluated, and suggestions for future studies were given.

1.1. Main Contribution of the Study

Optimization algorithms have their own strengths and weaknesses. While some algorithms have a strong exploitation feature, some algorithms have a strong exploration feature. The aim of this study is to merge an algorithm with a strong exploitation feature and an algorithm with a strong exploration feature to obtain a stronger algorithm. For this purpose, eight different hierarchical approaches from algorithms with strong exploitation or exploration feature are proposed. As far as we know, it is the first study that presents the largest number of hierarchical approaches.

2. RELATED WORK

Metaheuristic algorithms are generally used today to resolve optimization problems. Although these algorithms are used very often, they have some disadvantages such as being stuck in local points, not being able to balance between exploration and exploitation as the problem to be solved becomes difficult and the solution space expands [11]. In order to get rid of these disadvantages and to achieve more successful results, studies have been continuing to improve the optimization algorithms in recent years. The hierarchical approach is one of these studies.

Two hierarchical approaches as PSO-GA and GA-PSO were used in antenna optimization using GA and PSO algorithms by Robinson et al. It has been observed that the PSO-GA hierarchical approach surpasses the GA-PSO hierarchical approach and the singular states of GA, PSO algorithms [12].

Bellatreche et al. (2006) [13] used the hierarchical use of GA and simulation annealing (SA) algorithms for physical data warehouse design. It was implemented as GA-SA. Experimental studies have shown that the proposed approach is feasible.

The combined GA-PSO approach was used for optimum location and sizing of distributed generation resources by Moradi and Abedini. According to the results, it has been shown that the GA-PSO approach passes GA and PSO algorithms individually and finds the most suitable solution for the system [14].

Arunachalam et al. (2014) [15] have proposed the PSO-Firefly Algorithm (FA) hierarchical approach, which they call HPSOFF, for the Combined Economic and Emission Dispatch (CEED) problem. In their proposed hierarchical approach, FA uses the population generated from the PSO algorithm. Half of the maximum number of function evaluations (MaxFES) uses the PSO and the other half the FA algorithm. In the experimental results, they saw that the method they suggested gave more successful results.

The ant colony (ACO) algorithm and the ABC algorithm are hierarchically used as ACO-ABC for the resolution of optimization problems by Kıran. As a result of experimental studies, it has been demonstrated that the ACO-ABC hierarchical approach shows superior performance compared to ACO and ABC algorithms [2].

The hierarchical approach of ACO and PSO algorithms has been applied to the traveling salesman problem by Eldem. According to the results, it has been observed that the hierarchical approach performs better than the standard ACO and standard PSO algorithm [16].

ABC, DE, and PSO algorithms were used hierarchically to solve optimization problems by Gökalp and Uğur. It was determined that the ABC-DE-PSO hierarchical structure among the combinations of these three algorithms was successful. Later, this structure was compared with the singular states of the algorithms on the test functions and it was revealed that the proposed hierarchical structure showed superior success [17].

Gaidhane and Nigam (2018) [18] have proposed the Grey Wolf Optimizer (GWO)-ABC hierarchical approach. In addition, they have proposed a new method rely on chaotic mapping and opposition-based learning to begin the population. In their paper, to evaluate the performance of the GWO-ABC, they tested on 27 synthesis comparison functions with distinct features; and they compared the results with 5 algorithms. According to the results, they showed that his overall performance improved.

Lin et al. (2018) [19] have proposed the PSO-DE hierarchical approach for numerical benchmark problems and optimization of active disturbance rejection controller (ADRC) parameters. In their study, they used the DE algorithm to get rid of PSO stagnation. They tested it on CEC2005 and CEC2011 test functions to assess the performance of the proposed algorithm. In the results, they presented that the proposed approach performed well.

Jiang et al. (2020) [20] have presented the PSO-GSA hierarchical approach, which they called HPSO-GSA. In their proposed approach, the PSO algorithm begins. Then, when the swarm reaches the optimum level and cannot develop fitness, it is included in the GSA PSO. In order to assess the effectiveness of their proposed HPSO-GSA, simulations were carried out on comparison test functions. In their results, they showed that HPSO-GSA outperformed PSO, GSA, and other recently developed hybrid variants.

Sharma and Ghosh (2020) [21] proposed a hybrid ABC-PSO algorithm to determine the optimum capacitor size. The ABC algorithm consists of three stages: worker bees, onlooker bees and scout bees. In this work, each phase updates the capacitors size and the worst solution results at the end of the last phase are optimized using the PSO algorithm. In this study, they designed and used the ABC-PSO algorithm to minimize the total power loss and energy loss.

Karakoyun et al. (2020) [22] proposed a new algorithm based on gray wolf optimizer (GWO) and shuffled frog leaping (SFLA) algorithms for multi-objective optimization problems. They named it MOSG. They applied some modifications to improve the performance of the proposed algorithm and to use the GWO algorithm effectively. The results show that the proposed algorithm gives successful results in solving multi-objective problems.

Dixit et al. (2021) [23] proposed a hybrid DE-PSO-based COVID-19 prediction model using support vector machine (SVM) from chest X-ray images. They initially performed feature extraction and data preprocessing steps. Then, the selected features were optimized with the hybrid DE-PSO algorithm as a new feature optimization approach. These optimized attributes are sent to the SVM. In the results, they showed that their proposed model achieved 99.34% accuracy.

Kisengeu et al. (2021) [24] proposed the hybrid ABC-PSO algorithm. With the algorithm they proposed in this study, they made under-voltage load shedding (UVLS) to increase voltage stability. They solved the UVLS problem by incorporating the PSO's velocity and position finding capability into the ABC algorithm. They compared their proposed hybrid ABC-PSO algorithm with GA, ABC-ANN (artificial neural networks), PSO-ANN. As a result, they show that the proposed algorithm throws an optimal amount of load.

Parouha and Verma (2021) [25] proposed a new hybrid algorithm based on PSO-DE for unlimited optimization problems and applications. They named this algorithm ihPSODE. They have integrated the ihPSODE algorithm with the proposed new PSO (nPSO) and DE (nDE). In ihPSODE, they initially ranked the entire population by fitness function value and applied nPSO to half. They then administered nDE to the offspring generated by nPSO. Then they combined the two populations. They demonstrated the superiority of the proposed algorithms in the results.

Zhang et al. (2022) [26] proposed a new GSA-PSO algorithm for multi-purpose load distribution for microgrid with electric vehicles. They named it MGSA-PSO. They integrated the global memory capacity of the PSO into the GSA to improve the global search performance of the GSA in the MGSA-PSO structure. In the results, they demonstrated the success of the proposed MGSA-PSO algorithm by analyzing it on different numbers of electric vehicles.

Li et al. (2022) [27] proposed ABC-PSO algorithm for mobile robot path planning. They added the search process and optimization process in the ABC algorithm to the optimized PSO algorithm. Thus, they developed ABC-PSO algorithm, which has good global search capabilities and fast convergence. They show in the results that the proposed algorithm finds the optimal path of the robot quickly and efficiently, with short search time and less iteration.

3. MATERIALS AND METHODS

In this section, metaheuristic algorithms used in the study are defined.

3.1. Artificial Algae Algorithm – AAA

Artificial algae algorithm (AAA) is an optimization algorithm presented in 2015 rely on the characteristics and life behavior of moving micro-algae. AAA consists of three main stages. These; evolutional process, helical movement process, and adaptation process. The helical movement process is based on the helical movements of algae in the liquid and their attitude towards approaching the light. The evolutionary process is based on the proliferation of algae by mitosis. The adaptation process is based on the adaptation of the algae to their environment. In the algorithm, algae is the main component, and all population contain of algae colonies. The number of algae cells in each algae colony is equal to the problem size. Thus, each solution in the resolution space corresponds to an artificial algae colony [3]. Due to its success in many problems, AAA is used in many areas such as feature selection [28], clustering analysis [29], uncapacitated facility location problems [30].

3.2. Artificial Bee Colony – ABC

The artificial bee colony (ABC) algorithm is a population-based optimization algorithm developed in 2005 by modeling the intelligent behavior of bees with swarm intelligence during the food search process. There are two types of bees in the artificial bee colony. The first type of bees is employed bees. Other types of bees are unemployed bees. Onlooker bees are unemployed bees. The ABC algorithm makes some assumptions. The first is that only one bee receives the nectar of each resource. Thus, the number of employed bees equals the total number of food sources. Another assumption is that the number of employed bees is the same to the number of onlooker bees [2, 5].

3.3. Differential Evolution Algorithm – DE

Differential evolution algorithm was put forward by Price and Storn in 1995. Differential evolution algorithm is one of population-based optimization algorithms rely on genetic algorithms in general [9]. Crossover, mutation, and natural selection operators in GA are also included in DE. In DE, chromosomes are handled one by one, and a new individual is formed using three randomly selected chromosomes. These operations are performed with mutation and crossover operators [31, 32].

3.4. Genetic Algorithm – GA

Genetic algorithms are evolutionary algorithms that optimize optimization problems modeled by biological processes [8]. Genetic algorithms are optimization methods based on natural selection principles. The algorithm was set up by John Holland. Later, many studies on genetic algorithms were published [33]. GA parameters represent genes. The aggregate set of parameters constitutes the chromosome. Each chromosome represents a solution [8].

In the algorithm, firstly, the initial population is randomly generated, and the suitability values of this population are calculated. Then, with the natural selection process, crossover and mutation are used to produce solutions in the next generation. The fitness evaluation process is applied to each individual to perform the selection process. A simple genetic algorithm consists of five basic components. These; representations of solutions, the method of forming the initial population, the fitness evaluation function, using the genetic operators and control parameters [32].

3.5. Gravitational Search Algorithm – GSA

The gravitational search algorithm (GSA) is an optimization algorithm presented in 2009 inspired by Newton's laws of gravity and motion. GSA tries to find the optimal solution according to Newton's laws of gravity and motion by using a series of agents called masses. Each possible solution corresponds to an agent in the GSA. The mass of each agent is represented by its fitness value. According to the fitness function, the best and worst agent of the population is detected and used in the algorithm [34].

3.6. Particle Swarm Optimization – PSO

Particle Swarm Optimization (PSO) is an optimization algorithm improved in 1995 inspired by fish and birds traveling in the swarm [6]. The algorithm is basically based on swarm intelligence. Social information sharing among individuals is important in PSO. In the algorithm, each individual is called a particle. The population formed by the combination of these particles is called a swarm. When determining the position of each particle, it takes advantage of its previous experience and adjusts it to the best position in the swarm [35]. In PSO, each particle has its best position vector called *pbest* (personal best) and passed down from generation to generation. The vector called *gbest* (global best) shows the best position information that the particle swarm has so far [36].

4. PROPOSED HIERARCHICAL APPROACHES

The six algorithms used in the study were tested on CEC'17 test functions[10]. After examining the performance of these algorithms on CEC'17 test functions and determining the characteristics of their algorithms, binary hierarchical structures were created. Hierarchical structure means that the methods work consecutively.

The first algorithm in the hierarchical approach was determined as one of the AAA, ABC, and DE algorithms. Because the global search success of these algorithms is better than other algorithms. The second hierarchical algorithm is the use of one of the ABC, GSA, and PSO algorithms whose local search capabilities are successful. The flow chart of the designed hierarchical structures is presented in Figure 1.

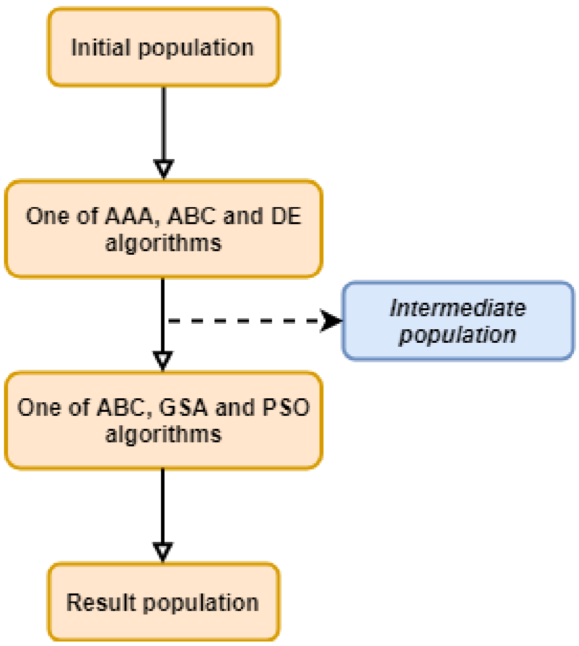


Figure 1. Flow chart of hierarchical approaches

As shown in the diagram in Figure 1, the algorithms were used successively in the hierarchical structures tested. The population developed by the first algorithm is called the intermediate population. This population is given as input to the second algorithm. In the first attempts of hierarchical approaches, the maximum number of function evaluations was equally distributed between the two algorithms as 50% -50%, and the results were examined. Then, 70% of the maximum number of function evaluations was determined to be the first algorithm and 30% to the second algorithm. This structure was compared with the structures where the maximum number of function evaluations was 50% - 50%. It was seen that the maximum number of function evaluations was 70% -30%, and the hierarchical structures gave more successful results. Therefore, the maximum number of function evaluations was taken as 70%- 30%. The results obtained after the completion of the second algorithm are the final results.

Table 1. Algorithm steps of AAA [3], ABC, DE [32], GA [32], GSA [34] and PSO [35]

AAA	ABC	DE
Step 1: Determination of parameters and initiation of algae colonies **REPEAT** **Step 2**: Helical movement stage **Step 3**: Evolutionary process **Step 4**: Adaptation process **Step 5**: Keep the best algae colony **UNTIL (number of iterations = Maximum number of iterations)**	**Step 1**: Determination of initial food sources **REPEAT** **Step 2**: Sending employed bees to food sources **Step 3**: Calculation of probability values **Step 4**: Selection of food source by onlooker bees **Step 5**: Resource release and explorer bee production **UNTIL (number of iterations = Maximum number of iterations)**	**Step 1**: Creating the initial population **REPEAT** **Step 2**: Mutation and regeneration **Step 3**: Crossover **Step 4**: Selection **UNTIL (number of iterations = Maximum number of iterations)**

GA	GSA	PSO
Step 1: Creating the initial population **REPEAT** **Step 2**: Calculation of the fitness values **Step 3**: Natural selection **Step 4**: Crossover **Step 5**: Mutation **UNTIL (number of iterations = Maximum number of iterations)**	**Step 1**: Creating the initial population **REPEAT** **Step 2**: Calculation of the fitness values **Step 3**: Finding the best and worst agent and updating the gravity value **Step 4**: Calculation of mass and acceleration of each agent **Step 5**: Updating speeds and locations **UNTIL (number of iterations = Maximum number of iterations)**	**Step 1**: Creating the initial population **REPEAT** **Step 2**: Calculation of the fitness values **Step 3**: The local best (pbest) is found for each particle. **Step 4**: Global best (gbest) is found **Step 5**: Positions and velocities are updated **UNTIL (number of iterations = Maximum number of iterations)**

In hierarchical trials, the CEC’17 test functions and their evaluation criteria were used. In this study, eight structures were tested. These eight structures are given in Table 2.

Table 2. Proposed structures

No	Proposed structures
1	AAA-ABC
2	AAA-GSA
3	AAA-PSO
4	ABC-GSA
5	ABC-PSO
6	DE-ABC
7	DE-GSA
8	DE-PSO

Hierarchical structures and singularities of algorithms were compared according to their average values over 20 independent runs in CEC’17 test functions.

5. EXPERIMENTAL STUDY

5.1. Experimental Setup

All tests in this study were made on CEC'17 test functions. The CEC’17 function set consists of 30 functions presented at the IEEE Evolutionary Computing Congress in 2017 and used to evaluate the performance of algorithms under equal conditions. The function is shown in Table 3. These functions have function groups defined in four different classes, single-mode (F1-F3), multi-mode (F4-F10), hybrid (F11-F20), and joined (F21-F30), and all functions are minimization problems. The search range [-100,100] is defined for all functions [37].

All algorithms were tested according to CEC’17 evaluation criteria. These criteria are given in Table 4. The specific parameters of each algorithm used in the algorithms are given in Table 5.

Table 3. CEC'17 test functions

Function Class	No	Function	F_i
Single mode	F1	Shifted and Rotated Bent Cigar Function	100
	F2*	Shifted and Rotated Sum of Different Power Function	200
	F3	Scrolled and Rotated Zakharov Function	300
Basic Multi-mode	F4	Scrolled and Rotated Rosenbrock Function	400
	F5	Shifted and Rotated Rastrigin Function	500
	F6	Scrolled and Rotated Extended Scaffer Function	600
	F7	Scrolled and Rotated Lunacek Bi_Rastrigin Function	700
	F8	Shifted and Rotated Discontinuous Rastrigin Function	800
	F9	Shifted and Rotated Levy Function	900
	F10	Shifted and Rotated Schwefel Function	1000
Hybrid	F11	Hybrid Function 1 (*N*=3)	1100
	F12	Hybrid Function 2 (*N*=3)	1200
	F13	Hybrid Function 3 (*N*=3)	1300
	F14	Hybrid Function 4 (*N*=4)	1400
	F15	Hybrid Function 5 (*N*=4)	1500
	F16	Hybrid Function 6 (*N*=4)	1600
	F17	Hybrid Function 7 (*N*=5)	1700
	F18	Hybrid Function 8 (*N*=5)	1800
	F19	Hybrid Function 9 (*N*=5)	1900
	F20	Hybrid Function 10 (*N*=6)	2000
Combined	F21	Combined Function 1 (*N*=3)	2100
	F22	Combined Function 2 (*N*=3)	2200
	F23	Combined Function 3 (*N*=4)	2300
	F24	Combined Function 4 (*N*=4)	2400
	F25	Combined Function 5 (*N*=5)	2500
	F26	Combined Function 6 (*N*=5)	2600
	F27	Combined Function 7 (*N*=6)	2700
	F28	Combined Function 8 (*N*=6)	2800
	F29	Combined Function 9 (*N*=3)	2900
	F30	Combined Function 10 (*N*=3)	3000
	Search Range: [-100,100]		

* The function F2 has been omitted because it exhibits large instability.

Table 4. Evaluation criteria of CEC'17 functions

Population size (*N*)	50
Dimension (*D*)	10, 30, 50, 100
Maximum function evaluation number (*MaxFES*)	10000 * D
Lower limit	-100
Upper limit	100
The number of runs (*run*)	20

Table 5. Parameters of algorithms

Algorithm	Parameters
AAA	Loss of energy (e) = 0.3 Shear force (K) = 2 Adaptation coefficient (A_p) = 0.2
ABC	Limit=100
DE	Step size (F_weight) = 1 Crossover probability constant (F_CR) = 0.9
GA	Crossover probability (p_c) = 0.9 Mutation probability (p_m) = 0.1 Stochastic Universal Sampling in Selection (SUS)
GSA	α parameter = 20 Gravity constant initial value (G_0) = 100
PSO	Inertia weight (w) = 1 Inertia Weight reduction ratio (wdamp) = 0.99 Learning Constants (c_1, c_2) = 2

The statistical results such as mean and standard deviation were used in all studies to evaluate the quality of the solutions. When comparing, the hierarchical approaches were compared according to their mean and standard deviation values, and the best hierarchical approach and the basic states of the algorithms were compared according to their mean values.

5.2. Experimental Results

In this study, hierarchical approaches are proposed for continuous optimization problems. First, these proposed hierarchical approaches were compared on the CEC'17 test functions in 10, 30, 50 and 100 dimensions. Then, according to the results obtained from this comparison, the hierarchical approach, which was found to be the most successful, and the basic states of the art algorithms used in the hierarchical approaches were again compared on the CEC'17 test functions in 10, 30, 50 and 100 dimensions. The results obtained in the hierarchical approaches are presented between Table 6 and Table 9. In addition, in the last row of these tables, there is the best value showing how many best results each hierarchical approach achieved.

Table 6. Results of hierarchical approaches for D = 10

F	AAA-ABC		AAA-GSA		AAA-PSO		ABC-GSA		ABC-PSO		DE-ABC		DE-GSA		DE-PSO	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
f1	1,06E+03	1,16E+03	1,20E+03	1,89E+03	6,22E+02	6,94E+02	8,92E+02	7,41E+02	3,30E+02	2,97E+02	**1,00E+02**	2,53E-09	**1,00E+02**	2,51E-09	**1,00E+02**	4,71E-09
f3	3,17E+02	3,07E+01	3,25E+02	2,78E+01	**3,00E+02**	9,83E-11	8,82E+03	3,40E+03	**3,00E+02**	1,27E-08	**3,00E+02**	2,61E-14	**3,00E+02**	1,30E-14	**3,00E+02**	0,00E+00
f4	4,04E+02	1,07E+00	4,04E+02	1,11E+00	4,03E+02	1,24E+00	4,01E+02	6,48E-01	**4,00E+02**	6,35E-01	**4,00E+02**	1,10E-08	4,03E+02	1,24E+01	**4,00E+02**	3,62E-12
f5	**5,05E+02**	1,51E+00	5,06E+02	2,35E+00	5,06E+02	2,05E+00	5,07E+02	2,09E+00	5,11E+02	2,61E+00	5,09E+02	2,89E+00	5,12E+02	4,05E+00	5,14E+02	7,03E+00
f6	**6,00E+02**	9,76E-14	**6,00E+02**	6,39E-14	**6,00E+02**	7,82E-14	**6,00E+02**	2,89E-06	**6,00E+02**	7,94E-08	**6,00E+02**	2,79E-02	6,01E+02	4,66E-01	**6,00E+02**	3,82E-01
f7	**7,14E+02**	1,64E+00	7,16E+02	1,82E+00	7,15E+02	2,05E+00	7,21E+02	3,23E+00	7,20E+02	3,14E+00	7,20E+02	4,25E+00	7,26E+02	4,73E+00	7,24E+02	9,65E+00
f8	8,05E+02	2,11E+00	8,05E+02	1,67E+00	**8,03E+02**	1,50E+00	8,07E+02	1,63E+00	8,11E+02	4,09E+00	8,12E+02	4,08E+00	8,11E+02	3,42E+00	8,15E+02	6,14E+00
f9	**9,00E+02**	4,92E-13	**9,00E+02**	2,29E-13	**9,00E+02**	3,69E-14	**9,00E+02**	3,05E-01	**9,00E+02**	6,90E-14	9,01E+02	1,82E+00	9,01E+02	5,99E-01	9,16E+02	2,63E+01
f10	1,20E+03	1,19E+02	1,27E+03	1,25E+02	**1,17E+03**	1,22E+02	1,44E+03	1,37E+02	1,38E+03	1,02E+02	1,34E+03	1,18E+02	1,51E+03	2,01E+02	1,40E+03	1,81E+02
f11	**1,10E+03**	7,49E-01	**1,10E+03**	1,01E+00	**1,10E+03**	9,01E-01	1,13E+03	2,54E+01	1,11E+03	5,89E+00	1,11E+03	5,21E+00	1,11E+03	8,45E+00	1,11E+03	1,39E+01
f12	1,01E+04	7,59E+03	9,32E+03	6,35E+03	8,37E+03	5,79E+03	5,80E+05	3,50E+05	1,29E+04	9,06E+03	**1,35E+03**	1,87E+02	1,48E+03	2,58E+02	1,45E+03	1,94E+02
f13	2,30E+03	2,41E+03	1,49E+03	2,25E+02	1,67E+03	5,82E+02	6,63E+03	3,33E+03	4,47E+03	3,32E+03	**1,31E+03**	6,07E+00	**1,31E+03**	2,08E+01	1,32E+03	2,10E+01
f14	1,46E+03	7,64E+01	1,44E+03	9,54E+01	1,42E+03	3,18E+01	3,18E+03	1,11E+03	1,48E+03	9,35E+01	**1,41E+03**	8,34E+00	**1,41E+03**	1,29E+01	**1,41E+03**	1,01E+01
f15	1,54E+03	7,86E+01	1,54E+03	4,12E+01	1,52E+03	3,93E+01	2,05E+03	5,71E+02	1,52E+03	1,68E+01	**1,50E+03**	2,83E+00	**1,50E+03**	2,18E+00	**1,50E+03**	2,81E+00
f16	**1,60E+03**	8,52E+00	**1,60E+03**	7,08E-01	**1,60E+03**	9,06E+00	1,62E+03	1,99E+01	**1,60E+03**	8,66E+00	1,62E+03	2,82E+01	1,64E+03	5,15E+01	1,66E+03	7,09E+01
f17	**1,70E+03**	3,83E+00	**1,70E+03**	6,87E-01	**1,70E+03**	4,98E-01	**1,70E+03**	2,91E+00	**1,70E+03**	2,84E+00	1,71E+03	8,21E+00	1,73E+03	2,97E+01	1,72E+03	3,63E+01
f18	2,92E+03	1,21E+03	2,77E+03	1,13E+03	2,49E+03	8,11E+02	4,33E+03	1,14E+03	5,01E+03	2,72E+03	1,82E+03	1,20E+01	**1,81E+03**	1,56E+01	1,82E+03	1,58E+01
f19	1,92E+03	6,09E+01	1,94E+03	6,89E+01	1,91E+03	1,05E+01	2,36E+03	2,81E+02	1,93E+03	2,94E+01	**1,90E+03**	1,19E+00	**1,90E+03**	1,66E+00	**1,90E+03**	1,03E+00
f20	**2,00E+03**	8,82E-03	**2,00E+03**	2,69E-05	**2,00E+03**	6,98E-02	**2,00E+03**	4,59E-01	**2,00E+03**	4,45E-01	**2,00E+03**	4,17E+00	**2,02E+03**	2,60E+01	2,01E+03	1,17E+01
f21	2,21E+03	2,28E+01	2,22E+03	4,36E+01	2,22E+03	4,41E+01	2,22E+03	6,12E+00	**2,20E+03**	1,42E+00	2,22E+03	1,37E+01	2,31E+03	2,64E+01	2,31E+03	3,69E+01
f22	2,28E+03	3,61E+01	**2,27E+03**	4,38E+01	**2,27E+03**	4,06E+01	2,29E+03	1,85E+01	2,28E+03	3,13E+01	2,29E+03	2,93E+01	2,30E+03	2,23E+01	2,30E+03	9,17E-01
f23	2,61E+03	2,19E+00	2,61E+03	3,11E+00	2,61E+03	1,91E+00	2,61E+03	3,95E+00	**2,58E+03**	9,40E+01	2,61E+03	4,17E+00	2,61E+03	4,18E+00	2,62E+03	6,05E+00
f24	2,56E+03	7,27E+01	2,61E+03	1,21E+02	2,63E+03	1,24E+02	2,51E+03	5,06E+00	**2,50E+03**	2,42E-06	2,54E+03	2,44E+01	2,74E+03	4,50E+00	2,73E+03	7,78E+01
f25	2,90E+03	2,48E-01	2,90E+03	5,42E-01	2,90E+03	4,49E-01	2,88E+03	5,17E+01	2,86E+03	9,37E+01	**2,78E+03**	9,27E+01	2,91E+03	2,04E+01	2,91E+03	2,06E+01
f26	2,81E+03	1,39E+02	2,84E+03	1,23E+02	2,88E+03	7,16E+01	2,80E+03	7,86E+01	**2,74E+03**	1,04E+02	2,84E+03	7,17E+01	3,01E+03	3,15E+02	2,94E+03	4,56E+01
f27	**3,09E+03**	7,39E-01	**3,09E+03**	4,86E-01	**3,09E+03**	5,88E-01	3,10E+03	3,19E+00	**3,09E+03**	2,71E+00	**3,09E+03**	3,12E+00	**3,09E+03**	1,09E+01	3,10E+03	1,71E+01
f28	3,13E+03	3,45E+01	3,13E+03	8,20E+01	3,10E+03	7,72E+01	3,12E+03	6,44E+01	3,10E+03	5,92E+00	**3,09E+03**	1,18E+02	3,27E+03	1,50E+02	3,27E+03	1,41E+02
f29	**3,14E+03**	4,61E+00	3,15E+03	1,15E+01	**3,14E+03**	9,07E+00	3,19E+03	2,74E+01	3,17E+03	3,16E+01	3,15E+03	2,27E+01	3,17E+03	3,22E+01	3,15E+03	1,50E+01
f30	**5,07E+03**	1,28E+03	6,42E+03	1,87E+03	7,01E+03	4,28E+03	1,69E+05	1,07E+05	1,20E+04	7,18E+03	8,62E+03	1,21E+04	1,26E+05	2,99E+05	2,92E+05	4,66E+05
Best	**11**		**8**		**12**		**4**		**12**		**13**		**9**		**7**	

Table 7. Results of hierarchical approaches for D = 30

F	AAA-ABC		AAA-GSA		AAA-PSO		ABC-GSA		ABC-PSO		DE-ABC		DE-GSA		DE-PSO	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
f1	6,35E+02	9,60E+02	**4,24E+02**	4,72E+02	7,80E+02	1,11E+03	1,04E+03	7,61E+02	6,61E+02	5,48E+02	8,74E+02	1,32E+03	4,47E+02	5,86E+02	8,63E+02	2,55E+03
f3	2,01E+04	7,35E+03	2,42E+04	1,01E+04	1,38E+03	1,56E+02	8,13E+04	8,66E+03	2,20E+03	5,34E+02	4,60E+02	2,40E+02	6,43E+02	5,61E+02	**4,39E+02**	2,09E+02
f4	4,74E+02	2,02E+01	4,59E+02	3,08E+01	4,66E+02	2,27E+01	4,43E+02	2,66E+01	**4,35E+02**	2,72E+01	4,75E+02	2,47E+01	4,77E+02	1,17E+01	4,77E+02	2,02E+01
f5	5,53E+02	1,31E+01	**5,46E+02**	9,56E+00	**5,46E+02**	1,20E+01	5,59E+02	8,67E+00	6,02E+02	1,80E+01	5,73E+02	1,54E+01	5,62E+02	1,33E+01	5,78E+02	1,80E+01
f6	**6,00E+02**	1,17E-13	**6,00E+02**	1,25E-13	**6,00E+02**	1,45E-13	**6,00E+02**	3,28E-06	**6,00E+02**	2,48E-07	6,10E+02	3,68E+00	6,02E+02	2,00E-01	6,02E+02	2,16E+00
f7	7,75E+02	8,54E+00	**7,74E+02**	1,08E+01	7,79E+02	1,20E+01	8,31E+02	1,56E+01	8,15E+02	1,71E+01	8,19E+02	1,51E+01	8,57E+02	1,89E+01	8,89E+02	4,49E+01
f8	8,60E+02	1,99E+01	8,51E+02	1,24E+01	**8,49E+02**	1,15E+01	8,52E+02	5,14E+00	9,07E+02	1,69E+01	8,79E+02	1,78E+01	8,74E+02	1,88E+01	8,85E+02	2,09E+01
f9	**9,01E+02**	6,58E-01	**9,01E+02**	1,83E+00	9,02E+02	2,15E+00	9,07E+02	9,66E-01	2,25E+03	1,05E+03	2,02E+03	7,14E+02	9,08E+02	8,87E-01	2,20E+03	8,05E+02
f10	**2,97E+03**	4,62E+02	3,03E+03	4,19E+02	2,99E+03	4,45E+02	4,00E+03	2,55E+02	4,14E+03	3,00E+02	3,91E+03	5,73E+02	3,61E+03	5,48E+02	4,02E+03	5,49E+02
f11	**1,13E+03**	9,37E+00	1,14E+03	1,55E+01	**1,13E+03**	9,14E+00	2,40E+03	6,68E+02	1,18E+03	4,04E+01	1,21E+03	4,92E+01	1,22E+03	7,04E+01	1,21E+03	5,07E+01
f12	3,75E+05	2,78E+05	4,79E+05	2,62E+05	3,13E+05	2,20E+05	2,78E+06	7,64E+05	5,04E+05	4,81E+05	**3,23E+04**	1,70E+04	4,73E+04	2,13E+04	4,52E+04	3,00E+04
f13	7,38E+03	6,98E+03	6,62E+03	7,66E+03	5,00E+03	6,05E+03	1,66E+05	9,32E+04	1,27E+04	8,30E+03	5,83E+03	1,37E+04	**2,65E+03**	2,26E+03	8,74E+03	1,84E+04
f14	2,99E+04	2,99E+04	1,63E+04	1,64E+04	1,45E+04	1,07E+04	5,26E+05	1,89E+05	2,56E+04	3,32E+04	1,48E+03	3,23E+01	**1,47E+03**	2,00E+01	1,48E+03	3,57E+01
f15	2,40E+03	1,54E+03	2,87E+03	1,62E+03	2,25E+03	9,23E+02	3,19E+04	1,28E+04	5,16E+03	5,38E+03	1,65E+03	1,22E+02	1,67E+03	1,30E+02	**1,62E+03**	8,71E+01
f16	**2,12E+03**	1,50E+02	2,25E+03	1,06E+02	2,15E+03	1,44E+02	2,47E+03	1,70E+02	2,31E+03	1,81E+02	2,58E+03	2,39E+02	2,39E+03	2,60E+02	2,58E+03	2,20E+02
f17	**1,84E+03**	9,87E+01	**1,84E+03**	8,67E+01	**1,84E+03**	7,26E+01	2,08E+03	1,57E+02	1,95E+03	1,16E+02	2,10E+03	1,64E+02	2,07E+03	1,88E+02	2,18E+03	2,18E+02
f18	1,32E+05	6,54E+04	1,31E+05	9,05E+04	1,24E+05	5,67E+04	7,46E+05	5,27E+05	3,07E+05	2,52E+05	3,27E+03	1,34E+03	4,16E+03	2,61E+03	**2,90E+03**	1,09E+03
f19	2,60E+03	9,66E+02	3,20E+03	2,97E+03	3,48E+03	2,13E+03	4,91E+04	1,86E+04	6,56E+03	5,04E+03	1,95E+03	6,20E+01	**1,94E+03**	3,60E+01	**1,94E+03**	4,15E+01
f20	**2,15E+03**	7,58E+01	2,17E+03	7,20E+01	2,19E+03	9,79E+01	2,47E+03	1,18E+02	2,35E+03	1,32E+02	2,27E+03	1,01E+02	2,33E+03	2,13E+02	2,43E+03	2,30E+02
f21	2,34E+03	2,66E+01	2,35E+03	3,63E+01	2,34E+03	3,49E+01	**2,33E+03**	3,29E+01	2,38E+03	6,74E+01	2,35E+03	5,74E+01	2,37E+03	1,94E+01	2,39E+03	2,37E+01
f22	2,46E+03	5,13E+02	**2,43E+03**	5,93E+02	2,61E+03	7,66E+02	3,12E+03	1,43E+03	2,46E+03	6,95E+02	2,88E+03	1,18E+03	4,64E+03	1,09E+03	4,45E+03	1,58E+03
f23	**2,68E+03**	9,65E+01	**2,68E+03**	8,60E+01	2,69E+03	6,98E+01	2,71E+03	1,22E+01	2,75E+03	2,01E+01	2,72E+03	1,53E+01	2,72E+03	2,03E+01	2,73E+03	1,92E+01
f24	2,88E+03	8,89E+01	2,90E+03	1,66E+01	2,90E+03	1,86E+01	**2,85E+03**	1,48E+02	2,97E+03	1,36E+02	2,87E+03	1,08E+02	2,90E+03	1,96E+01	2,91E+03	2,67E+01
f25	2,89E+03	1,50E+00	2,89E+03	1,26E+00	2,89E+03	1,58E+00	2,89E+03	1,44E+00	**2,88E+03**	1,62E+00	2,89E+03	7,14E+00	2,89E+03	1,43E+00	2,89E+03	1,05E+01
f26	3,40E+03	6,78E+02	3,19E+03	5,49E+02	3,86E+03	6,74E+02	3,10E+03	3,20E+02	**2,92E+03**	1,07E+02	3,42E+03	6,25E+02	4,36E+03	4,03E+02	4,55E+03	4,82E+02
f27	**3,20E+03**	6,41E+00	**3,20E+03**	5,76E+00	**3,20E+03**	8,97E+00	3,22E+03	6,01E+00	3,22E+03	7,49E+00	3,21E+03	1,47E+01	3,22E+03	1,58E+01	3,21E+03	1,24E+01
f28	3,21E+03	6,57E+00	3,20E+03	1,43E+01	3,21E+03	5,76E+00	3,21E+03	4,56E+00	3,21E+03	4,98E+00	**3,19E+03**	5,70E+01	**3,19E+03**	4,49E+01	**3,19E+03**	5,27E+01
f29	3,40E+03	8,69E+01	3,41E+03	8,40E+01	**3,38E+03**	6,71E+01	3,68E+03	8,84E+01	3,58E+03	7,93E+01	3,68E+03	1,76E+02	3,72E+03	1,59E+02	3,62E+03	1,92E+02
f30	7,88E+03	2,10E+03	7,25E+03	1,26E+03	7,92E+03	3,11E+03	8,12E+04	5,39E+04	1,18E+04	3,48E+03	**5,17E+03**	1,79E+02	5,19E+03	2,83E+02	5,26E+03	3,08E+02
Best	**9**		**9**		**7**		**3**		**4**		**3**		**4**		**5**	

Table 8. Results of hierarchical approaches for D = 50

F	AAA-ABC		AAA-GSA		AAA-PSO		ABC-GSA		ABC-PSO		DE-ABC		DE-GSA		DE-PSO	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
f1	3,03E+03	3,60E+03	**2,19E+03**	1,93E+03	2,93E+03	3,34E+03	1,23E+04	6,04E+03	3,39E+03	3,03E+03	1,17E+04	1,28E+04	1,51E+04	1,20E+04	1,47E+04	1,19E+04
f3	7,83E+04	1,77E+04	7,11E+04	1,49E+04	7,98E+03	1,98E+03	1,78E+05	1,03E+04	1,09E+04	2,10E+03	4,69E+04	1,84E+04	3,47E+04	1,91E+04	**5,46E+03**	1,23E+03
f4	4,55E+02	2,79E+01	**4,47E+02**	2,46E+01	4,54E+02	2,82E+01	4,57E+02	2,01E+01	4,55E+02	2,30E+01	5,33E+02	5,35E+01	5,61E+02	4,12E+01	5,49E+02	4,90E+01
f5	6,19E+02	2,13E+01	6,17E+02	2,38E+01	6,22E+02	2,65E+01	**6,07E+02**	1,14E+01	7,49E+02	3,23E+01	6,82E+02	2,72E+01	6,47E+02	2,51E+01	6,69E+02	3,87E+01
f6	**6,00E+02**	1,22E-13	**6,00E+02**	1,17E-13	**6,00E+02**	1,79E-13	**6,00E+02**	3,51E-06	**6,00E+02**	1,11E-06	6,17E+02	3,65E+00	6,03E+02	1,07E+00	6,11E+02	5,13E+00
f7	8,68E+02	2,38E+01	**8,66E+02**	2,19E+01	8,73E+02	2,56E+01	9,70E+02	1,55E+01	9,58E+02	2,77E+01	1,07E+03	1,05E+02	9,85E+02	1,46E+01	1,26E+03	1,46E+02
f8	9,22E+02	3,04E+01	**9,11E+02**	2,40E+01	9,23E+02	2,73E+01	9,15E+02	1,06E+01	1,03E+03	2,21E+01	9,80E+02	2,93E+01	9,57E+02	3,81E+01	9,77E+02	2,55E+01
f9	1,57E+03	6,27E+02	**9,15E+02**	1,39E+00	1,25E+03	4,00E+02	**9,15E+02**	1,72E+00	1,60E+04	4,24E+03	4,76E+03	1,31E+03	9,47E+02	1,00E+02	4,88E+03	2,06E+03
f10	**4,85E+03**	3,94E+02	4,87E+03	5,75E+02	4,99E+03	4,46E+02	6,41E+03	4,21E+02	6,60E+03	4,13E+02	6,17E+03	8,68E+02	6,12E+03	6,38E+02	6,47E+03	1,08E+03
f11	1,22E+03	5,71E+01	1,21E+03	3,76E+01	**1,20E+03**	2,61E+01	5,21E+03	1,27E+03	1,26E+03	4,83E+01	1,33E+03	6,44E+01	1,32E+03	7,53E+01	1,28E+03	6,49E+01
f12	3,25E+06	1,87E+06	3,06E+06	2,17E+06	2,47E+06	1,35E+06	1,07E+07	3,02E+06	3,42E+06	1,79E+06	3,64E+05	2,80E+05	5,13E+05	4,14E+05	**3,15E+05**	2,15E+05
f13	5,43E+03	5,92E+03	**3,65E+03**	2,88E+03	4,15E+03	4,74E+03	2,34E+05	2,31E+05	6,11E+03	6,70E+03	2,04E+04	1,43E+04	2,67E+04	1,30E+04	2,50E+04	1,11E+04
f14	2,87E+05	2,38E+05	1,32E+05	1,08E+05	8,53E+04	7,63E+04	6,67E+05	2,86E+05	6,80E+04	3,92E+04	3,26E+03	1,63E+03	2,76E+03	1,00E+03	**2,63E+03**	9,32E+02
f15	6,66E+03	5,11E+03	6,85E+03	5,37E+03	**6,12E+03**	5,08E+03	7,09E+04	4,35E+04	1,55E+04	5,46E+03	2,41E+04	1,11E+04	1,51E+04	1,22E+04	1,83E+04	1,12E+04
f16	**2,80E+03**	2,82E+02	**2,80E+03**	2,51E+02	2,81E+03	1,71E+02	3,12E+03	1,95E+02	3,02E+03	2,92E+02	3,37E+03	4,82E+02	3,10E+03	4,45E+02	3,30E+03	4,15E+02
f17	**2,41E+03**	1,95E+02	2,43E+03	1,96E+02	2,47E+03	2,12E+02	3,00E+03	2,43E+02	2,87E+03	2,10E+02	2,96E+03	3,86E+02	2,88E+03	3,47E+02	3,13E+03	3,25E+02
f18	6,17E+05	3,12E+05	6,20E+05	5,23E+05	6,38E+05	4,08E+05	1,65E+06	5,96E+05	1,42E+06	9,34E+05	1,23E+04	7,42E+03	1,19E+04	7,37E+03	**1,11E+04**	5,68E+03
f19	6,71E+03	4,98E+03	7,82E+03	4,17E+03	5,35E+03	2,48E+03	8,72E+04	2,74E+04	2,38E+04	1,03E+04	4,12E+03	4,25E+03	6,11E+03	7,86E+03	**2,92E+03**	9,65E+02
f20	2,62E+03	1,47E+02	**2,59E+03**	2,20E+02	2,61E+03	2,02E+02	3,07E+03	1,83E+02	3,01E+03	2,38E+02	2,90E+03	2,28E+02	2,96E+03	3,33E+02	2,96E+03	3,34E+02
f21	2,43E+03	2,10E+01	2,42E+03	1,82E+01	2,42E+03	2,15E+01	**2,40E+03**	1,62E+01	2,54E+03	3,27E+01	2,49E+03	3,84E+01	2,45E+03	2,71E+01	2,46E+03	3,74E+01
f22	6,32E+03	1,50E+03	**6,23E+03**	1,42E+03	6,61E+03	8,15E+02	8,67E+03	5,23E+02	8,27E+03	1,46E+03	8,69E+03	6,62E+02	7,65E+03	7,77E+02	8,03E+03	9,24E+02
f23	2,87E+03	3,19E+01	2,87E+03	2,76E+01	2,87E+03	2,44E+01	**2,86E+03**	1,70E+01	3,00E+03	2,66E+01	2,92E+03	3,52E+01	2,90E+03	4,58E+01	2,93E+03	4,24E+01
f24	3,12E+03	6,01E+01	3,12E+03	4,09E+01	3,11E+03	5,25E+01	3,24E+03	2,84E+01	3,43E+03	1,79E+02	**3,05E+03**	8,46E+01	3,07E+03	3,89E+01	3,08E+03	4,81E+01
f25	**3,00E+03**	2,86E+01	3,01E+03	1,90E+01	3,01E+03	2,52E+01	3,02E+03	1,47E+01	3,02E+03	1,00E+01	3,03E+03	4,77E+01	3,02E+03	2,77E+01	3,02E+03	3,42E+01
f26	4,41E+03	1,05E+03	4,77E+03	6,74E+02	5,14E+03	5,77E+02	**4,23E+03**	8,66E+02	4,96E+03	1,61E+03	5,35E+03	6,81E+02	5,58E+03	4,98E+02	6,03E+03	5,25E+02
f27	**3,26E+03**	2,77E+01	3,27E+03	2,19E+01	3,27E+03	2,45E+01	3,39E+03	2,84E+01	3,39E+03	3,06E+01	3,40E+03	9,25E+01	3,38E+03	9,36E+01	3,42E+03	9,19E+01
f28	**3,26E+03**	1,05E+01	3,27E+03	1,42E+01	3,28E+03	1,64E+01	3,29E+03	1,13E+01	3,28E+03	1,49E+01	3,30E+03	2,22E+01	3,34E+03	2,11E+02	3,30E+03	2,05E+01
f29	3,58E+03	1,31E+02	3,64E+03	1,50E+02	**3,57E+03**	1,77E+02	4,29E+03	1,86E+02	4,26E+03	1,86E+02	4,40E+03	3,43E+02	4,17E+03	2,95E+02	4,33E+03	3,38E+02
f30	6,97E+05	7,12E+04	6,91E+05	6,80E+04	6,94E+05	5,75E+04	1,49E+06	4,43E+05	8,80E+05	9,34E+04	**6,62E+05**	8,93E+04	7,31E+05	2,20E+05	7,64E+05	1,56E+05
Best	**7**		**10**		**4**		**6**		**1**		**2**		**0**		**5**	

Table 9. Results of hierarchical approaches for D = 100

F	AAA-ABC		AAA-GSA		AAA-PSO		ABC-GSA		ABC-PSO		DE-ABC		DE-GSA		DE-PSO	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
f1	2,89E+03	3,61E+03	**2,14E+03**	2,05E+03	2,79E+03	4,42E+03	1,69E+04	8,43E+03	7,46E+03	7,43E+03	4,09E+08	8,15E+08	2,32E+09	6,17E+09	1,01E+09	1,46E+09
f3	2,95E+05	5,45E+04	2,92E+05	3,52E+04	**8,47E+04**	1,99E+04	3,62E+05	1,97E+04	1,00E+05	1,03E+04	4,78E+05	6,68E+04	3,01E+05	5,28E+04	1,34E+05	2,80E+04
f4	**6,14E+02**	2,02E+01	6,18E+02	2,32E+01	6,22E+02	2,47E+01	6,26E+02	2,70E+01	6,28E+02	2,84E+01	7,09E+02	5,79E+01	7,48E+02	9,98E+01	7,36E+02	8,84E+01
f5	9,08E+02	6,14E+01	9,10E+02	4,70E+01	9,18E+02	6,46E+01	**8,32E+02**	1,87E+01	1,23E+03	8,20E+01	1,02E+03	7,18E+01	9,55E+02	6,96E+01	1,00E+03	7,11E+01
f6	**6,00E+02**	2,17E-13	**6,00E+02**	2,39E-13	**6,00E+02**	2,60E-13	**6,00E+02**	3,41E-06	**6,00E+02**	3,15E-06	6,36E+02	6,27E+00	6,09E+02	3,49E+00	6,30E+02	6,34E+00
f7	**1,18E+03**	5,35E+01	**1,18E+03**	5,59E+01	**1,18E+03**	4,10E+01	1,23E+03	1,50E+01	1,47E+03	6,08E+01	3,07E+03	4,13E+02	1,53E+03	1,40E+02	2,94E+03	4,13E+02
f8	1,22E+03	7,01E+01	1,22E+03	7,28E+01	1,20E+03	6,85E+01	**1,15E+03**	2,57E+01	1,57E+03	7,62E+01	1,32E+03	6,96E+01	1,22E+03	5,15E+01	1,32E+03	7,83E+01
f9	1,82E+04	4,23E+03	7,13E+03	3,83E+03	1,86E+04	7,48E+03	**3,09E+03**	1,03E+03	7,42E+04	8,30E+03	1,58E+04	5,02E+03	4,54E+03	1,56E+03	1,43E+04	3,22E+03
f10	1,15E+04	1,10E+03	**1,12E+04**	8,51E+02	1,14E+04	9,47E+02	1,37E+04	8,34E+02	1,52E+04	9,74E+02	1,41E+04	1,25E+03	1,28E+04	7,60E+02	1,41E+04	1,20E+03
f11	1,77E+04	6,20E+03	1,72E+04	5,29E+03	2,16E+03	2,27E+02	6,48E+04	8,27E+03	2,59E+03	3,67E+02	1,90E+03	2,98E+02	**1,74E+03**	1,55E+02	2,44E+03	2,01E+03
f12	1,09E+07	4,05E+06	1,21E+07	8,28E+06	1,07E+07	4,14E+06	3,67E+07	7,09E+06	2,45E+07	7,58E+06	**5,81E+06**	3,53E+06	6,60E+06	5,20E+06	5,98E+06	2,78E+06
f13	**2,97E+03**	1,75E+03	3,38E+03	2,13E+03	3,15E+03	2,64E+03	7,67E+04	5,74E+04	5,57E+03	2,04E+03	1,99E+04	1,25E+04	4,19E+05	1,81E+06	1,80E+04	1,31E+04
f14	1,66E+06	1,21E+06	9,58E+05	5,01E+05	1,38E+06	7,54E+05	9,57E+05	2,51E+05	1,53E+06	8,71E+05	**2,94E+04**	1,63E+04	3,39E+04	1,97E+04	3,34E+04	1,69E+04
f15	2,36E+03	7,28E+02	**2,00E+03**	4,28E+02	2,15E+03	4,79E+02	1,41E+05	8,14E+04	2,86E+03	1,19E+03	1,48E+04	1,15E+04	1,34E+04	1,07E+04	1,27E+04	1,00E+04
f16	4,77E+03	3,62E+02	**4,65E+03**	4,52E+02	4,78E+03	3,54E+02	5,66E+03	3,95E+02	5,07E+03	4,64E+02	5,56E+03	3,58E+02	5,41E+03	7,83E+02	5,39E+03	7,08E+02
f17	4,09E+03	2,76E+02	**3,99E+03**	3,72E+02	4,06E+03	2,98E+02	4,88E+03	5,40E+02	4,89E+03	4,48E+02	5,00E+03	4,97E+02	4,89E+03	5,23E+02	4,90E+03	5,58E+02
f18	2,22E+06	1,02E+06	1,63E+06	7,84E+05	1,73E+06	8,45E+05	1,09E+06	2,40E+05	4,22E+06	2,06E+06	2,40E+05	1,04E+05	2,20E+05	1,29E+05	**2,07E+05**	9,36E+04
f19	**2,33E+03**	4,53E+02	2,40E+03	5,31E+02	2,82E+03	8,42E+02	2,20E+05	8,88E+04	3,23E+03	1,10E+03	2,20E+04	1,40E+04	2,41E+04	1,46E+04	2,36E+04	1,45E+04
f20	4,20E+03	3,77E+02	**4,08E+03**	3,53E+02	4,32E+03	2,73E+02	5,48E+03	4,48E+02	4,86E+03	3,74E+02	4,67E+03	4,81E+02	4,41E+03	3,70E+02	4,65E+03	4,03E+02
f21	2,74E+03	5,25E+01	2,71E+03	6,45E+01	2,73E+03	6,47E+01	**2,67E+03**	2,15E+01	3,06E+03	7,28E+01	2,86E+03	7,83E+01	2,81E+03	8,29E+01	2,84E+03	6,58E+01
f22	**1,37E+04**	9,05E+02	**1,37E+04**	1,19E+03	1,39E+04	1,15E+03	1,80E+04	7,94E+02	1,83E+04	8,64E+02	1,64E+04	9,55E+02	1,53E+04	8,46E+02	1,62E+04	1,27E+03
f23	**3,08E+03**	3,92E+01	**3,08E+03**	3,81E+01	3,10E+03	4,90E+01	3,17E+03	2,23E+01	3,23E+03	3,35E+01	3,32E+03	8,38E+01	3,35E+03	7,80E+01	3,38E+03	8,57E+01
f24	**3,67E+03**	5,22E+01	3,68E+03	6,27E+01	3,68E+03	5,13E+01	3,68E+03	4,02E+01	3,90E+03	4,53E+01	3,88E+03	1,09E+02	3,89E+03	1,28E+02	3,92E+03	1,51E+02
f25	**3,19E+03**	4,32E+01	3,21E+03	4,64E+01	3,23E+03	4,52E+01	3,24E+03	2,53E+01	3,25E+03	2,97E+01	3,36E+03	9,27E+01	3,33E+03	9,61E+01	3,40E+03	7,48E+01
f26	1,01E+04	1,13E+03	9,84E+03	6,06E+02	1,02E+04	4,71E+02	**9,44E+03**	3,54E+02	1,33E+04	5,09E+02	1,27E+04	1,17E+03	1,21E+04	9,83E+02	1,23E+04	1,13E+03
f27	**3,36E+03**	3,58E+01	3,38E+03	2,98E+01	3,39E+03	2,20E+01	3,52E+03	3,86E+01	3,48E+03	3,12E+01	3,51E+03	7,81E+01	3,50E+03	6,34E+01	3,54E+03	5,65E+01
f28	3,37E+03	1,82E+01	**3,36E+03**	1,69E+01	3,37E+03	1,98E+01	3,41E+03	1,48E+01	3,40E+03	2,00E+01	3,62E+03	3,01E+02	4,09E+03	1,51E+03	4,79E+03	3,71E+03
f29	**5,52E+03**	2,91E+02	5,53E+03	3,56E+02	5,54E+03	2,78E+02	7,39E+03	2,74E+02	7,37E+03	3,69E+02	6,54E+03	5,97E+02	6,62E+03	5,11E+02	6,58E+03	5,57E+02
f30	9,47E+03	2,40E+03	1,02E+04	3,94E+03	**9,41E+03**	2,74E+03	4,64E+04	3,92E+04	1,97E+04	3,65E+03	1,49E+04	1,42E+04	1,79E+04	2,07E+04	2,87E+04	7,04E+04
Best	**11**		**11**		**4**		**6**		**1**		**2**		**1**		**1**	

Table 10. Friedman test results of all hierarchical approaches

			AAA-ABC	AAA-GSA	AAA-PSO	ABC-GSA	ABC-PSO	DE-ABC	DE-GSA	DE-PSO
D = 10		***Best***	11	8	12	4	12	**13**	9	7
	Friedman Test	***Mean Rank***	3.9310	4.2759	**3.5000**	5.5517	4.0690	3.6207	5.5172	5.5345
		Final Rank	3	5	**1**	8	4	2	6	7
		p-Value	**7.1387e-05**							
D = 30		***Best***	**9**	**9**	7	3	4	3	4	5
	Friedman Test	***Mean Rank***	3.4483	**3.3448**	3.4483	5.7586	5.5172	4.4655	4.6552	5.3621
		Final Rank	2.5	**1**	2.5	8	7	4	5	6
		p-Value	**9.9017e-06**							
D = 50		***Best***	7	**10**	4	6	**1**	2	0	5
	Friedman Test	***Mean Rank***	3.1552	**2.6897**	3.0000	5.5690	5.6207	5.5690	4.8966	5.5000
		Final Rank	3	**1**	2	6.5	8	6.5	4	5
		p-Value	**2.1195e-10**							
D = 100		***Best***	**11**	**11**	4	6	1	2	1	1
	Friedman Test	***Mean Rank***	2.9828	**2.5517**	3.0517	5.0345	5.8103	5.6897	5.1897	5.6897
		Final Rank	2	**1**	3	4	8	6.5	5	6.5
		p-Value	**8.2196e-12**							

Looking at the comparison of hierarchical approaches in 10 dimensions in Table 6, has become the most successful DE-ABC hierarchical structure. Here, the DE-ABC hierarchical approach has been the most successful due to the low-dimensional global search ability of DE and the low-dimensional local search ability of ABC. Then, the hierarchical structures AAA-PSO and ABC-PSO were the most successful. In Table 7, the comparison of hierarchical approaches in 30 dimensions, has become the most successful AAA-ABC and AAA-GSA hierarchical structures. Here, the AAA-ABC and AAA-GSA hierarchical approaches were most successful, with the global search capabilities of the AAA algorithm combined with the local search capabilities of the ABC and GSA. In 50 dimensions the comparison of hierarchical approaches in Table 8, has become the most successful AAA-GSA hierarchical structure. In Table 9 for 100 dimensions; has become the most successful AAA-ABC and AAA-GSA hierarchical structures. As the size of CEC'17 test functions increases, the problem becomes more difficult and algorithms have difficulty in solving these problems. It can be seen from these tables that the AAA-GSA hierarchical approach has managed to maintain its success even if the problem becomes difficult. For the AAA-GSA hierarchical approach, the combination of AAA's strong global search capability and GSA's strong local search resulted in a more powerful approach. This approach proved to be powerful by achieving success as the problem size increased.

When the results between Table 6 and Table 9 are examined, it is observed that some results are close to each other. For this reason, it is necessary to determine whether the results obtained by the methods have a statistically significant difference. At this point, the Friedman test is widely used in the literature. The Friedman test is a non-parametric statistical test used to rank the results of multiple methods [38]. In this study, the Friedman test was used both when ordering hierarchical approaches among themselves and when ordering the most successful hierarchical approach and the basic states of algorithms. The significance level (p-value) for the Friedman test was determined as 0.05. If this p-Value is less than 0.05, there is a statistically significant difference between the results. If the p-Value is not less than 0.05, there is no significant difference. Mean rank, final rank, and p-Value values obtained in Friedman rank test results of hierarchical approaches are given in Table 10. As seen in Table 10, the method with the lowest mean rank value is the most successful method. Looking at the Friedman test results in Table 10, while the AAA-PSO hierarchical approach in 10 dimensions was the first in the ranking with the lowest mean rank, the AAA-GSA hierarchical approach in 30, 50, and 100 dimensions were the first in the ranking with the lowest mean rank. As seen in the Table 10, each p-Value is less than 0.05. This indicates that there is a statistically significant difference between the results. Here, among the hierarchical approaches, the AAA-GSA hierarchical approach was found to be the most successful and powerful.

AAA-GSA, which is the most successful hierarchical approach determined in the first part of the experimental study, was compared with the basic algorithms on CEC'17 test functions. The results of this comparison are given in Table 11 and Table 14 for the 10, 30, 50, and 100 dimensions, respectively.

Table 11. Mean results for D=10

F	AAA	ABC	DE	GA	GSA	PSO	AAA-GSA
f1	5,48E+02	5,16E+02	**1,00E+02**	1,75E+03	3,26E+06	2,23E+03	1,20E+03
f3	3,01E+02	7,02E+03	**3,00E+02**	3,49E+03	1,24E+04	**3,00E+02**	3,25E+02
f4	4,03E+02	**4,00E+02**	**4,00E+02**	4,14E+02	4,06E+02	4,03E+02	4,04E+02
f5	**5,05E+02**	5,07E+02	5,18E+02	5,23E+02	5,13E+02	5,16E+02	5,06E+02
f6	**6,00E+02**	**6,00E+02**	6,01E+02	**6,00E+02**	6,01E+02	**6,00E+02**	**6,00E+02**
f7	**7,14E+02**	7,17E+02	7,25E+02	7,37E+02	7,29E+02	7,17E+02	7,16E+02
f8	**8,05E+02**	8,07E+02	8,17E+02	8,15E+02	8,13E+02	8,12E+02	**8,05E+02**
f9	**9,00E+02**	**9,00E+02**	9,19E+02	9,53E+02	9,01E+02	**9,00E+02**	**9,00E+02**
f10	**1,21E+03**	1,24E+03	1,49E+03	1,84E+03	1,89E+03	1,59E+03	1,27E+03
f11	**1,10E+03**	1,11E+03	1,11E+03	1,12E+03	1,39E+03	1,12E+03	**1,10E+03**
f12	9,91E+03	4,30E+04	**1,39E+03**	1,47E+06	4,31E+05	1,01E+04	9,32E+03
f13	1,51E+03	2,34E+03	**1,32E+03**	7,86E+03	9,14E+03	6,95E+03	1,49E+03
f14	1,44E+03	1,65E+03	**1,41E+03**	3,07E+03	3,73E+03	1,46E+03	1,44E+03
f15	1,54E+03	1,61E+03	**1,51E+03**	4,44E+03	9,84E+03	1,56E+03	1,54E+03
f16	**1,60E+03**	1,61E+03	1,64E+03	1,78E+03	1,96E+03	1,85E+03	**1,60E+03**
f17	**1,70E+03**	**1,70E+03**	1,73E+03	1,73E+03	1,79E+03	1,75E+03	**1,70E+03**
f18	2,84E+03	3,14E+03	**1,83E+03**	1,47E+04	4,65E+03	6,79E+03	2,77E+03
f19	1,92E+03	2,00E+03	**1,90E+03**	7,22E+03	6,11E+03	2,16E+03	1,94E+03
f20	**2,00E+03**	**2,00E+03**	2,01E+03	2,03E+03	2,16E+03	2,07E+03	**2,00E+03**
f21	2,24E+03	**2,21E+03**	2,31E+03	2,30E+03	2,32E+03	2,30E+03	2,22E+03
f22	2,28E+03	**2,25E+03**	2,29E+03	2,31E+03	2,31E+03	2,35E+03	2,27E+03
f23	2,61E+03	**2,60E+03**	2,62E+03	2,64E+03	2,62E+03	2,62E+03	2,61E+03
f24	2,68E+03	**2,50E+03**	2,71E+03	2,72E+03	2,55E+03	2,72E+03	2,61E+03
f25	2,90E+03	**2,67E+03**	2,92E+03	2,94E+03	2,93E+03	2,92E+03	2,90E+03
f26	2,86E+03	**2,66E+03**	2,92E+03	3,09E+03	2,88E+03	2,93E+03	2,84E+03
f27	3,09E+03	**3,08E+03**	3,09E+03	3,14E+03	3,10E+03	3,11E+03	3,09E+03
f28	3,13E+03	**3,07E+03**	3,30E+03	3,28E+03	3,11E+03	3,28E+03	3,13E+03
f29	**3,14E+03**	3,16E+03	3,16E+03	3,23E+03	3,22E+03	3,23E+03	3,15E+03
f30	**6,04E+03**	7,23E+03	3,95E+05	8,45E+05	2,69E+05	2,77E+05	6,42E+03
Total	**12**	**13**	**9**	**1**	**0**	**3**	**7**

Table 12. Mean results for D=30

F	AAA	ABC	DE	GA	GSA	PSO	AAA-GSA
f1	4,14E+02	2,84E+02	**1,06E+02**	3,57E+03	2,11E+07	4,97E+03	4,24E+02
f3	1,25E+04	1,13E+05	**3,01E+02**	4,02E+04	8,61E+04	7,00E+02	2,42E+04
f4	4,57E+02	**4,32E+02**	4,62E+02	4,96E+02	4,97E+02	4,91E+02	4,59E+02
f5	5,48E+02	5,83E+02	5,84E+02	6,40E+02	6,20E+02	6,04E+02	**5,46E+02**
f6	**6,00E+02**	**6,00E+02**	6,09E+02	**6,00E+02**	6,03E+02	6,05E+02	**6,00E+02**
f7	7,76E+02	8,05E+02	8,67E+02	9,08E+02	8,58E+02	8,04E+02	**7,74E+02**
f8	**8,49E+02**	8,94E+02	8,88E+02	9,11E+02	8,95E+02	8,99E+02	8,51E+02
f9	**9,01E+02**	2,22E+03	2,37E+03	2,59E+03	9,07E+02	2,04E+03	**9,01E+02**
f10	**3,01E+03**	3,45E+03	4,03E+03	4,48E+03	3,91E+03	4,37E+03	3,03E+03
f11	**1,13E+03**	1,76E+03	1,20E+03	1,23E+03	3,11E+03	1,21E+03	1,14E+03
f12	4,04E+05	8,63E+05	**3,19E+04**	1,59E+06	1,80E+06	1,84E+05	4,79E+05
f13	6,13E+03	2,89E+04	**2,42E+03**	1,80E+04	3,35E+05	2,05E+04	6,62E+03
f14	2,08E+04	1,36E+05	**1,48E+03**	9,83E+05	1,39E+05	1,67E+04	1,63E+04
f15	1,97E+03	6,75E+03	**1,64E+03**	6,87E+03	4,33E+04	7,01E+03	2,87E+03
f16	**2,17E+03**	2,18E+03	2,50E+03	2,90E+03	2,81E+03	2,48E+03	2,25E+03
f17	**1,84E+03**	1,89E+03	2,15E+03	2,43E+03	2,64E+03	2,01E+03	**1,84E+03**
f18	1,02E+05	3,13E+05	**2,68E+03**	1,70E+06	2,55E+05	2,35E+05	1,31E+05
f19	3,47E+03	1,53E+04	**1,94E+03**	7,63E+03	1,18E+05	7,51E+03	3,20E+03
f20	2,18E+03	2,25E+03	2,36E+03	2,58E+03	2,86E+03	2,42E+03	**2,17E+03**
f21	2,35E+03	**2,31E+03**	2,38E+03	2,44E+03	2,39E+03	2,40E+03	2,35E+03
f22	2,61E+03	**2,32E+03**	4,65E+03	3,69E+03	**2,32E+03**	3,93E+03	2,43E+03
f23	2,71E+03	2,72E+03	2,74E+03	2,85E+03	2,72E+03	2,81E+03	**2,68E+03**
f24	2,90E+03	**2,73E+03**	2,91E+03	3,30E+03	2,88E+03	2,95E+03	2,90E+03
f25	2,89E+03	**2,88E+03**	2,89E+03	2,90E+03	2,89E+03	2,90E+03	2,89E+03
f26	3,45E+03	**2,90E+03**	4,63E+03	5,65E+03	2,92E+03	4,22E+03	3,19E+03
f27	**3,20E+03**	3,21E+03	3,21E+03	3,27E+03	3,26E+03	3,24E+03	**3,20E+03**
f28	**3,20E+03**	**3,20E+03**	**3,20E+03**	3,21E+03	**3,20E+03**	3,21E+03	**3,20E+03**
f29	**3,41E+03**	3,54E+03	3,71E+03	3,96E+03	4,14E+03	3,67E+03	**3,41E+03**
f30	6,31E+03	2,23E+04	**5,21E+03**	9,55E+03	4,23E+05	7,99E+03	7,25E+03
Total	**10**	**8**	**10**	**1**	**2**	**0**	**10**

Table 13. Mean results for D=50

F	AAA	ABC	DE	GA	GSA	PSO	AAA-GSA
f1	**1,34E+03**	5,48E+03	1,10E+05	2,15E+03	4,11E+07	2,58E+03	2,19E+03
f3	5,17E+04	2,19E+05	**1,42E+04**	5,10E+04	1,73E+05	2,51E+03	7,11E+04
f4	4,46E+02	**4,42E+02**	5,09E+02	5,38E+02	5,82E+02	5,53E+02	4,47E+02
f5	6,21E+02	7,09E+02	6,69E+02	7,62E+02	7,40E+02	7,14E+02	**6,17E+02**
f6	**6,00E+02**	**6,00E+02**	6,19E+02	**6,00E+02**	6,06E+02	6,20E+02	**6,00E+02**
f7	8,75E+02	9,25E+02	1,23E+03	1,08E+03	9,69E+02	9,49E+02	**8,66E+02**
f8	9,29E+02	1,00E+03	9,85E+02	1,07E+03	1,05E+03	1,02E+03	**9,11E+02**
f9	1,34E+03	9,37E+03	5,00E+03	7,74E+03	**9,14E+02**	1,20E+04	9,15E+02
f10	**4,69E+03**	5,78E+03	6,42E+03	6,72E+03	5,98E+03	7,11E+03	4,87E+03
f11	**1,20E+03**	3,12E+03	1,30E+03	1,48E+03	3,66E+03	1,29E+03	1,21E+03
f12	2,32E+06	5,91E+06	**2,63E+05**	1,57E+06	1,02E+07	2,50E+06	3,06E+06
f13	**3,60E+03**	2,11E+04	2,52E+04	6,73E+03	1,19E+06	6,33E+03	3,65E+03
f14	2,35E+05	1,18E+06	**2,63E+03**	1,07E+06	1,06E+05	7,40E+04	1,32E+05
f15	**5,58E+03**	2,00E+04	1,50E+04	1,01E+04	1,47E+05	1,01E+04	6,85E+03
f16	**2,78E+03**	2,83E+03	3,31E+03	3,61E+03	3,47E+03	2,94E+03	2,80E+03
f17	2,45E+03	2,66E+03	2,89E+03	3,41E+03	3,45E+03	3,01E+03	**2,43E+03**
f18	6,35E+05	1,55E+06	**6,66E+03**	1,72E+06	7,29E+05	1,33E+06	6,20E+05
f19	7,78E+03	2,92E+04	**3,51E+03**	2,06E+04	1,96E+05	1,69E+04	7,82E+03
f20	2,98E+03	2,84E+03	2,96E+03	3,17E+03	3,22E+03	2,89E+03	**2,59E+03**
f21	2,43E+03	2,47E+03	2,47E+03	2,60E+03	2,52E+03	2,54E+03	**2,42E+03**
f22	6,73E+03	**6,16E+03**	8,24E+03	8,60E+03	9,38E+03	9,04E+03	6,23E+03
f23	**2,87E+03**	2,92E+03	2,92E+03	3,17E+03	3,08E+03	3,03E+03	**2,87E+03**
f24	3,12E+03	3,40E+03	**3,07E+03**	4,10E+03	3,17E+03	3,20E+03	3,12E+03
f25	**3,00E+03**	3,01E+03	3,04E+03	3,08E+03	3,07E+03	3,06E+03	3,01E+03
f26	5,15E+03	4,24E+03	5,90E+03	8,61E+03	**2,94E+03**	5,36E+03	4,77E+03
f27	3,29E+03	3,34E+03	3,38E+03	3,72E+03	3,72E+03	3,46E+03	**3,27E+03**
f28	3,29E+03	3,28E+03	3,29E+03	3,32E+03	3,31E+03	3,31E+03	**3,27E+03**
f29	3,65E+03	3,90E+03	4,20E+03	4,44E+03	4,65E+03	4,47E+03	**3,64E+03**
f30	**6,79E+05**	8,98E+05	7,89E+05	9,51E+05	1,18E+07	9,49E+05	6,91E+05
Total	**10**	**3**	**6**	**1**	**2**	**0**	**11**

Table 14. Mean results for D=100

F	AAA	ABC	DE	GA	GSA	PSO	AAA-GSA
f1	**1,55E+03**	7,09E+03	9,00E+08	5,28E+03	8,87E+07	3,72E+05	2,14E+03
f3	2,60E+05	5,45E+05	4,07E+05	2,27E+04	3,51E+05	**1,94E+04**	2,92E+05
f4	6,14E+02	**6,08E+02**	7,01E+02	6,41E+02	6,44E+02	7,52E+02	6,18E+02
f5	9,22E+02	1,19E+03	9,94E+02	1,17E+03	1,16E+03	1,11E+03	**9,10E+02**
f6	**6,00E+02**	**6,00E+02**	6,33E+02	**6,00E+02**	6,11E+02	6,50E+02	**6,00E+02**
f7	1,20E+03	1,44E+03	3,08E+03	1,85E+03	1,21E+03	1,39E+03	**1,18E+03**
f8	**1,22E+03**	1,53E+03	1,31E+03	1,50E+03	1,53E+03	1,40E+03	**1,22E+03**
f9	2,04E+04	5,00E+04	**1,44E+04**	1,93E+04	2,30E+03	5,18E+04	7,13E+03
f10	1,15E+04	1,33E+04	1,41E+04	1,39E+04	1,30E+04	1,49E+04	**1,12E+04**
f11	1,09E+04	7,14E+04	**1,82E+03**	6,61E+03	1,72E+05	2,45E+03	1,72E+04
f12	9,54E+06	3,26E+07	**2,98E+06**	4,63E+06	2,79E+07	1,80E+07	1,21E+07
f13	3,48E+03	1,52E+04	1,25E+04	8,96E+03	1,47E+06	1,01E+04	**3,38E+03**
f14	1,34E+06	1,00E+07	**2,23E+04**	9,09E+05	3,98E+05	8,10E+05	9,58E+05
f15	2,11E+03	1,62E+04	2,18E+04	3,71E+03	1,92E+05	4,88E+03	**2,00E+03**
f16	4,76E+03	4,93E+03	5,63E+03	6,01E+03	6,60E+03	5,12E+03	**4,65E+03**
f17	**3,95E+03**	4,52E+03	5,18E+03	5,43E+03	4,75E+03	4,88E+03	3,99E+03
f18	1,69E+06	6,90E+06	**1,55E+05**	7,58E+05	5,59E+05	2,71E+06	1,63E+06
f19	3,21E+03	1,06E+05	2,56E+05	4,41E+03	3,97E+05	5,35E+03	**2,40E+03**
f20	4,15E+03	4,88E+03	4,62E+03	5,23E+03	5,64E+03	5,04E+03	**4,08E+03**
f21	2,75E+03	3,00E+03	2,86E+03	3,09E+03	2,95E+03	3,01E+03	**2,71E+03**
f22	1,40E+04	1,63E+04	1,63E+04	1,69E+04	1,81E+04	1,81E+04	**1,37E+04**
f23	3,10E+03	3,17E+03	3,37E+03	3,47E+03	4,72E+03	3,73E+03	**3,08E+03**
f24	3,70E+03	3,87E+03	3,90E+03	4,27E+03	**3,54E+03**	4,17E+03	3,68E+03
f25	**3,20E+03**	3,23E+03	3,34E+03	3,32E+03	3,25E+03	3,38E+03	3,21E+03
f26	9,78E+03	1,17E+04	1,27E+04	1,90E+04	**4,38E+03**	1,25E+04	9,84E+03
f27	**3,38E+03**	3,46E+03	3,53E+03	3,73E+03	3,95E+03	3,75E+03	**3,38E+03**
f28	**3,35E+03**	3,38E+03	5,97E+03	3,37E+03	3,40E+03	3,48E+03	3,36E+03
f29	5,72E+03	6,92E+03	6,59E+03	7,07E+03	7,19E+03	6,93E+03	**5,53E+03**
f30	**8,73E+03**	2,33E+04	2,14E+04	1,17E+04	2,04E+06	1,11E+04	1,02E+04
Total	**8**	**2**	**5**	**1**	**2**	**1**	**15**

Table 15. The overall results of the hierarchical approaches

			AAA	ABC	DE	GA	GSA	PSO	AAA-GSA
D = 10		***Best***	12	13	9	1	0	3	7
	Friedman Test	***Mean Rank***	**2.3793**	2.7069	3.7759	6.0517	5.6897	4.8966	2.5000
		Final Rank	**1**	3	4	7	6	5	2
		p-Value	**4.6991e-18**						
D = 30		***Best***	10	8	10	1	2	0	10
	Friedman Test	***Mean Rank***	**2.3103**	3.5172	3.5517	6.0517	5.3276	4.7931	2.4483
		Final Rank	1	3	4	7	6	5	2
		p-Value	**6.0272e-15**						
D = 50		***Best***	10	3	6	1	2	0	11
	Friedman Test	***Mean Rank***	2.2414	4.0000	3.7414	5.5000	5.6552	4.6897	**2.1724**
		Final Rank	2	4	3	6	7	5	**1**
		p-Value	**3.4359e-14**						
D = 100		***Best***	8	2	5	1	2	1	15
	Friedman Test	***Mean Rank***	2.3276	4.6724	4.4310	4.5345	5.0000	4.9828	**2.0517**
		Final Rank	2	5	3	4	7	6	**1**
		p-Value	**5.7557e-11**						

Looking at the comparison of the AAA-GSA hierarchical approach in 10 dimensions and the basic states of the algorithms in Table 11; the AAA and ABC algorithms were the most successful. The AAA and ABC algorithms have proven their power in low dimensions. As seen in Table 12 the comparison of the AAA-GSA hierarchical approach in 30 dimensions and the basic states of the algorithms; the AAA-GSA hierarchical approach and the AAA, DE algorithms have been most successful. In Table 13 for 50 dimensions; the AAA-GSA hierarchical approach has been most successful. In 100 dimensions the comparison of the AAA-GSA hierarchical approach and the basic states of the algorithms in Table 14; the AAA-GSA hierarchical approach was again the most successful. With the increase in the size of the test functions and the difficulty of the problem, the AAA-GSA hierarchical approach has managed to maintain its success. Comparing the AAA-GSA hierarchical approach with the basic versions of the algorithms, the following conclusion can be drawn; the AAA-GSA hierarchical approach developed the AAA and GSA algorithms.

The average rank, final rank, and p-Value values obtained in the Friedman rank test results of the most successful hierarchical approaches and the basic states of the algorithms are given in Table 15. As seen in Table 15, the method with the lowest average rank is the most successful method. Looking at the Friedman test results in Table 15; while the AAA algorithm in 10 and 30 dimensions was the first in the ranking, the AAA-GSA hierarchical approach in 50 and 100 dimensions was the first in the ranking. As seen in Table 15, each p-Value is less than 0.05. This indicates that there is a statistically significant difference between the results. The fact that the AAA-GSA hierarchical approach is most successful in the dimensions where the problem becomes more difficult shows that this approach is strong and stable.

6. CONCLUSION

In this study, hierarchical structures are proposed based on the results of the AAA, ABC, DE, GA, GSA, and PSO algorithms on CEC'17 test functions. The results of the proposed hierarchical structures on the same test functions are examined, and the singular states of the algorithms are compared with these hierarchical models. When the results among the proposed hierarchical approaches are examined, the AAA-GSA hierarchical approach has been the most successful approach, keeping its success especially as the problem gets difficult. Then, since the AAA-GSA hierarchical approach was the most successful approach, the success of this approach and the basic states of the algorithms were compared. When the Friedman test results are also examined, it shows that there is a statistically significant distinction between the results. It is seen that the hierarchical structure of AAA-GSA generally has the first rank in the Friedman test results. It is seen from these results that; The AAA-GSA approach has been a very powerful approach to difficult problems. The AAA-GSA approach has improved these algorithms by surpassing the successes of AAA and GSA algorithms. The results of the proposed AAA-GSA approach are quite promising. But it is clear that all hierarchical approaches need further refinement.

Author Contributions: Ferda Nur Arıcı: Concept, Literature review, Data collection, Software, Writing-Review & Editing, Analysis of results, Ersin Kaya: Software, Writing-Review & Editing, Analysis of results, Investigation.
Conflict of Interest: No conflict of interest was declared by the authors.
Financial Disclosure: The authors declared that this study had received no financial support.

REFERENCES

[1] X.-S. Yang, Nature-inspired metaheuristic algorithms: Luniver press, 2010.

[2] M. S. Kıran, "Optimizasyon problemlerinin çözümü için yapay arı kolonisi algoritması tabanlı yeni yaklaşımlar," Selçuk Üniversitesi Fen Bilimleri Enstitüsü, 2014.

[3] S. A. Uymaz, "Yeni bir biyolojik ilhamlı metasezgisel optimizasyon metodu: Yapay alg

algoritması," Selçuk Üniversitesi Fen Bilimleri Enstitüsü, 2015.

[4] F. Glover and M. Laguna, "Tabu search," in Handbook of combinatorial optimization, ed: Springer, 1998, pp. 2093-2229.

[5] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Technical report-tr06, Erciyes university, engineering faculty, computer …2005.

[6] J. Kennedy and R. Eberhart, "Particle swarm optimization," in Proceedings of ICNN'95-International Conference on Neural Networks, 1995, pp. 1942-1948.

[7] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: optimization by a colony of cooperating agents," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 26, pp. 29-41, 1996.

[8] P. J. Angeline, "Evolution revolution: An introduction to the special track on genetic and evolutionary programming," IEEE Intelligent Systems, pp. 6-10, 1995.

[9] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," Journal of global optimization, vol. 11, pp. 341-359, 1997.

[10] F. N. Arıcı and E. Kaya, "Comparison and characterization of meta-heuristic algorithms on benchmark functions," Academic Perspective Procedia, vol. 2, pp. 508-517, 2019 2019.

[11] H. Haklı, "Sürekli fonksiyonların optimizasyonu için doğa esinli algoritmaların geliştirilmesi," Selçuk Üniversitesi Fen Bilimleri Enstitüsü, 2013.

[12] J. Robinson, S. Sinton, and Y. Rahmat-Samii, "Particle swarm, genetic algorithm, and their hybrids: optimization of a profiled corrugated horn antenna," in IEEE Antennas and Propagation Society International Symposium (IEEE Cat. No. 02CH37313), 2002, pp. 314-317.

[13] L. Bellatreche, K. Boukhalfa, and H. I. Abdalla, "Saga: A combination of genetic and simulated annealing algorithms for physical data warehouse design," in British National Conference on Databases, 2006, pp. 212-219.

[14] M. H. Moradi and M. Abedini, "A combination of genetic algorithm and particle swarm optimization for optimal DG location and sizing in distribution systems," International Journal of Electrical Power & Energy Systems, vol. 34, pp. 66-74, 2012.

[15] S. Arunachalam, T. AgnesBhomila, and M. R. Babu, "Hybrid particle swarm optimization algorithm and firefly algorithm based combined economic and emission dispatch including valve point effect," in International Conference on Swarm, Evolutionary, and Memetic Computing, 2014, pp. 647-660.

[16] H. Eldem, "Karınca Koloni Optimizasyonu (KKO) ve Parçacık Sürü Optimizasyonu (PSO) Algortimaları Temelli Bir Hiyerarşik Yaklaşım Geliştirilmesi," Yüksek Lisans, Bilgisayar Mühendisliği, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, 2014.

[17] O. Gokalp and A. Uğur, "An order based hybrid metaheuristic algorithm for solving optimization problems," in 2017 International Conference on Computer Science and Engineering (UBMK), 2017, pp. 604-609.

[18] P. J. Gaidhane and M. J. Nigam, "A hybrid grey wolf optimizer and artificial bee colony algorithm for enhancing the performance of complex systems," Journal of computational science, vol. 27, pp. 284-302, 2018.

[19] G.-H. Lin, J. Zhang, and Z.-H. Liu, "Hybrid particle swarm optimization with differential evolution for numerical and engineering optimization," International Journal of Automation and Computing, vol. 15, pp. 103-114, 2018.

[20] S. Jiang, C. Zhang, and S. Chen, "Sequential Hybrid Particle Swarm Optimization and Gravitational Search Algorithm with Dependent Random Coefficients," Mathematical Problems in Engineering, vol. 2020, 2020.

[21] S. Sharma and S. Ghosh, "FIS and hybrid ABC-PSO based optimal capacitor placement and sizing for radial distribution networks," Journal of Ambient Intelligence and Humanized Computing, vol. 11, pp. 901-916, 2020.

[22] M. Karakoyun, A. Ozkis, and H. Kodaz, "A new algorithm based on gray wolf optimizer and shuffled frog leaping algorithm to solve the multi-objective optimization problems," Applied Soft Computing, vol. 96, p. 106560, 2020.

[23] A. Dixit, A. Mani, and R. Bansal, "CoV2-Detect-Net: Design of COVID-19 prediction model based on hybrid DE-PSO with SVM using chest X-ray images," Information sciences, vol. 571, pp. 676-692, 2021.

[24] S. M. Kisengeu, C. M. Muriithi, and G. N. Nyakoe, "Under voltage load shedding using hybrid ABC-PSO algorithm for voltage stability enhancement," Heliyon, vol. 7, p. e08138, 2021.

[25] R. P. Parouha and P. Verma, "An innovative hybrid algorithm for bound-unconstrained optimization problems and applications," Journal of Intelligent Manufacturing, pp. 1-64, 2021.

[26] X. Zhang, Z. Wang, and Z. Lu, "Multi-objective load dispatch for microgrid with electric vehicles using modified gravitational search and particle swarm optimization algorithm," Applied Energy, vol. 306, p. 118018, 2022.

[27] Y. Li, X. Song, and W. Guan, "Mobile robot path planning based on ABC-PSO algorithm," in 2022 IEEE 6th Information Technology and

Mechatronics Engineering Conference (ITOEC), 2022, pp. 530-534.

[28] B. Turkoglu, S. A. Uymaz, and E. Kaya, "Binary Artificial Algae Algorithm for feature selection," Applied Soft Computing, vol. 120, p. 108630, 2022.

[29] B. Turkoglu, S. A. Uymaz, and E. Kaya, "Clustering analysis through artificial algae algorithm," International Journal of Machine Learning and Cybernetics, vol. 13, pp. 1179-1196, 2022.

[30] E. Kaya, "BinGSO: galactic swarm optimization powered by binary artificial algae algorithm for solving uncapacitated facility location problems," Neural Computing and Applications, pp. 1-20, 2022.

[31] T. Keskintürk, "Diferansiyel gelişim algoritması," 2006.

[32] D. Karaboğa, Yapay Zeka Optimizasyon Algoritmalari: Nobel Akademi Yayıncılık, 2017.

[33] G. G. Emel and Ç. Taşkın, "Genetik Algoritmalar ve Uygulama Alanlari," Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, vol. 21, pp. 129-152, 2002.

[34] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "GSA: a gravitational search algorithm," Information sciences, vol. 179, pp. 2232-2248, 2009.

[35] M. Y. ÖZSAĞLAM and M. ÇUNKAŞ, "Optimizasyon problemlerinin çözümü için parçaçık sürü optimizasyonu algoritması," Politeknik Dergisi, vol. 11, pp. 299-305, 2008.

[36] S. ÇINAROĞLU and H. Bulut, "K-ortalamalar ve parçacık sürü optimizasyonu tabanlı kümeleme algoritmaları için yeni ilklendirme yaklaşımları," Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, vol. 33, pp. 413-424, 2018.

[37] N. H. Awad, M. Z. Ali, P. N. Suganthan, J. J. Liang, and B. Y. Qu, "Problem Definitions and Evaluation Criteria for the CEC 2017 Special Session and Competition on Single Objective Real-Parameter Numerical Optimization," 2016.

[38] M. Friedman, "A comparison of alternative tests of significance for the problem of m rankings," The Annals of Mathematical Statistics, vol. 11, pp. 86-92, 1940.

Research Paper
Academic Platform Journal of Engineering and Smart Systems (APJESS) 10(3), 140-148, 2022
ISSN: 2822-2385
Received: 12-Dec-2021 Accepted: 28-Jul-2022
https://doi.org/10.21541/apjess.1035793

Assessment of Wind Energy Potential and Current Usage Status in Türkiye and in the World

*[1]Faruk KÖSE, [2]Suleyman KOSE

[1] Department of Mechanical Engineering, Konya Technical University, Konya, Türkiye, fkose@ktun.edu.tr
[2] Energy 2050, Department of Mechanical Engineering, University of Sheffield, Sheffield, UK, skose1@sheffield.ac.uk

Abstract

While many developed countries in the world set 2050, Türkiye has determined 2053 as the year of net zero emissions target. Increase in the usage of renewable energy sources such as wind for the energy production is one of the main methods to achieve this goal. Among the renewable energy sources used for electricity production, wind energy has been the most increasing energy source in Türkiye and around the world in recent years. Today, more than 100 countries in the world generate electricity from wind energy. China, the United States of America and Germany are the top 3 countries with the highest installed capacity of wind energy for a long time. Based on 2022 data, Türkiye ranks 7th in Europe and 12th in the world with respect to the installed wind power plant capacity. Among the renewable energy sources, wind energy in Türkiye has the largest installed power and electricity production after hydroelectricity. In this study, the potential and installed capacity of the wind energy for the different countries in the world and Türkiye are analysed, the shares of the wind energy in current energy production and its contributions to the total energy consumption has been examined. The conditions that will enable the world including Türkiye to meet the energy consumption completely from renewable energy sources and the possible contribution of wind energy to all renewable resources have been estimated.

Keywords: Energy, Renewable Energy, Türkiye Wind Energy Potential, Wind Energy, World Wind Energy Potential.

1. INTRODUCTION

The rapid increase in the world population, industrialization and the intensive use of technological tools and equipment cause energy consumption to increase by 4-5% on average every year [1]. Environmental and health problems have occurred as a result of the long-term use of fossil-based energy sources, which are used to meet current and increasing energy needs. In addition, although there has not been much change in the reserve life of fossil energy resources (estimated reserve life 42-50 years for oil, 60-65 years for natural gas and 150-200 years for coal) in the last two decades, there is a concern that these resources will be depleted within 50-200 years. In recent years, this has accelerated the efforts to reduce the use of fossil-based energy sources and to find clean energy sources to replace them [2-4]. Within the scope of these studies, the use of natural gas, which is the least polluting and harmful among the other fossil fuels, has been increased as a near-term target, while the efforts to increase the use of renewable energy sources, which have no harmful emissions, have accelerated. It is known that the use of renewable energy sources such as solar, wind, geothermal and biomass for drying, grain grinding, heating and health purposes as conventional method, apart from electricity generation, is as old as human history. Today, utilization from renewable energy sources for electricity generation, which is the most effective form of their usage, is aimed to produce cleaner energy by reducing the usage of fossil resources.

The most widely engaged renewable energy sources; hydraulic, wind, solar, geothermal and biomass energies at the present time and their usage is increasing regularly. Wind energy, one of the most important renewable energy sources, occurs as a result of the different heating of the ground surfaces by the sun's rays. The difference in heating of the ground surfaces leads to a difference in the temperature, pressure and humidity of the air and the pressure differences seen between the two different regions gives rise to the movement of air from high pressure to low pressure, leading to wind formation. Some advantages of wind energy are that wind energy is renewable and clean, does not require very high technology for energy production in some systems, and is free and abundant in the atmosphere. In order to make the most effective use of wind energy, which will exist as long as the sun and the world exist, it must be converted into

* Corresponding Author

electrical energy. Wind turbines are used for this purpose [1, 2]. In wind turbines, the kinetic energy of the wind is first converted to mechanical energy by the turbine blades and then to electrical energy by the generator. In recent years, developments in propeller type wind turbine technology (such as gearless turbines, with increased unit power and efficiency) have made wind energy competitive with fossil fuel energies by reducing the cost of electricity generation from wind energy.

Many studies are found in the literature on the situation of wind energy in Türkiye and in the world. Senel et al. [1] examined the wind energy potential in Türkiye and the world, in addition to the development and usage status of wind energy systems as of the end of 2014. Kose [2] examined the formation and characteristic features of wind energy with Türkiye's wind energy potential. Köse et al. [5] performed an analysis by using the wind energy measurement data which was obtained from measurements performed at the Alaaddin Keykubat Campus of Konya Selcuk University and annual water data regarding Konya Altinapa dam. Their calculation showed that a wind power plant built in the region of Konya Akyokus and a water turbine power plant installed on the Akyokus purification plant can meet the annual electrical energy demand of the plant.

In the report published by the International Renewable Energy Agency (IRENA), the statistical information about world total and each country’s separately renewable energy potentials and their current situations associated with renewable energy usage for the year 2021 and before are given [6]. In the report published by Lee et al. [7] the statistical information about world total and each countries separate wind energy potentials and their current situations associated with wind energy usage for year of 2020 and previous years are shared.

According to the BP annual statistical report, electricity generation values for all countries of the world and the world total from 1985 to the end of 2020 using wind energy use are given [8]. Kose et al. [9] investigated the most suitable wind turbine and turbine tower height that can be built in the Konya Selcuk University Alaeddin Keykubat campus region based on their wind energy measurements performed in this region. Their analysis showed that the proposed wind turbine would pay itself back in 6,5 years.

Karabag et al. [10] examined the current and future status of renewable energy resources in light of large databases provided by national and international renewable energy institutions. They evaluated the latest situation in the world and Türkiye in the transition to 100% renewable energy, and made predictions about how close to the target of 100% renewable energy the transition is.

The government's purchase prices in $cent/kWh for licensed electricity produced in Türkiye from renewable energy sources according to the regulation are given in Table 1. [11]. The electricity produced by the facilities within the scope of unlicensed electricity generation activities from the same sources has been started to be purchased by the Republic of Türkiye Energy Market Regulatory Authority (EMRA) as of 10/5/2019. This purchasing is in TL kuruş/kWh based on the retail one-time active energy price of its own subscriber group. [11]. As seen in Table 1, considering the costs and efficiencies of the production systems, the electricity produced is purchased with the tariff of 7,3$cent/kWh, 7,3$cent/kWh, 10,5$cent/kWh, 13,3$cent/kWh, 13,3$cent/kWh from the electricity production facilities based on hydroelectric, wind, geothermal, biomass and solar energy respectively.

Table 1. The government’s electricity purchase prices of licensed electricity produced from renewable energies sources according to the regulation in Türkiye [11]

Production Facility Type based on Renewable Energy Source	Price Tariff (US Dollar cent/kWh)
Hydroelectric enrgy generation facility	7,3
Production facility based on wind energy	7,3
Production facility based on geothermal energy	10,5
Biomass based production facility (including landfill gas)	13,3
Solar energy based production facility	13,3

Another practice in Türkiye that encourages electricity production from the renewable energy sources including wind is that an additional incentive fee is paid for each kWh electricity produced during the first 5 years of operation of the facility, if each system component used in these systems is domestically produced [11, 12]. The additional incentive fees paid for wind turbine parts by parts is given in Table 2, and the highest fee is 1.3 USD cents/kWh if all the mechanical parts in the rotor and nacelle groups are domestic production. With such additional incentives, domestic production of the entire system is aimed, starting from the parts of the turbine.

Table 2. Incentive fees to be paid depending on the parts in case the wind turbine parts are domestically produced in Türkiye [11, 12]

Domestic Manufacturing	Incentive Fees for Domestic production (US Dollar cent/kWh)
1- Blade	0,8
2- Generator and power electronics	1,0
3- Turbine tower	0,6
4- All mechanical parts in rotor and nacelle groups	1,3

The current study intends to systematically analyse the current situation of wind energy usage in the world and in Türkiye. The outline of the paper is as follows: world wind energy potential and its usage examined in detail in Section 2. Türkiye’s wind energy potential and to what extent this potential is being utilized are discussed in Section 3. The conclusion is provided in Section 4.

2. WORLD WIND ENERGY POTENTIAL AND ITS USAGE

Before examining the world wind energy potential and its current usage status, the world's total current energy use and the shares of different energy sources used for energy production should be examined. In Figure 1, between 1990 and 2019, the world's total energy uses are shown at 5-year intervals [11]. Based on the data provided by the International Energy Agency (IEA), this chart shows the highest energy supply is from oil with 187.364.800 TJ in 2019; coal with 162.375.732 TJ, third: natural gas with 140.784.380 TJ, fourth: biofuel and waste with 56.813.210 TJ, fifth: nuclear with 30.461.171 TJ, sixth: hydraulic power with 13.194.639 TJ and seventh: wind and solar with 13.417.236 TJ. As a result, the sum of all of them 606.411.168 TJ, is the world's total energy consumption in 2019 [11]. The last three of these are renewable energies and they have a 14,1% share of the total, while the total of wind and solar power have a 2,21% share of the total.

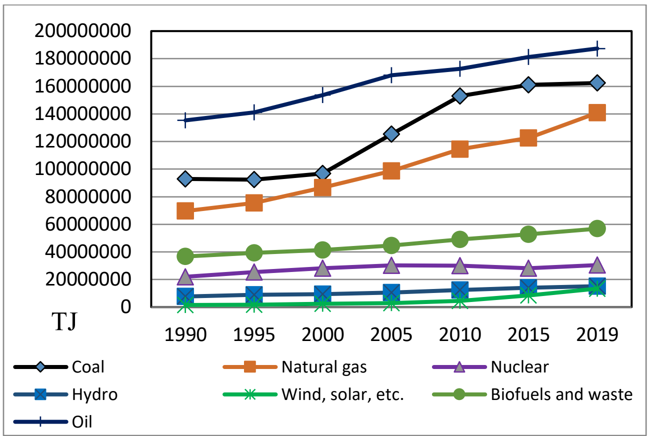


Figure 1. World 1990-2019 total energy supply (TJ) [11]

The values of the total and resource-based distribution of world primary energy consumption for 2019 and according to three different scenarios for 2050 are given in Table 3. [13]. According to this table, the total primary energy consumption, which was be 627 EJ in 2019, will have lowest value of 653 EJ based on the net zero emission scenario, 692 EJ based on the accelerated scenario, and highest value of 760 EJ based on the new momentum scenario created considering countries with new economies growing. With respect to the primary energy sources type, the lowest possible values also are obtained with net zero emissions scenario for the fossil fuels; Compared to 2019 values, it is aimed to decrease to 22,8% in oil, 43,6% in natural gas and 10,7% in coal.

Considering the data in Table 3, it is estimated that energy production from renewable energy sources will increase 1.71 times in hydroelectricity and 5,65 times in renewable energies including biomass in 2050 compared to 2019. In the same period, it is aimed that nuclear energy production will increase by 2 times and 82% of the total primary energy can be produced from nuclear and renewable energies.

Table 3. World primary energy by fuel, consumption level for 2019 and demands of 2050 with different scenarios, EJ [13]

Primary energy by fuel	2019 (EJ)	Accele rated	Net Zero	New Momentum
Total	**627**	**692**	**653**	**760**
Oil	193	87	44	154
Natural gas	140	94	61	181
Coal	158	25	17	103
Nuclear	25	40	49	27
Hydro	38	61	65	48
Renewables (incl. bioenergy)	74	384	418	247

Various studies have been carried out by the International Energy Agency (IEA) in order to determine the world wind energy potential. In these studies, the world technical wind potential was calculated as 53.000 TWh/year, based on the prediction that 4% of the regions with a wind capacity above 5,1 m/s would be used due to practical and social constraints [1]. In Figure 2, the wind speed values for the places which have 80m above the ground are given in m/s and miles/h units. [14]. Likewise, Figure 3, the comparison of technical wind potentials of world continents are illustrated [1]. Wind energy potential is highest in North America with 14.000 TWh/year and lowest in Oceania with 3.000 TWh/year. These data show that North America, Eastern Europe, Russia and Africa have 66% of the world wind energy potential.

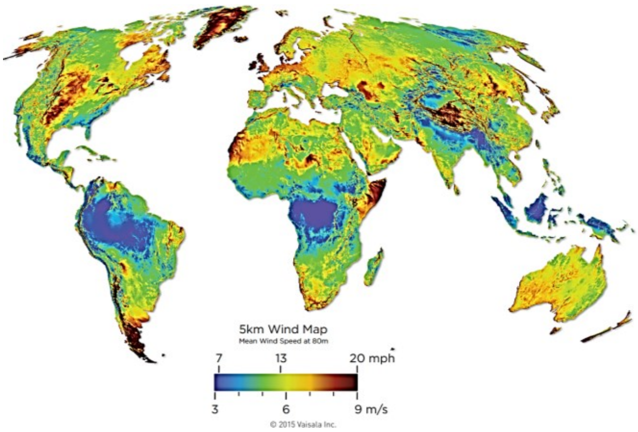


Figure 2. Technical wind potentials of world continents [14]

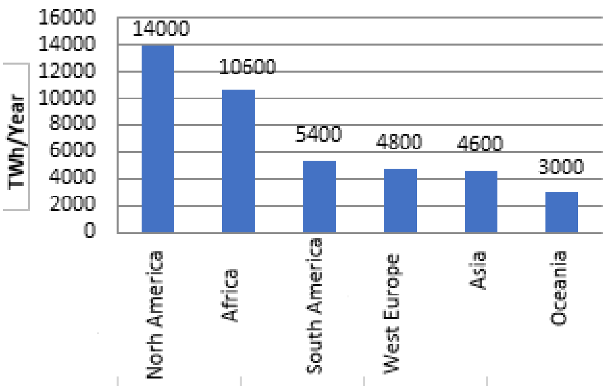
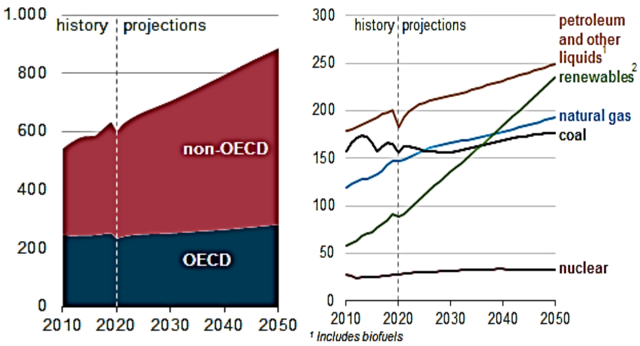


Figure 3. Technical wind potentials of world continents [1]

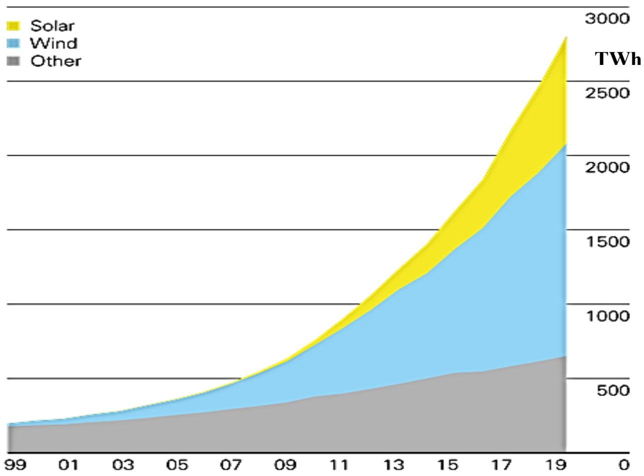
According to Lee et al. [7] as can be seen in Figure 4 a), by 2050, % 50 increases in global energy consumption is expected due to non–OECD economic growth and population increases. Furthermore, they claim that although liquid fuels are the largest primary source for energy production in the reference state, energy production from renewable sources also reaches about the same levels. In Figure 4 (b), electricity generation from renewable sources has been converted to Btu at a rate of 8.124 Btu/kWh. When the total energy consumption of OECD member and non-member countries is analysed as it can be seen in Figure 4 (a), it is noticed that energy consumption of non-OECD countries will increase tremendously due to increasing population and industrialization.



(a) (b) [2] *8,124 Btu/kWh*

Figure 4. 2010-2050 (a) world energy consumption, (b) world primary energy consumption by energy source, (quadrillion BTU) [7]

Total world electricity generation from wind power by the end of 2020 was 1591,2 TWh/year [6]. This value is approximately 3% of the world wind capacity given in Figure 3 and 5,3 times the electricity consumption of Türkiye, which was 305 billion kWh/year in 2020. World total renewable energy production between 1999 and 2019 is given in Figure 5. Accordingly, it is seen that energy production from wind power has exceeded 50% of total renewable energy production in recent years.



Year (1999-2019

Figure 5. Renewable energy generation by source between 1999-2019, (TWh) [8]

For the global wind industry, 2020 was the best year in history with 53% annual growth. More than 93 GW of wind power was installed in a challenging year disrupting both the global supply chain and project processes [7]. World total wind power increased to 743 GW with 93 GW of new installations in 2020. This equals 1.1 billion tons of CO_2 reduction per year.

The installed wind energy capacity of the world has increased 3 times in the last 10 years, reaching from 283 GW to 837 GW as illustrated in Figure 6 [13], which has made wind energy one of the most cost competitive and durable energy sources.

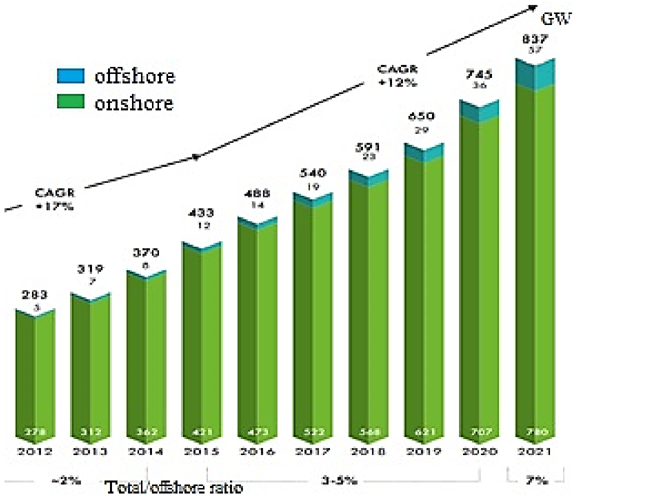


Figure 6. Development in the world total wind installed power over a 10-year period (2012-2021) [13]

Regionally, records have been performed in terms of onshore installations in Asia Pacific North America and Latin America in 2020-2021. As of 2022, a total of 55 GW of new onshore wind energy capacity has been gained from these three regions [13].

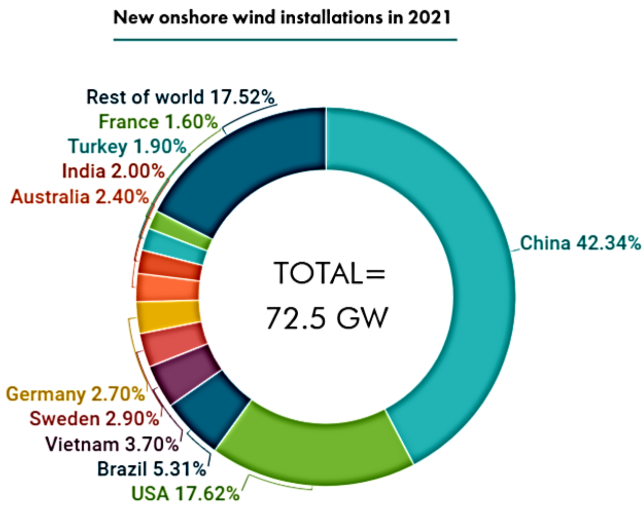


Figure 7. New onshore wind power installation rates by countries for 2021 [13]

As can be seen from Figure 7, when we look at the new wind installations, Türkiye is seen among the top 10 countries on a global scale. In Türkiye, there was a new wind capacity increase of 686 MW in 2019, and the increase in 2020 almost doubled the previous year with 1224 MW. The total wind capacity in Türkiye increased to 9.559 MW in 2020 and then in 2021 10.681 MW. As an offshore wind farm, 21,1 GW of worldwide capacity was commissioned last year, making 2021 the best year ever as shown in Figure 8. China has broken the record, achieving 80% all new world offshore wind power installations. Steady growth was recorded in Europe. The United Kingdom took the lead, followed by Denmark and Netherlands. The remaining new offshore wind installations in 2021 were made by the Vietnam and rest of the world. World offshore wind installation exceeded 57 GW in 2021 and represents 7% of total world wind capacity [13].

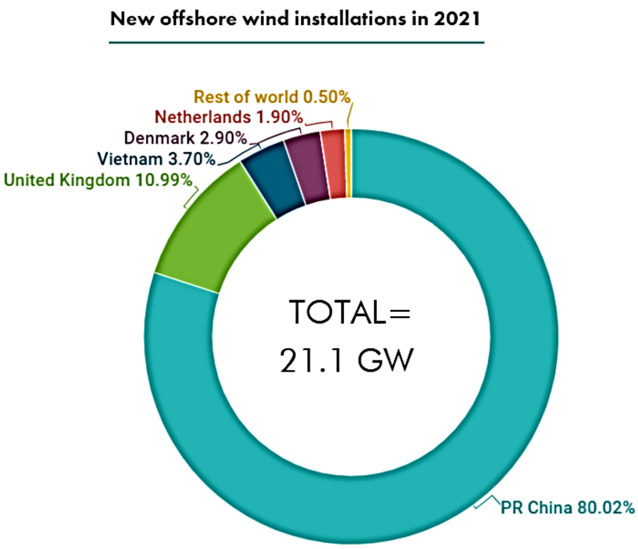


Figure 8. New offshore wind power installation rates by countries for 2021 [13]

The world and European wind installed power rankings taken into account considering 2022 values, China, USA and Germany are in the first three places in the world, respectively, and Germany, Spain and England are in the top three rankings in the European ranking. Türkiye ranks 7th in Europe and 12th in the world with its 10.681 MW of installed power as presented in Table 4 [3, 6, 7, 13].

Table 4. World and European wind installed power ranking

Rank	World Ranking Country	Installed Power (MW)	European Ranking Country	Installed Power (MW)
1	China	338.309	Germany	64.542
2	US	134.396	Spain	27.089
3	Germany	64.542	England	26.586
4	India	40.084	France	19.131
5	Spain	27.089	Sweden	11.915
6	England	26.586	Italy	10.839
7	France	21.580	**Türkiye**	**10.681**
8	Brazil	19.131	Netherland	6.992
9	Canada	14.255	Denmark	6.840
10	Sweden	11.915	Portugal	5.239
11	Italy	10.839	Belgium	2.843
12	**Türkiye**	**10.681**	Ireland	2.830
	Total	**719.407**	**Total**	**195.527**

Some scenarios and plans have been made by some international organizations such as IRENA, IEA and EWEA to produce the world's energy need from clean and renewable energy sources instead of fossil-based sources that disrupt the ecological balance and harm human health. Some of these are based on the 2015 Paris agreement targets, some are based on 2030 and some are based on 2050 [6-8]. According to IRENA's Energy Transition Scenario, world wind additional annual capacity increase would need to reach approximately 180 GW to meet the Paris agreement targets [6]. According to the IEA's Net Zero scenario by 2050, it needs to increase to 160 GW by 2025 and to 280 GW by 2030. This value corresponds to 3 times the increase in 2020 as illustrated in Figure 9. Total annual global investment in clean energy and efficient systems infrastructure should increase from US$380 billion in 2020 to US$1,6 trillion in 2030, according to the IEA [15].

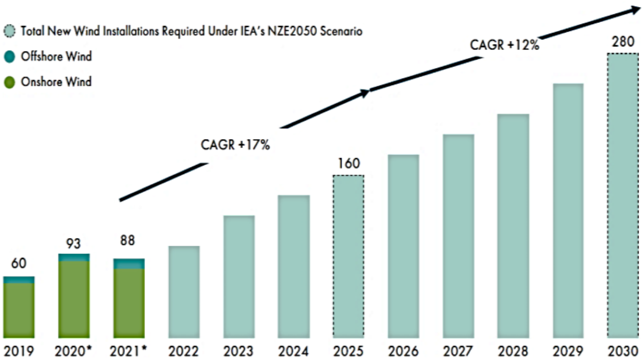


Figure 9. New global wind energy installations required by 2030 (GW) (CAGR: compound annual growth rate) [8]

To reach the net zero targets, the wind market expects the addition of new onshore and offshore wind power with a capacity of over 469 GW in the next five years. That indicates around 94 GW of new installations per year by 2025, according to current policies [8]. Governments are expected to significantly increase their targets following COP26 (26th UN Climate Change Conference of the Parties) [17]. Combined with China's net zero target by 2060 and the United States' intention to reach net zero by 2050 are extremely important in terms of reducing the emissions of greenhouse gases because these countries adopt two-thirds of the global economy and represent 63% of global greenhouse gas emissions. By 2030, according to IRENA, the average LCOE (levellised cost of electricity) of both onshore and offshore wind electricity generation costs is projected to decrease by 25% for onshore and 55% for offshore, from 2018 levels. Annual wind installations must increase significantly to reach net zero by 2050 [6, 8, 18].

3. TURKIYE WIND ENERGY POTENTIAL AND ITS USAGE STATUS

There are different power estimations about the installed power of wind energy in Türkiye depending on the annual average wind speeds at which power plants can be established. According to the Wind Energy Potential Atlas (REPA) [16] prepared by the Ministry of Energy and Natural Resources in 2007 (Figure 10), the wind power plant that can be installed has been determined for the total country, regions and provinces based on the lands which is suitable for the establishment of a wind power plant having an altitude of 50 m with wind speed of 7,0 m/s and above as shown in Table 5 and Table 6. According to these tables, the total capacity for terrestrial areas is 37.836 MW, and for offshore 10.013 MW, a total of 47.894 MW, which is approximately 48.000 MW. Among the regions, the Ege region has the highest value with 18.975 MW, followed by the Marmara region with 12.704 MW, which shows that there are too many wind farms established in these two regions today.

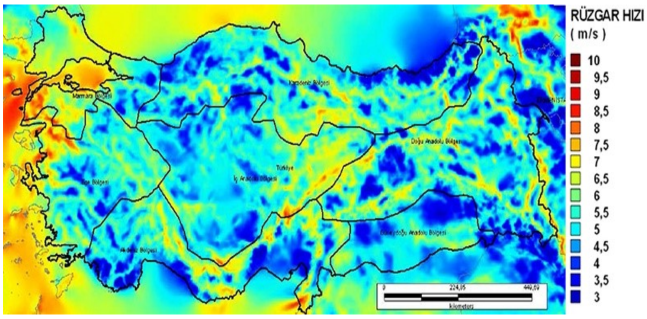


Figure 10. Türkiye 70 m wind speed potential map [16]

Table 5. Türkiye's wind farm capacities for 50 m altitude and annual average wind speed of 7,0 m/s and above [16]

Wind Classification	Annual Power Density, W/m^2	Annual Average Wind Speed, m/s	Total Capacity, MW
4	400-500	7,0-7,5	29.259,36
5	500-600	7,5-8,0	12.994,32
6	600-800	8,0-9,0	5.399,92
7	>800	>9,0	195,84
	Total Capacity		**47.894,44** terrestrial: 37.836 above the sea: 10.013

Table 6. Wind plant capacities of regions of Türkiye for 50 m altitude and annual average wind speed of 7,0 m/s and above (MW) [16]

Region Name	**Capacity (MW)**
Marmara	12.704,0
Aegean	14.975,0
Mediterranean	5.335,0
Central Anatolia	914,0
Black Sea	2.472,0
Eastern Anatolia	986,0
Southeast Anatolia	0,0
Total Capacity	37.386,0

Although Türkiye's wind potential values are determined as 48.000 MW based on the lands having 50 m altitude and wind speed of 7,0 m/s and above according to 2007 REPA values, when 100 m altitude and wind speed of 6.0 m/s and above areas are taken into consideration, the potential is 83.000 MW appears to be out. In recent city-based potential studies for larger powerful and more efficient wind turbines, it has been determined as 115.129 MW. Considering the rate at which this potential is used, the usage rate is 8,3% for the installed power of 9.559 MW according to the values of 2020, and 9,2% for the 10.585 MW power in October 2021. In Table 7, the 5 cities with the highest potential and current capacity utilizations status are given. The city with the highest potential is Balıkesir with 13.827 MW, followed by Çanakkale with 13.013 MW [1].

Table 7. 5 cities with the highest wind potential in Türkiye [1]

City	**Theoretical Potential MW**	**In operation MW**	**Under constr. MW**	**Licenced MW**	**Sub-license MW**	**Total Process MW**	**Process/ Theory Ratio**
Balıkesir	13.827	1.294	87	0	0	1.382	10 %
Çanakkale	13.013	797	256	40	162	1.255	10 %
İzmir	11.854	1.680	55	0	23	1.758	15 %
Manisa	5.302	691	11	10	0	712	13 %
Samsun	5.222	48	4	9	0	61	1 %

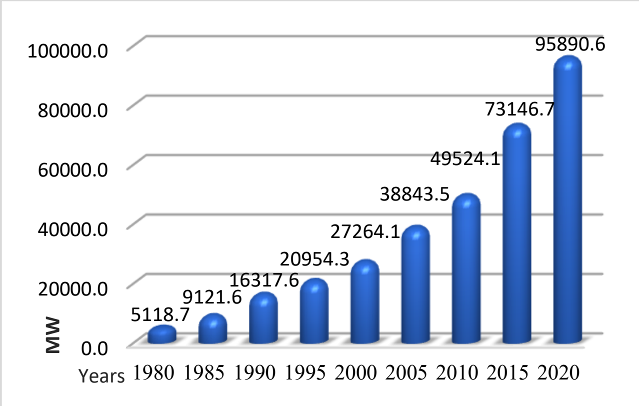


Figure 11. Development of Türkiye's installed power over the years [15, 20]

The development of Türkiye's electrical installed power between 1980 and 2020 is given in Figure 11, and the power which was 5.118 MW in 1980 increased approximately 19 times and reached 95.890 MW after 40 years. The first wind farm in Türkiye was established in 1998 in Izmir. Türkiye's total electricity installed power values and electricity generation from renewable sources in 2020 are given in Table 8 and Table 9 respectively. As of the end of 2020, Türkiye has 8.832 MW of installed wind power and the ratio of this power to the installed electricity is 9,21% and the energy produced with this power is 24.828 GWh and its ratio to the total energy generation is approximately 8,1%. Some of the 269 power plants that have been commissioned have not yet reached the installed capacity of the license and their construction is still on going. With the full capacity commissioning of these power plants, an additional wind turbine with a capacity of 2.091 MW will be commissioned and the installed power will reach 11.650 MW. In addition, the license capacity of 60 power plants, of which no units have been commissioned yet but progress has been made in their installation, is 165 MW. When all of these projects are completed, the installed capacity of Türkiye's wind power plants will be 11.814 MW.

Table 8. Türkiye's total electricity installed power values in 2020 [20]

Source	Installed Power MW	Ratio, %
Import Coal	8.841,9	9,22
Bituminous Coal	782,5	0,82
Lignite	9.988,7	10,42
Liquid Fuels	189,4	0,20
Multi-Fuel	4.889,1	5,10
Waste Heat	397,5	0,41
Natural Gas	21.599,4	22,53
Renewable Waste + Waste	1.105,3	1,15
Wind	8.832,4	9,21
Solar	6.667,4	6,95
Water Dam	22.925,0	23,91
Water Stream, Lake	8.058,9	8,40
Geothermal	1.613,2	1,68
TOTAL	**95.890,6**	**100**

The share of wind energy in Türkiye's total electricity installed power is approximately 10 percent, while wind energy has 19,3 percent of the total renewable energy installed power. When the wind energy capacity is evaluated on a city basis, İzmir has the highest capacity with approximately 1700 MW, followed by Balıkesir with 1300 MW, Çanakkale with approximately 850 MW, Manisa with 750 MW and Istanbul with 420 MW. In addition, Hatay and Kırklareli with 415 MW each, Aydın with 400 MW, Afyonkarahisar with 325 MW, Kayseri with 272 MW (10th), Osmaniye with 260 and Konya with 242 MW (12th), Muğla with 220 MW, Bursa with 205 MW and Mersin with approximately 200 MW installed wind capacity are among the 15 cities with the highest energy installed capacity.

The installed power shares of different energy sources in Türkiye's electricity production for the years 2010 and 2020 are given in Figure 12. These energy sources used for electricity production are hydroelectric, natural gas, coal (stone, lignite and imported coal), wind, diesel, liquid fuels, geothermal, biogas and solar the values given in Table 8 is direct agreement with the 2020 values of Figure 11.

Table 9. Türkiye’s electricity generation from renewable sources in 2020 [19, 20]

Source	Generation (GWh)	Contribution (%)
Renewable Waste+Waste	4.459,9	3,47
Wind	24.828,2	19,34
Solar	10.950,2	8,53
Water Dam	57.463,9	44,77
Water Stream, Lake	20.630,4	16,07
Geothermal	10.027,7	7,81
Total Renewable	**128.360,4**	**100,00**
Renewable (%)	**----**	**41,85**
Türkiye Total	**306.703,1**	**100**

Furthermore, the fact that the energy produced by hydroelectric power plants is higher than all the energy obtained by the use of fossil fuels in 2020 values, the rate of renewable energy approaching 50% and the rate of electricity generation from wind energy approaching 50% in renewable generation shows the importance of renewable and wind energy (Table 9).

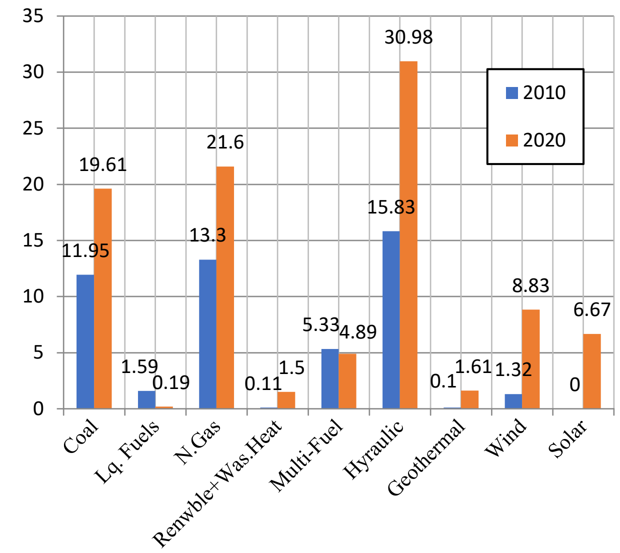


Figure 12. Electricity installed capacity of Türkiye from primary energy sources in 2010 and 2020 (GW) [19, 20]

When the graph of Türkiye's wind total installed power and annual installation amounts between (1998-2020) given in Figure 13 is examined, it is seen that the total installed power has increased approximately 7 times in the last 10 years and 2 times in the last 5 years. Annual installation amounts have grown rapidly after 2012, after reaching the highest value in 2016, it decreased slightly until 2018 and started to increase rapidly as of 2019.

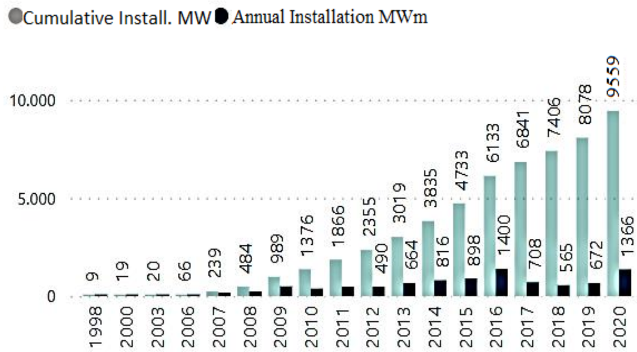


Figure 13. Development of Türkiye’s wind power over the years [15, 20]

In recent years, there have been significant developments in the technologies of wind turbines used to convert wind energy into electrical energy. The first of these developments has been achieved by the widespread use of turbines without gearbox and their getting cheaper. The second important development has been achieved by increasing the rotor diameter that is directly related with the turbine power, by constructing turbines with a power over 10 MW. While the first of these developments will increase the efficiency of the turbines, the second will increase the tower height of the turbine and increase the power coefficient value with higher wind speeds.

4. CONCLUSION

In this study, first the energy needs and current energy usage status of the world and Türkiye were determined, then the wind energy potentials were obtained and by using these wind energy potentials, it has been investigated to what extent the energy demands can be met. In addition, measures and scenarios were analysed to reduce the use of fossil-based energy, which has harmful effects on the ecological balance of the world and living health, and to increase the use of wind and other renewable energy resources to replace them. The results obtained in the study are summarized below.

Currently, 80% of the world's general energy needs and 85% of Türkiye's total energy needs are met from fossil fuels. The share of wind energy in Türkiye's total electricity installed power is approximately 10 per cent, while wind energy has 19,3 per cent of the total renewable energy installed power.

The world technical wind potential is 53000 TWh/year, with the highest being North America with 14000 TWh/year from the continents and Oceania having the lowest 3000 TWh/year. In addition, North America, Eastern Europe, Russia and Africa have 66% of the world's wind energy potential. The total installed wind power of the world in 2021 is 837 GW, and the installed power of Türkiye is 10,7 GW. With this power value, Türkiye ranks 12th among world countries and 7th among European countries. As of the end of 2021, Türkiye has 10.681 MW of wind power, which is 9,2% of the country's electrical installed power, and the energy generation by using this power is 24.828 GWh and its ratio to the total energy production is approximately 10%.

While the world's total electrical energy production in 2020 is 26.823,2 TWh/year, Türkiye's production is 305,5 TWh/year. Energy production from renewable sources was 3.147,0 TWh/year for the world and 49,8 TWh/year for Türkiye.

According to IRENA's Energy Transition Scenario, additional annual wind capacity increase would need to reach approximately 180 GW to have a chance to meet the Paris targets. According to the IEA's Net Zero scenario by 2050, it needs to increase to 160 GW by 2025 and to 280 GW by 2030. This value is 3 times the increase in 2020.

Developments seen in turbine technologies and the production of domestic turbines in Türkiye will increase the use of wind energy, making wind electricity cheaper both in the world and in Türkiye. As a result, it is estimated that it will be possible to reach the target of 100% zero emissions and 100% renewable energy production by using all renewable energy sources, even if not only with wind energy, until 2053 or 2070.

Author contributions: Concept – Faruk Köse, Süleyman Köse; Data Collection & Processing - F.K., S. K.; Literature Search - F.K., S. K.,; Writing - F.K., S. K.
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. C. Senel, E. Koc, 2015. "Wind Energy in the World and Türkiye Condition-General Evaluation", Journal of Engineers and Machinery, volume 56, issue 663, p. 46-56.

[2] F. Kose, Renewable Energy Resources (and Systems), S.Unv. Faculty of Engineering and Architecture Lecture Note. Publication No: 51, Konya, 2002.

[3] Anonymous-1, "Statistical Review of World Energy 2021" 70th edition, 2021, (bp.com), BP p.l.c.1St James's Square London SW1Y 4PD/UK.

[4] bp Statistical Review of World Energy 2007, 56th edition. Url: https://refman.energytransitionmodel.com/publications/344, Date of access: 12.01.2022.

[5] F., Kose, M.N., Kaya, (2013). "Analysis on meeting the electric energy demand of an active plant with a wind-hydro hybrid power station in Konya Türkiye Konya water treatment plant." Renewable Energy, 55, 196-201.

[6] IRENA, "Renewable Energy Statistics 2021", The International Renewable Energy Agency, Abu Dhabi.

[7] J. Lee, F. Zhao, "GWEC- Global Wind Report 2021, Global Wind Energy Council, Rue Belliard 51-53, 1000 Brussels, Belgium.

[8] EWEA, "Wind energy scenarios for 2030", A report by the European Wind Energy Association, 2015. Url: ewea.org/, Date of access: 10.12.2021.

[9] F. Kose, M.H. Aksoy, M. Ozgoren (2014).,"An Assessment of Wind Energy Potential to Meet Electricity Demand and Economic Feasibility in Konya, Türkiye", International Journal of Green Energy, 11(6), 559-576.,

[10] N. Karabağ, C. B. Çobanoğlu Kayıkcı, A. Ongen, (2021),"%100 The World and Türkiye on the Road to Renewable Energy Transition,", European Journal of Science and Technology No.21, pp.230-240, 2021.

[11] "The Law on the Use of Renewable Energy Resources for the Purpose of Electricity Generation", Law Number: 5346, Date of Enactment: 10/5/2005 Release date in the Government gazette: Date: 18/5/2005 Number: 25819 (Last Revision Date: 11.01.2022).

[12] "Regulation on the encouragement of domestic production of the components used in the facilities producing electrical energy from renewable energy sources", Release Date in Official Government Gazette: 24.06.2016 Official Government Gazette Number: 29752.

[13] Joyce Lee, Feng Zhao, "GWEC-Global Wind Report 2022", Global Wind Energy Council, Rue de Commerce 31 1000 Brussels, Belgium, www.gwec.net, Date of access: 04.04.2022.

[14] Global wind map, https://www.vaisala.com/en/lp/free-wind-and-solar-resource-maps#. Date of access 13.06.2022.

[15] IEA, "World Energy Outlook 2021", Url: https://www.iea.org/reports/world-energy-outlook-2021, Date of access: 27.10.2021.

[16] REPA, "Turkish Wind Energy Potential Atlas", Ministry of Energy and Natural Resources, 2007, Ankara.

[17] GWEA, "Global Wınd Energy Manıfesto For Cop26", Url: www.windareyouin.com, Date of access: 18.10.2021.

[18] S. Dale, "bp Energy Outlook: 2022 edition", web: https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/energy-outlook/bp-energy-outlook-2022.pdf, Date of access: 14.06.2022.

[19] EİGM, Url: https://enerji.gov.tr/eigm-yenilenebilir-enerji-kaynaklar-ruzgar, Date of access: 25.11.2021.

[20] TEİAŞ, (Türkiye Electricity Transmission Co.), https://www.teias.gov.tr/, Date of access: 27.12.2021.

Research Paper
Academic Platform Journal of Engineering and Smart Systems (APJESS) 10(3), 149-156, 2022
ISSN: 2822-2385
Received: 07-Apr-2022 Accepted: 02-Aug-2022
https://doi.org/10.21541/apjess.1100238

Unsupervised Learning Approach for Detection and Localization of Structural Damage using Output-only Measurements

*[1]Burcu GÜNEŞ

[1] Department of Civil Engineering, Istanbul Technical University, Istanbul, Turkey, bgunes@itu.edu.tr

Abstract

Interrogation of the vibration data collected from the sensors embedded throughout the structure without relying on a finite element model of the system for monitoring the health of structural systems has received significant attention in the recent years especially with the current advancements in sensor technology. The data-driven methods explored within this context falls into the realm of statistical pattern recognition field requiring extraction of damage detection features and a statistical decision-making process for identification of damage. Machine learning algorithms provide statistical means for making such decisions. In this study, an unsupervised machine learning approach, one-class support vector machine (OC-SVM), requiring training data only from the undamaged state of the structure is explored for damage detection purposes. The coefficients of the autoregressive (AR) model are extracted as damage sensitive features and used as the required training data. The trained classifier is then used with the data obtained from the same structure at different damage states for classification. Damage detection in the form of recognizing outliers or anomalies not belonging to the target class, is followed by damage localization within the given sensor resolution using statistical means. To this end, Itakura distance measuring the distance between two sets of linear predictor coefficients of the AR processes, is utilized as damage location indicator. Numerical simulations are performed on a truss and a beam structure with several damage scenarios including realistic levels of measurement noise and modeling error. Results show that the proposed approach can successfully detect existence of damage and the statistical measure shows promising performance for further localization of the damaged region.

Keywords: Structural Health Monitoring, Unsupervised Learning, Support Vector Machines, Time Series Modelling, Statistical Pattern Recognition

1. INTRODUCTION

Monitoring the health of civil engineering systems due to deterioration under normal operating conditions or after an extreme event is vital to take the necessary preventive measures to protect these systems against collapse, reducing maintenance cost and prolonging their service lives. Structural Health Monitoring (SHM) aims to assure the structural safety of civil infrastructure by evaluating the integrity of these systems and providing warning signs as soon as the condition of the structure deteriorates. Several approaches that estimate the state of the structure with different levels of refinement have been proposed in the literature. According to Rytter [1], damage characterization includes four stages: (1) damage detection, (2) damage localization, (3) damage quantification, and (4) prediction of the remaining service life.

Vibration-based damage identification is a subdiscipline of the SHM field that accomplishes these tasks through measured vibration responses of the structure. The vibration-based damage detection can be classified in two general classes of model-based and data-driven approaches [2]. Data-driven approaches utilizing solely the recorded vibration data stand out as the more convenient and accessible alternative in regards to their ease of implementation over the model-based approaches requiring refined finite element model of the structure. Especially for complex structural systems with a large number of degrees of freedom, obtaining a refined finite element model of the structure is a difficult, computationally expensive task and from the material heterogeneity and mechanical behavior aspects, may not even be feasible or practical. Data-driven approaches also establish a model, but this is usually a statistical representation of the system. Released from the physical model requirement, data-driven damage detection depend on statistical analysis to determine the significance

* Corresponding Author

of the changes in the data and have the potential to be constructed in a fully automated manner with minimum user interaction.

Time series modeling and outlier analysis can be listed among the statistical approaches available for structural diagnosis based on collected vibration data. With these approaches, diagnosis stage usually starts by a data processing step where data collected from a sensor network deployed throughout the structure are transformed into features that are sensitive to damage. This transformation retains only the information necessary for diagnosis while discarding any other information. To this end, Worden et al. [6] performed an outlier analysis on a spring–mass–dashpot system for damage detection. Sohn et al. [3] and Sohn and Farrar [4] applied time-series modeling and utilized the residual error of the prediction model as a damage sensitive feature. Wandji [5] exploited autoregressive (AR) modeling to propose a goodness-of-fit test to distinguish damaged data. Mattson and Pandit[6] used the standard deviation of the residual of the AR model of vibration data as damage detecting feature. In another study by Nair et al.[7], the coefficients of the AR model of the ambient vibration response data was explored as damage sensitive features and damage states were identified through hypothesis testing. Kar and Mohanty [8] utilized time domain signals and diagnosed ball bearing faults by employing Kolmogorov–Smirnov (K-S) test. Distance between autoregressive moving average (ARMA) models were investigated by Zheng and Mita [9] for damage detection. Nair and Kiremidjian [10] detected damage by investigating the migration of extracted AR coefficients in which the feature vector is modeled by a Gaussian mixture model. In a later study, Gul and Catbas [11] predicted the data of the damaged structure, utilizing the autoregressive models with exogenous input (ARX) developed for the healthy structure, and proposed the difference between the predictions and the measured values as the damage feature. Roy et al. [12] explored combination of ARX model coefficients, K-S test statistical distance and the model residual error as damage sensitive features for the potential of detection and localizing damage. Entezami et al. [13] developed damage indices for damage localization that are based on damage sensitive features extracted from time series modeling of the measured vibration data.

Following the feature extraction from each data set, comparison of these features with the baseline values facilitates to arrive at a damage decision. In this context, the process can essentially be viewed as a statistical pattern recognition problem that is part of machine learning concerned with classification [14]. More specifically, this means, once these features are extracted at the initial nominal state of the structure, a machine learning approach can be employed to train a relationship between these features and the baseline state.

Despite many implementations of the time series models and machine learning techniques in the context of SHM majority of the algorithms available in the literature are custom-made for certain data sets. The lack of unsupervised techniques capable of detecting structural damage without relying on any prior knowledge about the structure's condition; the lack of strategies capable of detecting structural damage automatically in a robust manner and the lack of algorithms suitable for performing real-time detection can be listed among the obstacles several issues remain as aspects hindering the practical application of these damage detection algorithms to civil engineering structures. This study aims to address these issues and provide a technique that can be implemented in an unsupervised, automated and decentralized manner that will be suitable for real-time SHM.

For cases where training data are available only from a single condition, the baseline or the undamaged state, the unsupervised machine learning approaches can be adopted to train the decision-making algorithm. This fits very well for civil engineering applications since each structural system is unique with its own structural and dynamic properties requiring its own individual training data. It is not practical or even possible to expect that labeled training data are acquired and made available for various damage scenarios . One-class support vector machine (OC-SVM) classifier, also known as novelty detection approach, is an unsupervised learning algorithm classifying just one-class objects. It has been used successfully in many applications such as image retrieval, audio surveillance, biometric traits and more recently in SHM applications. The recent applications of machine learning methods utilized for vibration-based damage detection in civil structures are reviewed in [15, 16].

In the proposed approach, the extracted features from the AR model are utilized to train the OC-SVM. With the learned relationship, the trained classifier exploits the data obtained from the same structure at different states and arrives at a damage decision, by classifying the data as either belonging to the learned state (undamaged) or not (damaged). This completes the first stage of the damage characterization problem. For the next stage of damage localization, in the extracted feature space of AR coefficients, a statistical distance metric is employed. This metric, defined as Itakura distance, initially developed to measure the similarities between AR models of the voice segments [17]. This metric is calculated for each sensor location and the larger value is interpreted as an indication to the proximity of damage.

The proposed methodology is investigated numerically on two different structural systems: a truss and a beam type structure. Several damage scenarios are simulated including reduction in member rigidities and fracture of the members for the truss system and loss of connection rigidities for the beam model. The acceleration responses under ambient excitation data are simulated at the designated sensor locations. Modeling error and measurement noise are incorporated into the numerical work for simulating field data.

2. MATERIALS AND METHODS

2.1. Time Series Modeling for Feature Extraction

Among the time series models available for fitting a mathematical model to time series data, AR model stands out as the most widely used time series model for extracting damage sensitive system. The AR models assume the measurements have noise, and modeling error is sampled from Gaussian distribution.

For a linear, stationary, and univariate time series, the AR model is formulated as follows:

$$y(t)=\sum_{i=1}^{p} a_i\, y(t-i)+e(t) \tag{1}$$

where $y(t)$ is the measured vibration response at a single sensor at time t, $a_i=[a_1, a_2, \ldots, a_p]$ represents the AR model coefficients, p is the order of the model, and $e(t)$ is the residual error at time that corresponds to the difference between the measured and the predicted time series data obtained by the AR model [18, 19].

By fitting an AR model to time-domain response data acquired from each sensor at the nominal state of the structure, one can extract the AR parameters for the corresponding sensor location representing the healthy state. Measured data from the same set of sensors at an unknown state of the structure are also processed to extract the AR parameters fitting to that undetermined state. These parameters are compared as damage sensitive features to be compared for classification using a support machine algorithm described in the following section.

2.2. One-Class Support Vector Machine

The OC-SVM is a specialized version of the standard support vector machines with the objective of finding an optimal hyperplane in which most of the training samples are included in a minimum volume and separating and distinguishing one-class objects from all others. Objects falling outside the constructed hyperplane are considered as outliers.

Scholkopf et al. [20, 21] proposed the OC-SVM algorithm to train the classifier using only the features belonging to the target class. The decision function for the hypersphere that encloses the maximum number of training samples takes the following form

$$f(x) = sgn\left\{\sum_{i=1}^{m} \alpha_i K(x, x_i) - \rho_i\right\} \tag{2}$$

where m is the number of training samples and α_i are the Lagrange multipliers obtained from optimization of the following set of equations:

$$min_{\alpha} \left\{\frac{1}{2} \alpha_i \alpha_j K(x_i, x_j)\right\} \tag{3}$$

subject to

$$0 \le \alpha_i \le \frac{1}{v_m} \tag{4}$$

$$\sum_{i}^{m} \alpha_i = 1 \tag{5}$$

ρ in equation (2) is the distance of the hypersphere from the origin. ν is the percentage of the samples considered as outliers and $K(\ )$ defines the OC-SVM kernel that allows projection of samples from the original to the feature space. Among the various kernels, radial basis function is the most used kernel allowing to determine the radius of the hypersphere through the kernel parameter γ as follows:

$$K(x, x_i) = \exp\left(-\gamma d(x, x_i)\right) \tag{6}$$

such that

$$d(x, x_i) = \|x - x_i\|^2 \tag{7}$$

2.3. Statistical Control Metrics: Itakura Distance Measure for Damage Localization

Following the decision on existence of structural damage, a metric implemented for damage localization is needed for completion of the damage identification process. The available literature proposes parametric-based damage indices that uses the parameters of the prediction models for identifying damage location and [13, 22, 23] or some other measures correlating locations of a sensor network with the damage features [9, 24-26]. In this study, a statistical measure defined as Itakura distance [17] is explored as a possible means for fusing the information from all sensors deployed throughout the structure and localizing damage. The theoretical basis for this statistical measure are summarized below.

Suppose that $x(t)$ and $y(t)$ are two time series corresponding to the baseline state (target class) and damaged state (outlier), respectively. Assume that the coefficients of the corresponding two AR processes of order p are given by

$$a_x = [1 \quad a_{x1} \quad a_{x2} \quad \cdots \quad a_{xp}] \tag{5a}$$

$$a_y = [1 \quad a_{y1} \quad a_{y2} \quad \cdots \quad a_{yp}] \tag{5b}$$

The means square error (MSE) for the baseline process corresponding to $x(t)$ is

$$MSE_{xx} = a_x^T\, R_x a_x \tag{6}$$

where R_x is the autocorrelation matrix of signal $x(t)$ of size p+1.

Supposing that $x(t)$ pass through the AR model parametrized by a_y, the MSE of the signal is

$$MSE_{xy} = a_y^T\, R_x a_y \tag{7}$$

The Itakura distance indicating how far the outlier state parametrized by a_y, is from the baseline state parametrized by a_x, is defined as

$$D_{Ix,y} = log\, \frac{a_y^T R_x a_y}{a_x^T R_x a_x} \tag{8}$$

Similarly, how well the signal $y(t)$ is modeled by the AR-parameters of $x(t)$, is given by

$$D_{Iy,x} = log\, \frac{a_x^T R_y a_x}{a_y^T R_y a_y} \tag{9}$$

Combining the two distances to obtain a symmetric distance measure, one gets the Itakura distance (D_I) as

$$D_I = \frac{1}{2}\left(D_{Ix,y} + D_{Iy,x}\right) \qquad (10)$$

2.4. Methodology

The proposed damage detection and localization methodology is depicted in Figure 1.

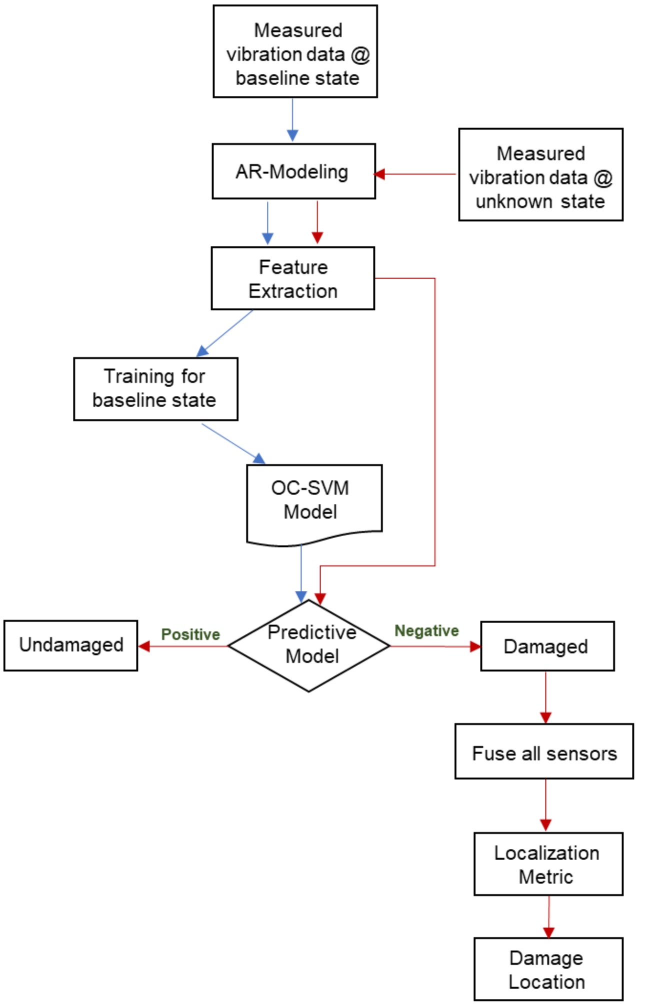


Figure 1. Damage identification methodology

The methodology starts by the training stage of the OC-SVM for the baseline (undamaged) state of the structure. The measured acceleration signals in response to ambient excitations from the structure at this state are processed including detrending and normalization procedures leading to the extraction of AR parameters. After sufficient data required for training is collected, the OC-SVM classifier is attained for each sensor location as the predictive model. The completion of the training stage is followed by the prediction stage for the data collected from the structure later at an unknown state. With the new datasets of unknown state, AR parameters are identified and these newly identified features are tested with the previously established predictive model. This comparison with the OC-SVM classifier outputs a binary decision regarding the classification of data as belonging to the baseline (undamaged) state or not. Once damage is detected damage localization stage is invoked by calculating the localization index, namely Itakura distance, for all sensor locations. The sensor with the highest value of the localization index is marked as the sensor closest to the possible damaged location.

3. NUMERICAL EVALUATION

This section presents the numerical simulations conducted on two different structures to investigate the performance of the proposed damage detection algorithm and the localization metric. The two structures, 14-degree-of-freedom (DOF) truss and 24 DOF fixed-ended beam considered in this study are depicted in Figures 1(a) and 2(a), respectively. For both systems, masses are assumed to be lumped at the degrees of freedom and damping is assigned to be proportional, 2% for all modes. At the undamaged state, the members have identical rigidities.

The locations of the acceleration sensors together with the damage scenarios simulated for these structures are also depicted in these figures. Note that for each system, input motion is taken as ambient excitations acting in the vertical direction that are assumed to be unmeasured and output measurements as the acceleration responses simulated in the same direction as the input. This leads to the deployment of 7 acceleration sensors for the truss and 8 sensors for the fixed-ended beam, for output measurements as shown in Figures 1(b) and 2(b). The simulated damage scenarios in the form of loss of member and connection rigidity are summarized in Table 1 for both structural systems. The analytically computed natural frequencies of these systems for the simulation cases are determined through eigenvalue analysis and the first four modes are listed in Table 2.

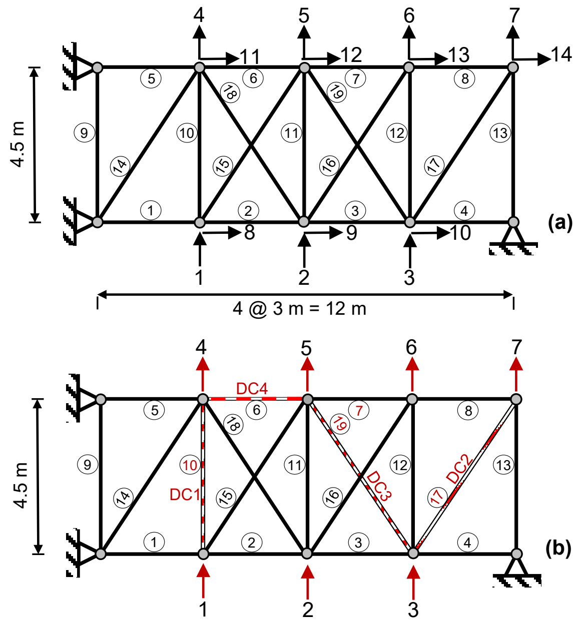


Figure 1. (a) 14 DOF truss model, (b) Damage scenarios and sensor locations for input and output

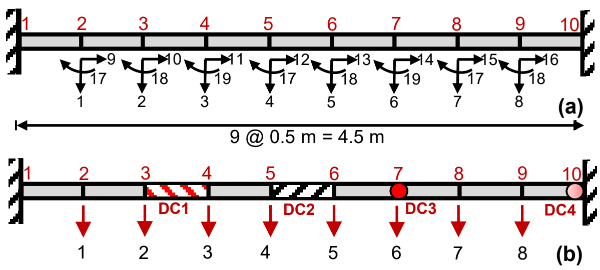


Figure 2. (a) Fixed-ended beam model, (b) Damage scenarios and sensor locations for input and output

Table 1. Summary of the damage scenarios

Configuration	*TRUSS*	*FIXED-FIXED BEAM*
DC 1	Bar 10 - 30% loss of axial rigidity	Section 34- 30% loss of flexural rigidity
DC 2	Bar 17 - 90% loss of axial rigidity	Section 34- 90% loss of flexural rigidity
DC 3	Bar 19 - 90% loss of axial rigidity	Plastic hinge at joint 7
DC 4	Bar 6 - 50% loss of axial rigidity	Plastic hinge at joint 10

Table 2. Natural frequencies of the systems considered in the study

Configuration	*TRUSS*				*FIXED-FIXED BEAM*			
	f_1 (Hz)	*f_2 (Hz)*	*f_3 (Hz)*	*f_4 (Hz)*	*f_1 (Hz)*	*f_2 (Hz)*	*f_3 (Hz)*	*f_4 (Hz)*
Healthy	4.152	11.238	18.300	24.891	2.000	5.495	10.715	17.531
DC 1	4.123	11.185	18.162	24.777	1.990	5.316	10.496	17.336
DC 2	3.242	6.771	17.029	24.878	1.519	5.153	8.676	13.868
DC 3	3.969	10.663	17.326	24.870	0.604	3.834	4.193	10.821
DC 4	4.152	11.176	17.970	24.398	1.378	4.455	9.257	15.706

In both cases, the input ambient excitations are modeled as Gaussian white noise processes and the output acceleration measurements are simulated with a sampling time of 0.04 sec. for a duration 300 sec. which is deemed sufficient to capture the frequency range of interest (2-25 Hz). A total of 100 simulations are performed at the nominal (healthy) state of the structures, 50% of which are used as the baseline training data for the one-class classifier. To gain a statistical sense of performance, 100 simulations for each damage scenario are examined. In the simulated acceleration measurements, sensor noise is contemplated using a random number generator with a level ranging between 2-10 % of the RMS of the response measured on DOF 1 such that

$$SNR = \sigma_{noise}^2 / \sigma_{signal}^2 \ldots\ldots \quad (13)$$

where σ_{signal}^2, σ_{noise}^2 are the variance of the simulated acceleration and the noise signals and *SNR* is the desired signal to noise ratio.

In order to mimic the variability in the recorded measurements due to operating conditions of existing structural systems, modeling error is also introduced to the simulations by assigning the modulus of elasticity of the members as the true values multiplied by a random scalar that is uniformly distributed between [0.9 and 1.1].

4. RESULTS AND DISCUSSION

Damage detection performance of the trained OC-SVM for the truss system and the beam in the 2-dimensional, and the 5-dimensional feature space are summarized in Tables 3 and 4, respectively. In the results presented healthy configuration is the baseline state for which training data was available and DC stands for the '*damage case'*. Examination of the data shows that, even with only two features, the damage detection performance of the OC-SVM is excellent for the truss system with the exception of a single sensor location for a specific damage scenario, namely sensor 7 for DC 2. Increasing the feature space slightly, from 2 to 5, for this specific case results in significant improvement. For the remainder of the data, the performance of the 2-dimensional versus 5-dimensional feature space data are comparable. For the fixed-ended beam simulations, although sensors close to the damage locations detect the existence of damage, the ones further away failed to label the damage cases as '*not belonging*' to the target state.

Table 3. Percent accuracy in damage detection of the truss system using OC-SVM

2 Features

Config.	*S 1*	*S 2*	*S 3*	*S 4*	*S 5*	*S 6*	*S 7*
Healthy	96	98	94	94	98	94	94
DC 1	97	100	100	100	100	88	100
DC 2	100	100	100	92	100	100	30
DC 3	100	100	100	100	100	99	97
DC 4	97	100	100	100	100	96	100

5 Features

Config.	*S 1*	*S 2*	*S 3*	*S 4*	*S 5*	*S 6*	*S 7*
Healthy	98	98	96	92	94	96	94
DC 1	100	100	100	100	100	100	95
DC 2	100	100	100	96	100	100	87
DC 3	100	100	100	100	100	100	92
DC 4	100	100	100	100	100	100	99

Table 4. Percent accuracy in damage detection of the fixed ended beam using OC-SVM

2 Features

Config.	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**
Healthy	90	82	82	74	82	82	96	94
DC 1	94	93	96	79	48	27	90	52
DC 2	65	46	89	89	72	99	66	43
DC 3	33	83	30	41	64	94	100	66
DC 4	20	12	11	87	24	99	99	32

5 Features

Config.	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**
Healthy	94	94	88	74	80	80	94	92
DC 1	92	98	98	93	97	99	98	66
DC 2	91	100	99	99	98	99	99	86
DC 3	49	99	99	92	63	100	100	67
DC 4	43	100	91	83	90	100	96	69

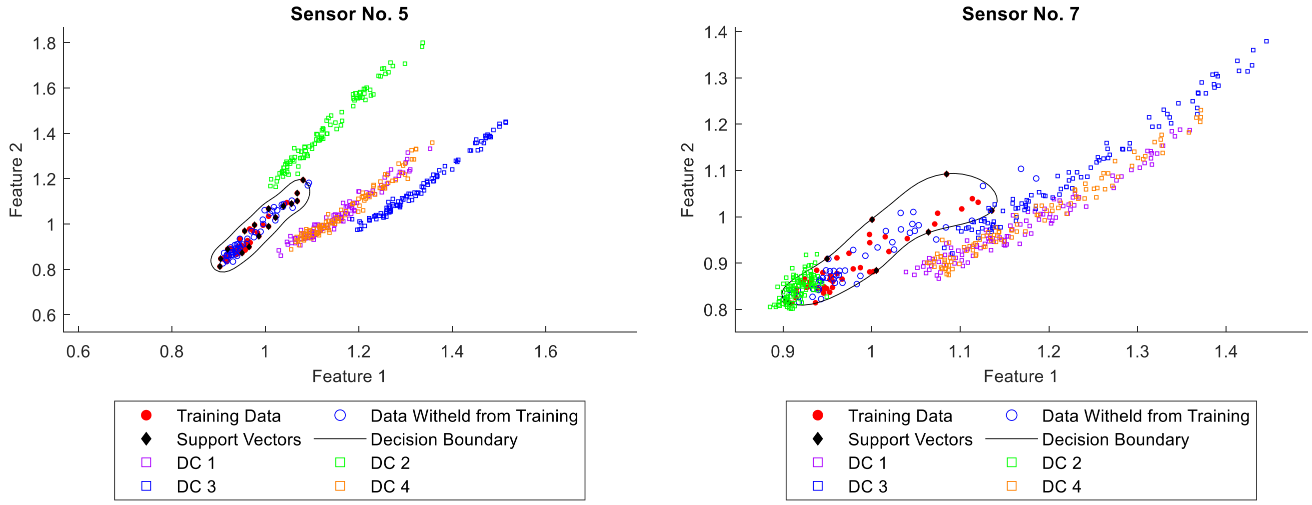


Figure 3. Training, validation, damage data and OC-SVM decision boundary for the truss system

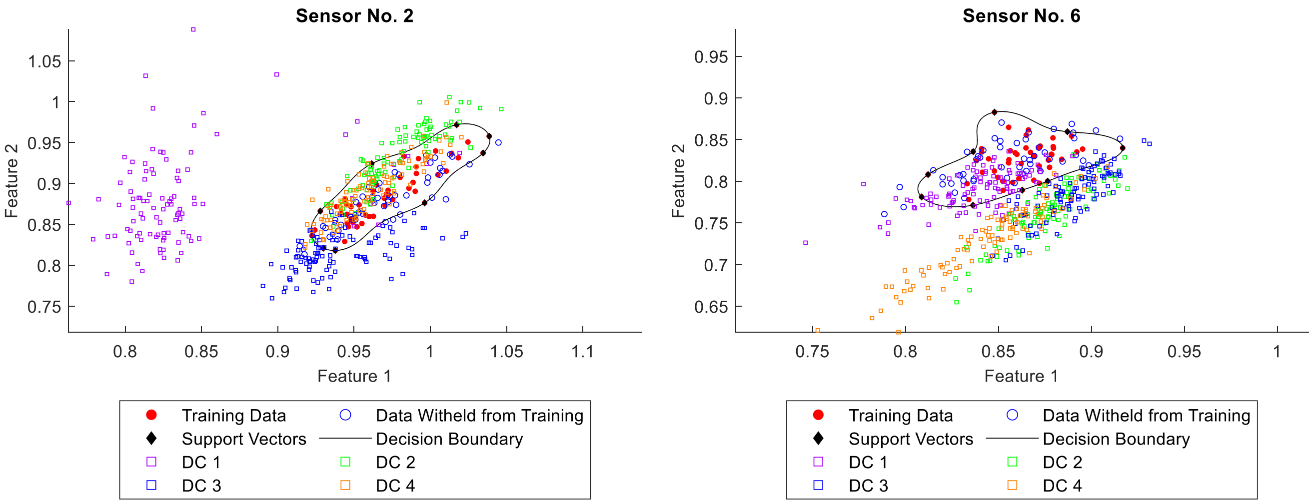


Figure 4. Training, validation, damage data and OC-SVM decision boundary for the fixed ended beam

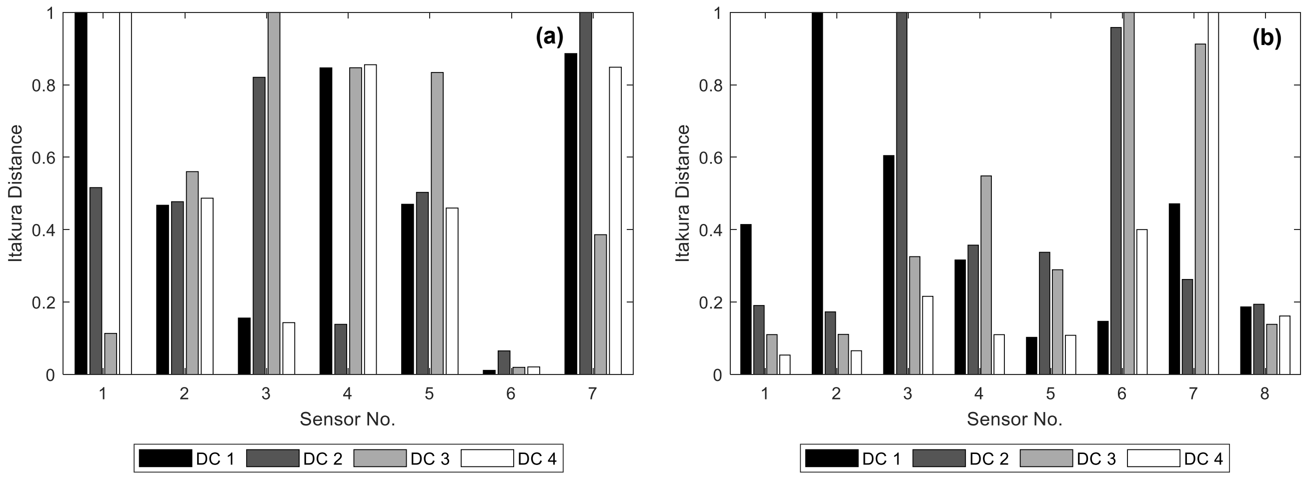


Figure 5. Normalized Itakura distance (a) truss system, (b) fixed-ended beam

Although increasing the feature space leads to some improvement for these misclassifications, it is not consistently significant enough for all the simulation cases to conclude with that generalization.

For visualizing the distribution of the damaged state data in relation to the decision boundary, the 2-dimensional feature space data including the training data, the support vectors and the decision boundary generated by the trained OC-SVM for selected sensors are provided for truss and beam systems in Figures 3 and 4, respectively. Damage localization using Itakura distance measure is carried out for both systems and the normalized values of the metric are plotted in Figure 5.

Although the plots are not sharp and distinctive enough to pinpoint the damaged region, the sensor giving the largest value of 1 is considered as the region closest to damage. The localization results based on these normalized values are tabulated in Table 5. For both systems, three out of the four damage scenarios, true damage locations are included in the identified set of probable damage locations. This is an expected accuracy for a data driven methodology utilizing an unsupervised learning approach and not exploiting a finite element model or data from different damage states. With its limited information, solely data obtained at the nominal state of the structure, perfect spatial discrimination cannot always be guaranteed. Regardless, it is evident that the approach. holds great promise for addressing the most important task of damage existence.

Table 5. Damage localization results

TRUSS

Config	Probable Location	True Location
DC 1	Sensor 1: Bar 1, **10,** 15, 2	Bar 10
DC 2	Sensor 7: Bar 8, 13, **17**	Bar 17
DC 3	Sensor 3: Bar 3, 4, 12, 17, **19**	Bar 19
DC 4	Sensor 1: Bar 1, 10, 15, 2	Bar 6

FIXED-ENDED BEAM

Config	*Probable Location*	*True Location*
DC 1	Sensor **2**	2, 3
DC 2	Sensor **3**	4, 5
DC 3	Sensor **6**	6
DC 4	Sensor **7**	8

5. CONCLUSIONS

Machine learning algorithms, with their capability of lending themselves easily for automation and implementation to different structural systems and sensor networks, have great potential for applications in the field of structural health monitoring. Considering the requirements of the ageing and degrading civil infrastructure, reliable detection of damage existence has the utmost importance within the damage identification problem. Although it is the first stage of the damage identification problem, it is the most important one since it precedes all other stages of locating and assessing the severity of damage and estimating the remaining service life. This study addresses this problem and presents a methodology for detection of structural damage using an unsupervised machine learning approach. The proposed approach exploits learning the undamaged state through a OC-SVM for deciding on the existence of damage and it couples this decision-making algorithm with a statistical distance measure for further localizing the damage. Simulations are performed on a beam and a truss system, two structure classes inherently different from each other, to test the robustness of the technique incorporate most of the complications that are encountered in actual applications, i.e., measurement noise on output-only measurements, modeling error, limited sensors and damages with different severity. The results demonstrate that the utilized machine is able to accurately detect damage by means of a low-dimensional feature vector obtained using the AR parameters of the output data. Damage localization by means of Itakura distance metric, include the true damage locations in the identified list of potential damage locations except for the two minor damage scenarios investigated in this study.

The proposed approach provides an unsupervised approach for damage detection that works solely on ambient vibration data collected under normal operation conditions. It overcomes the need to rely on a physical model and any apriori knowledge about the structure's condition. The strategy is capable of detecting structural damage automatically in a robust manner and operates with a localization index providing spatial information regarding damaged region within the provided sensor resolution.

The applicability of the method to damage scenarios including multiple components, changing boundary conditions and failure mechanisms include complications from the localization aspect of the damage identification problem that await future research.

Author Contributions: Concept – B.G.; Data Collection &/or Processing – B.G.; Literature Search – B.G.; Writing – B.G.
Conflict of Interest: No conflict of interest was declared by the author.
Financial Disclosure: The author declared that this study has received no financial support.

REFERENCES

[1] A. Rytter, Vibrational Based Inspection of Civil Engineering Structures. University of Aalborg, 1993.

[2] R. J. Barthorpe, On model-and data-based approaches to structural health monitoring (Doctoral dissertation), University of Sheffield, 2010.

[3] H. Sohn, C. R. Farrar, H. F. Hunter, K. Worden, Applying the LANL Statistical Pattern Recognition Paradigm for Structural Health Monitoring to Data from a Surface-Effect Fast Petrol Boat. Los Alamos National Laboratory Report LA-13761-MS. Los Alamos National Laboratory, Los Alamos, NM, 2001.

[4] H. Sohn and C. R. Farrar, "Damage diagnosis using time series analysis of vibration signals." Smart Materials and Structures, vol. 10, no. 3, pp. 446-451, 2001, doi: 10.1088/0964-1726/10/3/304.

[5] J. N. Wandji, "A nonparametric goodness-of-fit test for a class of parametric autoregressive models." Journal of Statistical Planning and Inference, vol. 71, no. 1, pp. 57-74, 1998, doi: 10.1016/s0378-3758(98)00010-x.

[6] S. G. Mattson and S. M. Pandit, "Statistical moments of autoregressive model residuals for damage localisation." Mechanical Systems and Signal Processing, vol. 20, no. 3, pp. 627-645, 2006, doi: 10.1016/j.ymssp.2004.08.005.

[7] K. K. Nair, A. S. Kiremidjian, and K. H. Law, "Time series-based damage detection and localization algorithm with application to the ASCE benchmark structure." Journal of Sound and Vibration, vol. 291, no. 1, pp. 349-368, 2006, doi: 10.1016/j.jsv.2005.06.016.

[8] C. Kar and A. Mohanty, "Monitoring gear vibrations through motor current signature analysis and wavelet transform." Mechanical Systems and Signal Processing, vol. 20, no. 1, pp. 158-187, 2006, doi: 10.1016/j.ymssp.2004.07.006.

[9] H. Zheng and A. Mita, "Two-stage damage diagnosis based on the distance between ARMA models and pre-whitening filters." Smart Materials and Structures, vol. 16, no. 5, pp. 1829-1836, 2007, doi: 10.1088/0964-1726/16/5/038.

[10] K. Krishnan Nair and A. S. Kiremidjian, "Time Series Based Structural Damage Detection Algorithm Using Gaussian Mixtures Modeling." Journal of Dynamic Systems, Measurement, and Control, vol. 129, no. 3, pp. 285-293, 2006, doi: 10.1115/1.2718241.

[11] M. Gul and F. N. Catbas, "Damage Assessment with Ambient Vibration Data Using a Novel Time Series Analysis Methodology." Journal of Structural Engineering, vol. 137, no. 12, pp. 1518-1526, 2011, doi: 10.1061/(asce)st.1943-541x.0000366.

[12] K. Roy, B. Bhattacharya, and S. Ray-Chaudhuri, "ARX model-based damage sensitive features for structural damage localization using output-only measurements." Journal of Sound and Vibration, vol. 349, pp. 99-122, 2015, doi: 10.1016/j.jsv.2015.03.038.

[13] A. Entezami and H. Shariatmadar, "An unsupervised learning approach by novel damage indices in structural health monitoring for damage localization and quantification." Structural Health Monitoring, vol. 17, no. 2, pp. 325-345, 2017, doi: 10.1177/1475921717693572.

[14] K. Worder, G. Manson, and N. Fieller, "Damage detection using outlier analysis." Journal of Sound and Vibration, vol. 229, no. 3, pp. 647-667, 2000, doi: 10.1006/jsvi.1999.2514.

[15] M. Flah, I. Nunez, W. Ben Chaabene, and M. L. Nehdi, "Machine Learning Algorithms in Civil Structural Health Monitoring: A Systematic Review." Archives of Computational Methods in Engineering, vol. 28, no. 4, pp. 2621-2643, 2020, doi: 10.1007/s11831-020-09471-9.

[16] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj, and D. J. Inman, "A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications." Mechanical Systems and Signal Processing, vol. 147, p. 107077, 2021, doi: 10.1016/j.ymssp.2020.107077.

[17] F. Itakura, "Minimum prediction residual principle applied to speech recognition." IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 23, no. 1, pp. 67-72, 1975, doi: 10.1109/tassp.1975.1162641.

[18] G. E. Box, G. M. Jenkins, G. C. Reinsel, Time series analysis: Forecasting and control, Prentice-Hall, Upper Saddle River, NJ, 1994.

[19] L. Ljung, L. System identification: Theory for the user, Prentice-Hall, Upper Saddle River, NJ, 1999.

[20] B. Schölkopf, R. C. Williamson, A. Smola, A., J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection." Advances in neural information processing systems, vol. 12, pp. 582–588, 1999.

[21] B. Schölkopf, A. J. Smola, and F. Bach, Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press, 2002.

[22] Q. Zhang, "Statistical damage identification for bridges using ambient vibration data." Computers & Structures, vol. 85, no. 7, pp. 476-485, 2007, doi: 10.1016/j.compstruc.2006.08.071.

[23] M. Gul and F. N. Catbas, "Structural health monitoring and damage assessment using a novel time series analysis methodology with sensor clustering." Journal of Sound and Vibration, vol. 330, no. 6, pp. 1196-1210, 2011, doi: 10.1016/j.jsv.2010.09.024..

[24] F. N. Catbas, H. B. Gokce, and M. Gul, "Nonparametric analysis of structural health monitoring data for identification and localization of changes: Concept, lab, and real-life studies." Structural Health Monitoring, vol. 11, no. 5, pp. 613-626, 2012, doi: 10.1177/1475921712451955.

[25] A. Mosavi, D. Dickey, R. Seracino, and S. Rizkalla, "Identifying damage locations under ambient vibrations utilizing vector autoregressive models and Mahalanobis distances." Mechanical Systems and Signal Processing, vol. 26, pp. 254-267, 2012, doi: 10.1016/j.ymssp.2011.06.009.

[26] H. Zheng and A. Mita, "Localized Damage Detection of Structures Subject to Multiple Ambient Excitations Using Two Distance Measures for Autoregressive Models." Structural Health Monitoring, vol. 8, no. 3, pp. 207-222, 2009, doi: 10.1177/1475921708102145.

Research Paper
Academic Platform Journal of Engineering and Smart Systems (APJESS) 10(3), 157-167, 2022
ISSN: 2822-2385
Received: 18-Feb-2022 Accepted: 02-Sep-2022
https://doi.org/10.21541/apjess.1075099

Modelling of Factors Influencing the Citation Counts in Statistics

*[1]Olcay ALPAY, [2]Nazan DANACIOĞLU, [3]Emel ÇANKAYA

[1] Department of Statistics, Sinop University, Sinop, Turkey, olcayb@sinop.edu.tr
[2] Department of Statistics, Sinop University, Sinop, Turkey, nazand@sinop.edu.tr
[3] Department of Statistics, Sinop University, Sinop, Turkey, ecankaya@sinop.edu.tr

Abstract

Citation is considered as the most popular quality assessment metric for scientific papers, and it is thus important to determine what factors promote the citation count of a paper in comparison to the others in the same field. The main aim of this study is to model the citation counts of the research published in SCI or SCI-Expanded journals of Statistics field which has a growing number of scientific works in Turkey. Modeling aspect is here highlighted to represent the right-skewed nature of the citations. Due to the fact that distribution of citation counts involves a great number of zeros, this study serves for an additional aim that is to model the counts with advanced discrete regression models for a more precise prediction. Data collected for this study consist of the citation counts of the papers produced by 261 Statisticians in between 2005-2017. Discrete models varying from Poisson to Zero-Inflated or Hurdle were constructed by possible influential factors, such as the publication age, the number of references, the journal category etc. Predictive performances of alternative discrete models were compared via AIC and Vuong test. Results suggested that Zero Inflated Negative Binomial and Hurdle Negative Binomial mixture models are the best forms to predict the zero inflation of citation counts. In addition, the influential factors of the final model were interpreted to make some suggestions to Statisticians to increase the citation counts of their work.

Keywords: Citation Count; Count Data Regression; Negative Binomial Regression; Zero-Inflated Regression

1. INTRODUCTION

Citation counts and related indicators are known to be used as a vital performance criterion in the academic evaluation of scientific articles, journals, researchers, and universities. Scientific journals are usually classified according to the journal impact factor, which is a scale depending on the citation counts of the articles they publish. Therefore, it has recently been questioned why some articles are more cited than the others and which factors affect the citation counts. There are many studies in the literature about citation, especially about estimating citation counts. An explanatory analysis of citation can inform us both about how conductive to citation success the personal characteristics of the authors are (such as their research experience, academic title, gender, etc.) and about the importance of the role of bibliometric features in raising a study's citation rate (such as the length of an article and its number of co-authors) [1]. Putting aside the basic descriptive analyses, the modeling citation count for predictive purposes attracts great attention. Attempts to achieve this aim are based on two separate approaches. The first considers regression modelling that can evaluate the skewness of citation counts with zero inflation. In this respect, generalized linear models (GLMs) have been found particularly useful to model such properties (e.g., Onodera and Yoshikane [2]). The second approach concentrates on classification of publications based on the magnitude of their citation counts. Defined also as machine learning methods, decision trees [3], support vector machines [4], and neural network [5] have been the most frequently used methods for this purpose. However, such methods serving for classification have the drawback of using vague information to define classification boundaries. Besides, classification approach is a simplified form of citation analysis without concentrating on citation patterns or features. To the best of our knowledge, there is only one recent study trying to ease such shortcomings [5].

In this study, we prefer the approach of regression modeling. In this respect, a series of linear modeling were applied by Vaio et al. [1] where the dependent variable is the number of times an author was cited, and the independent variables were the bibliometric variables collected for the basic sample. The question of to what extent the future number of citations that a paper will receive was addressed by Mingers et al. [6] if it is known how many citations it has received so

* Corresponding Author

far. Based on retrospective cohort study, Lokker et al. [7] compared 20 articles and journals in terms of citation counts determined by McMaster online ranking system for two years. The potential of a publication to create a scientific change was studied by Chen [8] and they proposed the structural variation model for estimating citation counts. An empirical pilot analysis to the time-dependent distribution of the percentages of never-cited papers was performed by Hu and Wu [9] in a series of different, consecutive citation time windows following their publication in selected six sample journals. They study the influence of paper length on the chance of papers' getting cited. Multiple linear regression was also proposed as a suitable method for the log-transformed data (citation count+1) based on the simulated discrete log-normal data [10]. However, it is well known that the citation count data is right-skewed with a potential number of zeros and log transformation is not the best strategy for modeling such data. Thus, there have been attempts to utilize the generalized linear modeling like in Maliniak et al. [11] who reported a significant influence of gender variable amongst the many others by means of Negative Binomial Model. Besides they highlighted huge gap between the genders. For this conclusion, Zigerell [12] commented that more data is necessary to come through this result and stated that the gender gap is more prevalent in elite journals. Applicability of different models by right-censoring the data was also assessed in Santos and Irizo [13] so as to deal with the skewness of citation counts.

Motivated by extracting what features are responsible for particularly zero counts, we here prefer the GLM approach in modeling counts rather than classification approach. It is also of our interest to compare the predictive performances of existing discrete regression models. Understanding of the intriguing factors that influence citations can be the goal of different scientific disciplines. However, literature review we performed revealed that such modeling strategies have been applied to limited number of scientific disciplines. For example, Qian et al. [14] applied negative binomial regression models to study the effect of various factors on the citation rates in Computer science. Politics, Economy, and Business appears to be attractive fields ([1], [6], [11]). Ahlgren et al. [15] used a very large publication set, however, across all disciplines regardless of which field it is.

Various studies indicate that the citation behavior of a paper differentiates according to the scientific field, or even sub-fields. For example, the number of citations per paper is detected to be much higher in social sciences than natural sciences [16]. In some disciplines, there may be "hot" topics or sub-fields that can influence the paper to be highly-cited. For example, papers published on analytic chemistry, organic chemistry and physical chemistry receive more citations than those on biochemistry [17]. Therefore, disciplines have different citation manners and factors affecting the citation counts of the papers vary accordingly [18]. In order to give an insight about the factors affecting the citation counts, we present a collection of studies belonging to different disciplines with varying methodological aspects in Table A1. It can be seen that the considered factors can be categorized as paper-related (e.g., characteristics of title, abstract, references of the study topic), author-related (e.g., number of authors, authors' academic rank or gender) and journal-related (e.g., the scope or impact factor of the journal). As far as Statistics science is concerned, however, such a study of citation behavior for the papers of this discipline appears lacking in the literature.

Therefore, this study aims to fulfill this gap by modeling the citation counts of publications in the SCI or SCI-Exp. journal lists, belonging to Turkish academicians in the field of Statistics. It offers a modest attempt to identify some of the factors that determine the citation counts of authors who published their work in this field. Besides, on the contrary to the work listed above, we put here more focus on the modeling the articles with zero citation counts. Discrete regression models like basically Poisson or Negative Binomial could be adequate for moderate size of zero counts; however, advanced discrete regression strategy is necessary to achieve the excess of zeros. Amongst the alternative zero-inflated models, we are here motivated to perform a comparison and choose the best one.

2. DISCRETE REGRESSION MODELS

The number of occurrences of any event as a result of the trials carried out in a specified process can be expressed as count data. The count data is the type of data that the observations can only take non-negative integer values (0, 1, 2, 3…). It is well known that the application of linear regression modeling (suitable for continuous response) for count data reveal inefficient and inconsistent predictions. General Linear Models (GLM), however, is a combination of linear and nonlinear regression models that take into account the non-normality of count data [19]. For example, Poisson Regression (PR) is the first natural choice [20] and other discrete models commonly preferred to describe the relationship between variables are Negative Binomial Regression (NBR), Zero-Inflated Regression (ZIP-ZINB) and Hurdle Regression (HP-HNB).

2.1. Poisson Regression (PR)

Let Y be a non-negative integer valued random variable and has a Poisson distribution with mean (μ) parameter set as $\mu = \lambda$. Then the probability function of Y is given as

$$f(y;\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!} \quad , \ y = 0,1,2,.... \text{ and } \ \lambda > 0 \tag{1}$$

with the expected value (E) and variance (V) of the function in (1) equal to

$$E(Y) = V(Y) = \lambda \tag{2}$$

Regression modelling can be constructed through the natural log link function as

$$log(\lambda) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k \tag{3}$$

where β_i are unknown regressors and x_i are the predictors. Conditional expected value is then obtained by exponentiating both sides of (3) as below:

$$E(Y_i|x_i) = exp(x_i^{'}\beta) \tag{4}$$

2.2. Negative Binomial Regression (NB)

NB regression provides a facility to slacken the assumption that the mean is equal to the variance, essential for the Poisson model. The classical NBR model is a mixture of Poisson and Gamma distribution.

$$P(Y = y_i) = \frac{\Gamma(y_i+\theta^{-1})}{\Gamma(\theta^{-1})\Gamma(y_i+1)}\left(\frac{\theta\lambda_i}{1+\theta\lambda_i}\right)^{y_i}\left(\frac{1}{1+\theta\lambda_i}\right)^{\theta^{-1}} \quad y_i = 0,1,.... \quad (5)$$

The expected value and variance of the distributional form of (5) are

$$E(Y) = \mu \text{ and } V(Y) = \mu(1+\mu\theta) \quad (6)$$

where θ is a dispersion parameter and μ is the mean parameter.

For a standard counting model, if the data contains more zeros than expected, it is called as zero-inflation. In this case, two-part mixed models will be preferred to fit the data. These

- Zero-inflated models
- Hurdle models

2.3. Zero-Inflated Regression Models (ZIP-ZINB)

Let π is the structural zero ratio, then zero-inflated regression models can be expressed as follows:

$$P(Y = y_i) = \begin{cases} \pi_i + (1-\pi_i)P(S_i = 0) & y_i = 0 \\ (1-\pi_i)P(S_i = y_i) & y_i > 0 \end{cases} \quad (7)$$

where $P(S)$ is the probability function of random variable S for which any discrete distribution can be selected. Generally, Poisson or Negative Binomial distribution is preferred, and inserting the probability function of these in (7) results in ZIP and ZINB models respectively.

2.4. Hurdle Regression Models (HP-HNB)

This model is also the mixture of the two components, the first of which includes the binary responses showing the positive counts (1) against the zero counts (0); the second includes only positive counts.

(i) Hurdle Poisson (HNB) Model:

$$P(Y = y_i) = \begin{cases} w_i & y_i = 0 \\ (1-w_i)\frac{e^{-\mu_i}\mu^{y_i}}{(1-e^{-\mu_i})y_i!} & y_i > 0 \end{cases} \quad (8)$$

(ii) Hurdle Negative Binomial (HNB) Model:

$$P(Y = y_i) = \begin{cases} w_i & y_i = 0 \\ (1-w_i)\frac{\Gamma(y_i+\theta^{-1})}{y_i!\Gamma(\theta^{-1})}\frac{(1+\theta^{-1}\mu_i)^{-(y_i+\theta^{-1})}\theta^{-y_i}\mu_i^{y_i}}{1-(1+\theta^{-1}\mu_i)^{\theta^{-1}}} & y_i > 0 \end{cases} \quad (9)$$

In both models defined in (8) and (9), μ is the mean parameter, θ is the dispersion parameter and $w_i = P(Y_i = 0)$.

3. MODEL SELECTION CRITERIA

In this study, three information criteria based on log-likelihood, and Vuong statistic were used in deciding the appropriate model.

Information Criteria:

- $AIC = -2LL + 2k$
- $AICc = AIC + 2k(k+1)/(n-k-1)$
- $BIC = -2LL + k(ln(n))$

where AIC = Akaike information criterion, AIC_c = Corrected AIC, BIC = Bayesian information criterion, LL = log-likelihood, k = Number of parameters in the estimated model and n = Sample size.

Vuong Statistics:

Assessment of fitting performance of Model 1 vs Model 2 by means of their corresponding probability functions $P_1(.)$ and $P_2(.)$ respectively can be achieved by Voung statistic (V) as below:

$$V = \frac{\overline{m}\sqrt{n}}{S_m} \sim N(0,1) \text{ and } m_i = log\left(\frac{P_1(Y_i|X_i)}{P_2(Y_i|X_i)}\right) \quad (10)$$

- $V >$ 1.96 ⇒ 1^{st} model is preferred.
- $V <$ -1.96 ⇒ 2^{nd} model is preferred.

where $\overline{m}$ = mean and S_m= standard deviation of m_i values.

4. APPLICATION

There are 32 Statistics Departments in state and private universities in Turkey. Additionally, we searched for Statisticians employed in other departments like Econometrics, Actuary, Biostatistics, etc. and found 261 academicians in total. SCI & SCI-Exp. publications (2005-2017) of all those Statisticians were searched through Scopus, and all the analyses were performed in SPSS and R.

Response variable: Citation counts of 1529 papers published in the period of 2005-2017.

Factors:

- Age (Age of paper)
- Ref (Number of references in the paper)
- Ath (Number of authors in the paper)
- FCA (First citation age)
- Pg (Number of pages in the paper)
- TL (Title length)

Journal Categories: Artificial Intelligence (AI), Biology (Bio), Chemistry (Chm), Computer (Comp), Econometric (Eco), Education (Edu), Energy (Engy), Engineering (Eng), Environment (Env), Fuzzy (Fuz), Management (Man), Mathematics (Math), Medicine (Med), Operational Research (OR), Other (Oth), Physics (Phy), Social (Soc), Statistics (Stat)

Data were first analyzed descriptively using the demographic features. It is observed that authors of the papers are almost equally distributed in gender (Figure 1). Distribution of authors in terms of academic title is 37% assistant professors, 23% associate professors and 40% professors respectively (Figure 2). Besides, only 28% of those work for private universities (Figure 3).

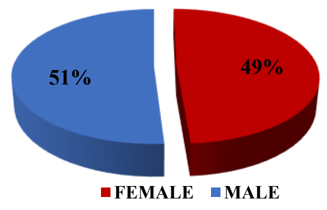


Figure 1. Distribution of authors by gender

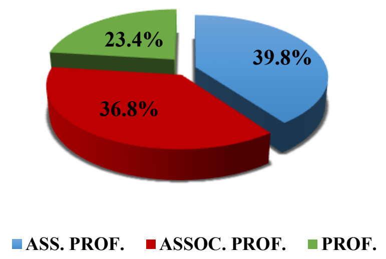


Figure 2. Distribution of authors by title

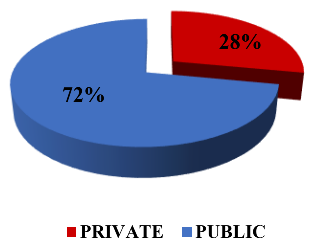


Figure 3. Distribution of authors by type of workplace

When the academicians are evaluated according to their experiences, Figure 4 reveals that majority of those (85.5%) have less than 20 years of experience.

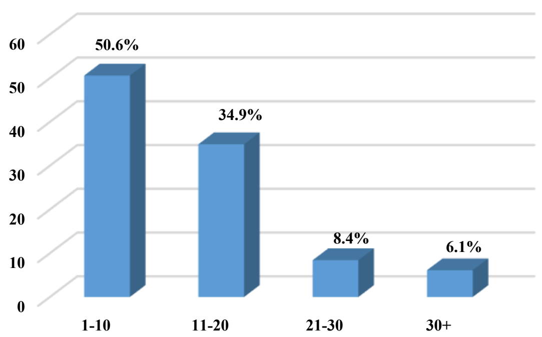


Figure 4. Distribution of authors by experience (years)

It can be observed that 79.7% of the authors have published less than 10 articles (Figure 5).

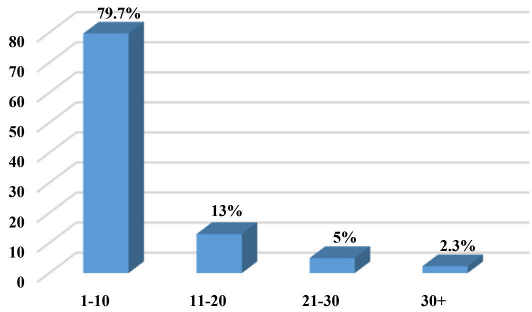


Figure 5. Distribution of authors by the number of publications

Visual presentation of how the total amount of publications is distributed amongst the authors of different titles is presented in Figure 6. To exemplify, 39.8% of our sample consists of assistant professors (Figure 2), 38.2% of whom owns 1-10 papers; 0.8% publish 11-20 papers; 0.8% publish 21-30 papers and none (0%) publishes 30+ papers as seen with blue bars of Figure 6. Publications with an amount of 1-10 seem to be highly produced by assistant professors. Although the percentages of more than 11, 21 or 30 publications are low in total, it can be seen that the number of publications naturally increases in line with the title of authors (Figure 6).

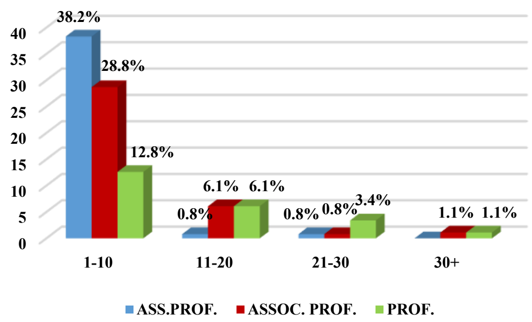


Figure 6. Number of publications by authors' title

A simultaneous evaluation of the authors with respect to their working experiences and the number of publications they realized is performed by the crosstabulation produced as in Table 1. For interpretation of such a table, let's concentrate on the authors who have 1-10 years of experience for the moment. The total of this row shows that 132 authors have such an experience, which corresponds to 50.6% of the grand total (that is 261 as seen in the right bottom corner). Cell frequencies in this row show that how 132 authors are distributed amongst the number of publication categories. For example, 118 authors of those own 1-10 number of papers. This table also presents the percentages of cell frequencies within the experience (the row total), within the number of SCI publication (the column total) and also within the grand total. For the above example, we can say that 89.4% of the authors with 1-10 years of experience (row total=132) publish 1-10 number of papers. We can additionally say that 56.7% of all authors owning 1-10 number of papers (column total=208) have 1-10 years of experience. Besides, we can also say that 45.2% of the whole sample (grand total=261) has 1-10 years of experience and also 1-10 number of papers. Therefore, such a presentation enables one to make interpretation within the row or column category as well as within the whole sample. According to totals within the whole sample in Table 1, it appears that half of authors (50.6%) have an experience less than 10 years and also the majority of the authors (79.7%) produced 1-10 number of papers. Amongst the authors experienced up to 10 years, the percentage of having 1-10 SCI publications appears to be 89.4%. Although it is not as high as this percentage, a similar pattern is observed for each experience category. That is the majority of each experience group owns 1-10 number of papers (see %within experience). If we look at "% within SCI" values, more than half of the authors with 1-10 number of papers (56.7%) have the experience less than 10 years. However, productivity corresponding to more than 10 papers (i.e., 11-20, 21-30, 30+) belongs mainly to the authors having 11-20 years (see the percentages of 47.1%, 46.2% and 83.3% for this experience group). Surprisingly, the productivity of the academicians decreases as they get more experienced.

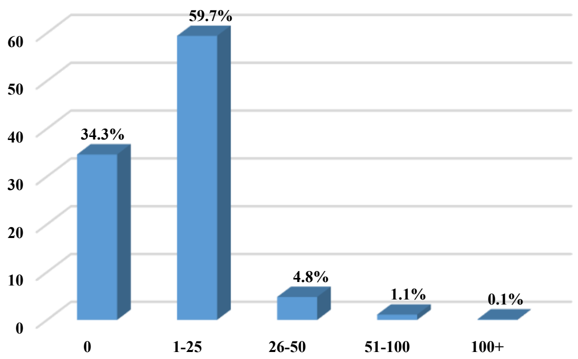


Figure 7. Distribution of citation counts

Table 1. Number of SCI publications according to experiences of authors

			# of SCI Publications				
			1-10	11-20	21-30	30+	Total
Experience	1-10	Count	118	10	3	1	132
		% within experience	89.4%	7.6%	2.3%	.8%	
		% within SCI	56.7%	29.4%	23.1%	16.7%	
		% of Total	45.2%	3.8%	1.1%	.4%	50.6%
	11-20	Count	64	16	6	5	91
		% within experience	70.3%	17.6%	6.6%	5.5%	
		% within SCI	30.8%	47.1%	46.2%	83.3%	
		% of Total	24.5%	6.1%	2.3%	1.9%	34.9%
	21-30	Count	14	5	3	0	22
		% within experience	63.6%	22.7%	13.6%	.0%	
		% within SCI	6.7%	14.7%	23.1%	.0%	
		% of Total	5.4%	1.9%	1.1%	.0%	8.4%
	30+	Count	12	3	1	0	16
		% within experience	75.0%	18.8%	6.3%	.0%	
		% within SCI	5.8%	8.8%	7.7%	.0%	
		% of Total	4.6%	1.1%	.4%	.0%	6.1%
Total		Count	208	34	13	6	261
		% of Total	79.7%	13.0%	5.0%	2.3%	100.0%

Table 2. Citation counts according to experiences of faculty member

			Citation Counts						
			0-25	26-50	51-100	101-150	151-200	200+	Total
Experience	1-10	Count	100	16	6	6	2	2	132
		% within experience	75.8%	12.1%	4.5%	4.5%	1.5%	1.5%	
		% within citation	57.5%	51.6%	26.1%	60.0%	22.2%	14.3%	
		% of Total	38.3%	6.1%	2.3%	2.3%	.8%	.8%	50.6%
	11-20	Count	51	12	12	3	3	10	91
		% within experience	56.0%	13.2%	13.2%	3.3%	3.3%	11.0%	
		% within citation	29.3%	38.7%	52.2%	30.0%	33.3%	71.4%	
		% of Total	19.5%	4.6%	4.6%	1.1%	1.1%	3.8%	34.9%
	21-30	Count	13	1	5	1	0	2	22
		% within experience	59.1%	4.5%	22.7%	4.5%	.0%	9.1%	
		% within citation	7.5%	3.2%	21.7%	10.0%	.0%	14.3%	
		% of Total	5.0%	.4%	1.9%	.4%	.0%	.8%	8.4%
	30+	Count	10	2	0	0	4	0	16
		% within experience	62.5%	12.5%	.0%	.0%	25.0%	.0%	
		% within citation	5.7%	6.5%	.0%	.0%	44.4%	.0%	
		% of Total	3.8%	.8%	.0%	.0%	1.5%	.0%	6.1%
Total		Count	174	31	23	10	9	14	261
		% of Total	66.7%	11.9%	8.8%	3.8%	3.4%	5.4%	100.0%

As far as citation counts are concerned, the distribution is obtained as in Figure 7, highly right-skewed as expected. It can be seen that there is a high percentage of publications with zero citation counts (n=525 corresponding to 34.3%).

When academicians are evaluated according to the number of citations they receive for their publications, it can be seen that the least number of citations in total belong to professors (Table 3). On the contrary, papers having more than 100 citations appear to be produced by associate professors and professors, presumably due to the most aged papers belong to those academicians.

Table 3. Citation counts by title of authors

Citation Counts	Ass. Prof.	Assoc. Prof.	Prof.
0-25	33.7%	23.4%	9.6%
26-50	4.2%	5.4%	2.3%
51-100	0.8%	3.4%	4.6%
101-150	1.1%	0.8%	1.9%
151-200		1.1%	2.3%
200+		2.7%	2.7%
Total	39.8%	36.8%	23.4%

Crosstabulation of citation count groupings against the experiences of authors is presented in Table 2. It is observed that majority of the authors own papers having less than 50 citations regardless of their experiences. Astonishingly, although only 5.4% of the authors manage to receive more than 200 citations, majority of this percent belongs to those having 11-20 years of experience. It can be concluded that the highly influential papers are more likely to be produced within 11-20 years of experience.

Table 4. Expected percentages and zeros according to all fitted models

	PR	NB	ZIP	ZINB	HP	HNB
Expected counts	133	520	525	541	525	525
Expected percentages (%)	8	34	34.3	35.4	34.3	34.3

(Observed number of zeros is 525 corresponding to 34.3%)

As mentioned above the distribution of citation counts is right-skewed in nature and generalized linear models are more suitable to represent this skewness. Besides, the high percentage of zero citation counts requires particular attention in modeling. Motivated by determining the influential factors on citation counts, PR, NB, ZIP, ZINB, HP, and HNB models were fitted the data and estimated parameters are listed in Table A2. Significant factors ($p<0.05$) are here indicated as bold.

Performance assessment in predicting zero counts for all the considered models is presented in Table 4. Noting that the observed percentage of zeros is 34.3%, Zero Inflated Poisson and Hurdle Models seem to be better than others for this purpose. As expected, PR model is not appropriate for this type of skewed data.

Table 5. Information criteria results

	PR	NB	ZIP	ZINB	HP	HNB
AIC	15586.0	7474.4	13245.5	7365.8	13251.2	7452.6
AICc	15586.8	7475.2	13248.7	7369.1	13254.4	7455.9
BIC	15615.6	7504.0	13229.2	7347.4	13234.8	7436.2
-LL	7768.0	3712.2	6574.8	3633.9	6577.6	3678.3

Model comparison in terms of information criteria as given in Table 5 suggested that ZINB and HNB models are the most suitable ones. As ZINB is slightly better than HNB, Vuong statistics would be more decisive to select the final model. This test is based on the principle that the differences between the likelihoods indicate which model fits the data better. Therefore, only significant pairwise comparisons of models achieved via Vuong test were presented in Table 6. Note that the estimated values (V) give the indication of model preferences based on the comparisons as described in Section 3.

Table 6. Estimated Vuong Statistics (v) with model preferences

Model 1 \ Model 2	NB	ZIP	ZINB	HNB
PR	-12.96 *Model 2> Model 1	-9.36 Model 2> Model 1	-13.27 Model 2> Model 1	-13.12 Model 2> Model 1
NB		10.61 Model 1> Model 2		-3.31 Model 2> Model 1
ZIP				-10.73 Model 2> Model 1
ZINB				3.93 Model 1> Model 2

For interpretation, *e.g., estimated $V = -12.96$ implies Model 2 (NB) is significantly better than Model 1 (PR).

Results indicate that ZINB is better than HNB, thus considered as the final model for the data at hand. Mathematical form of the final model can be presented as

$$\begin{aligned}\mu = exp(-1.353 &+ \mathbf{0.247}Age + \mathbf{0.017}Ref + 0.007Ath - 0.014FCA + \mathbf{2.062}Pg - \mathbf{0.002}TL + \mathbf{1.911}AI \\ &+ \mathbf{1.05}Bio + 1.221Chm + \mathbf{2.062}Comp + \mathbf{1.043}Eco + \mathbf{1.314}Edu + \mathbf{2.402}Engy + \mathbf{2.114}Eng \\ &+ \mathbf{1.316}Env + \mathbf{1.829}Fuz + 0.827Man + \mathbf{1.334}Math + \mathbf{1.676}Med + \mathbf{2.276}OR + \mathbf{1.242}Oth \\ &+ 1.062Phy + \mathbf{0.989}Soc + \mathbf{1.274}Stat)\end{aligned} \tag{10}$$

for the count part, and

$$\begin{aligned}(\pi/(1-\pi)) = exp(&7.046 - \mathbf{2.378}Age - 0.03Ref - 0.034Ath + 0.029FCA - 3.485Pg + 0.004TL - 19.7AI \\ &- 0.494Bio - 0.378Chm - 3.485Comp - 2.134Eco - 3.297Edu - \mathbf{4.884}Engy - \mathbf{4.367}Eng \\ &- 3.48Env - 2.113Fuz - 3.861Man - 3.36Math - \mathbf{3.408}Med - 1.648OR - 4.107Oth \\ &- 0.954Phy - 3.957Soc - 2.903Stat)\end{aligned} \quad (11)$$

for the binary part. The regression coefficients given as bold are the significant factors at 0.05 significance level.

5. CONCLUSION

There is ever growing interest for the quality assessment of research papers and citation count has been considered as the best indicator for this purpose. Although this measure is frequently used in variety of disciplines, there is a lack of interest for Statistics science. In this respect, our study fulfilled this gap and presented influential factors that affect the citation counts of the papers in Statistics. This study also investigates the suitability of the models within GLM framework. Due to the high percentage of zero citation counts and skewness of its distribution ZINB model is concluded to be the best amongst the others to model the citation counts.

In the final model, the citation counts are observed to be positively related to the age of the paper, the number of references and the number of pages in the paper. It is natural to observe the amount of citation to increase as the paper gets more aged. The variety of references as a reflection of the authors' knowledge also increases the frequency of the citation. The number of references in a paper has been stated as a good predictor for the citation behavior in many studies (e.g., [2], [8], [17], [21]). Besides, the length of the paper seems to rise up the citation counts as also stated by Stremersch et al. [22]. However, the question of what the optimal number is requires a special methodological attention and the answer differs from discipline to discipline. Vaio et al. [1] detected the optimal number of pages as 36 for Economic History, however, Robson and Mousquès [23] stated this number as 11 for Environmental Modeling discipline.

As reported in many studies, we here also detected that the study topics or sub-fields of Statistics have high impact on the citation of the papers. Fields of Artificial Intelligence, Biology, Computer, Econometric, Education, Energy, Engineering, Environment, Fuzzy, Mathematics, Medicine, Operational Research, and Social studies are observed to be the most influential subjects.

On the other hand, our model suggests that title length affects the citation count negatively. That is, the papers with longer titles receive significantly less citation than those with shorter titles. Such an effect has also been stated by a variety of studies in the literature (e.g., [24], [25]). However, it must be noted that this influence can depend on the discipline of the study as the papers of Medicine with longer titles were detected to receive more citations [26].

In addition, binary part of our model suggests that the probability of observing zero citation can be reduced by publishing the papers in the disciplines of Energy, Engineering, and Medicine. It can also be concluded that the higher the age of paper is, the lesser zero counts is. This is a natural result as the paper gets older, the probability of receiving citation increases. This effect is also reflected in positive count part of the model as the coefficient for the age of the paper variable is significant and positive.

In order to increase the citation counts, academicians studying in the field of Statistics are recommended to keep the title of the paper short, increase the number of references, and produce immense papers with more pages. Additionally, they can lead their interest to the statistical applications of particularly Operational Research, Computer Science, Artificial Intelligence, Fuzzy as well as Energy, Engineering, and Medicine.

For those who are interested, the study of the optimal number of pages for the papers of Statistics discipline can be suggested as a future work. Besides, models constructed here can be reproduced by enlarging the variety of the factors that presumably have impact on the citation behavior in Statistics.

Author contributions: The authors contributed equally to the study.
Conflict of Interest: No conflict of interest was declared by the authors.
Financial Disclosure: This paper is a part of Sinop University Scientific Research Project: FEF-1901-16-07.
Acknowledgments: The authors would like to thank the Editors and anonymous reviewers for their valuable comments.

REFERENCES

[1] G. Di Vaio, D. Waldenström and J. Weisdorf, Citation success: Evidence from economic history journal publications. Explorations in Economic History, vol. 49, no. 1, pp. 92-104, 2012.

[2] N. Onodera and F. Yoshikane, Factors affecting citation rates of research articles. Journal of the Association for Information Science and Technology, vol. 66, no. 4, pp.739–764, 2015

[3] M. Y. Wang, G. Yu, and D.R. Yu, Mining typical features for highly cited papers. Scientometrics, vol. 87 no. 3, pp. 695–706, 2011.

[4] L. Fu and C. Aliferis, Using content-based and bibliometric features for machine learning models to predict citation counts in the biomedical literature. Scientometrics, vol. 85, no.1, pp. 257-270, 2010.

[5] X. Ruan, Y. Zhu, J. Li and Y. Cheng, Predicting the citation counts of individual papers via a BPneural network. Journal of Informetrics, 14, 2020.

[6] J. Mingers, F. Macri and D. Petrovici, Using the h-index to measure the quality of journals in the field of Business and Management. Information Processing & Management, vol. 48, no. 2, pp. 234-241, 2012.

[7] C. Lokker, K.A. McKibbon, R.J. McKinlay, N.L. Wilczynski and R.B. Haynes, Prediction of Citation Counts for Clinical Articles at Two Years Using Data Available Within Three Weeks of Publication: Retrospective Cohort Study. Bmj, vol. 336, no. 7645, pp. 655-657, 2008.

[8] C. Chen, Predictive effects of structural variation on citation counts. Journal of the Association for Information Science and Technology, vol. 63, no. 3, pp. 431-449, 2012.

[9] Z. Hu and Y. Wu, Regularity in the time-dependent distribution of the percentage of never- cited papers: An empirical pilot study based on the six journals. Journal of Informetrics, vol. 8, no. 1, pp. 136–146, 2014.

[10] M. Thelwall and P. Wilson, Regression for Citation Data: An Evaluation of Different Methods. Journal of Informetrics, vol. 8, no. 4, pp. 963–971, 2014.

[11] D. Maliniak, R. Powers and B.F. Walter, The gender citation gap in international relations. International Organization, vol. 67, no. 4, pp. 889-922, 2013.

[12] L.J. Zigerell, Is The Gender Citation Gap in International Relations Driven by Elite Papers?. Research & Politics, April-June, 1-7, 2015.

[13] J.B. Santos and F.J.O. Irizo, Modelling Citation Age Data with Right Censoring. Scientometrics, vol. 62, no. 3, pp. 329-342, 2005.

[14] Y. Qian, W. Rong, N. Jiang, J. Tang and Z. Xiong, Citation regression analysis of computer science publications in different ranking categories and subfields. Scientometrics, vol. 108, pp. 1–24, 2017.

[15] P. Ahlgren, C. Colliander and P. Sjogarde, Exploring the relation between referencing practices and citation impact: A large-scale study based on Web of Science data. Journal of the Association for Information Science and Technology, vol. 69, no.5, pp. 728–743, 2018.

[16] P.F. Skilton, Does the human capital of teams of natural science authors predict citation frequency?. Scientometrics, vol. 78, no. 3, pp. 525–542, 2009.

[17] L. Bornmann, H. Schier, W. Marx, and H.D. Daniel, What factors determine citation counts of publications in chemistry besides their quality?. Journal of Informetrics, vol. 6, no. 1, pp. 11–18, 2012.

[18] I. Tahamtan, A.S. Afshar, and K. Ahamdzadeh, Factors affecting number of citations: a comprehensive review of the literature. Scientometrics, vol. 107, no. 3, pp. 1195-1225, 2016.

[19] J. Hardin and J. Hilbe, Generalized linear models and extensions. Texas, USA: Stata Corporation, 2012.

[20] E. Arıcan, Nitel yanıt değişkene sahip regresyon modellerinde tahmin yöntemleri. Master Thesis. Cukurova University, Institute of Science, 2010. https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp.

[21] F. Didegah, and M. Thelwall, Determinants of research citation impact in nanoscience and nanotechnology. Journal of the American Society for Information Science and Technology, vol. 64, no. 5, pp. 1055–1064, 2013.

[22] S. Stremersch, I. Verniers, and P.C. Verhoef, The quest for citations: Drivers of article impact. Journal of Marketing, vol. 71, no. 3, pp. 171-193, 2007.

[23] B.J. Robson, and A. Mousquès, Predicting citation counts of environmental modelling papers. 7th International Congress on Environmental Modelling and Software - San Diego, California, USA - June 2014.

[24] I. Ayres, and F.E. Vars, Determinants of citations to articles in elite law reviews. The Journal of Legal Studies, vol. 29(S1), pp. 427-450, 2000.

[25] S. Stremersch, N. Camacho, S. Vanneste, and I. Verniers, Unraveling scientific impact: Citation types in marketing journals. International Journal of Research in Marketing, vol. 32, no. 1, pp. 64-77, 2015.

[26] M. E. Falagas, A. Zarkali, D.E. Karageorgopoulos, V. Bardakas, and M.N. Mavros, The impact of article length on the number of future citations: A bibliometric analysis of general medicine journals. PLoS One, 8(2), e49476, 2013.

[27] X. Bai, F. Zhang and I. Lee, Predicting the citations of scholarly paper. Journal of Informetrics, vol. 13, pp. 407–418, 2019.

[28] H. Beydokhti, N. Riahinia, H.R. Jamali, S. Asadi and S.M. Riahi, Factors Affecting the Number of Citations to Clinical Therapeutic Articles Mentioning Level of Evidence. Mod Care J. vol. 17, no. 2, 2020.

[29] Z. Su, Prediction of future citation count with machine learning and neural network. 2020 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 14-16 April 2020.

Appendix A

Table A1. A collection of study examples on citation

Author	Discipline	Method	Influential factors
Santos and Irizo [13]	• Applied Economics	• Maximum likelihood estimation with data censored to the right (Log-normal, Weibull and Log-logistic)	• Log-logistic model is the best fit
Lokker et al. [7]	• Medicine	• Multiple regression	• The number of authors • Selection for abstraction in a synoptic journal • Clinical relevance score • Number of pages • Structured abstract • Number of cited references • Original papers • Multicentered study • Study about therapy
Fu and Aliferis [4]	• Biomedical publications • Cardiology • Endocrinology • Gastroenterology • Hematology • Medical Oncology • Nephrology • Pulmonary disease • Rheumatology	• Machine learning methods (SVM models) • Logistic regression-based classifier	• Citation history of authors • Certain topics • High impact journal
Wang et al. [3]	• Astronomy • Astrophysics	• Data mining	• The number of citations that papers obtain • Authors with high reputations receive disproportionately more citations than authors with low reputations • The reputation of a journal
Vaio et al. [1]	Economic history	• Poisson and negative binomial regression modeling	• Professors at economics and history departments • Length of papers • Co-authors • Conference and workshops presentations • Anglo-Saxon and German history • Gender
Chen [8]	• Terrorism • Mass extinction • Complex network analysis • Knowledge domain visualization	• Negative binomial models of complex network analysis • Zero-inflated negative binomial regression models	• The number of coauthors • The number of references
Maliniak et al. [11]	• Teaching, • Research, • International Policy	• Network analysis • Negative-binomial model	• Women tend to cite themselves less than men • Men tend to cite men more than women • Age of paper • Co-authorship employed by research university
Hu and Wu [9]	• Information Science • Multi-disciplinary Science	• Negative exponential model	• The length of a paper

Table A1. (Continues) A collection of study examples on citation

Author	Discipline	Method	Influential factors
Onodera and Yoshikane [2]	• Condensed matter physics, • Inorganic and nuclear chemistry • Electric and electronic engineering • Biochemistry and molecular biology, • Physiology, • Gastroenterology)	• Negative binomial multiple regression	• The price index • Number of references
Oian et al. [14]	• Computer science	• Negative binominal regression model	• Classification of a publication • Number of authors • Maximum h-index of all authors of a paper • Average number of papers published by a publication
Ahlgren et al. [15]	• Bibliometric	• Quantile regression	• Number of cited references • References to more recent publications
Ruan et al. [5]	• Library • Information • Documentation	• Four-layer back propagation (BP) neural network model	• Citations in the first two years • First-cited age • Paper length • Month of publication • Self-citations of journals
Bai et al. [27]	• Physics	• PPI model • Multi-feature model	• Inherent quality of scholarly paper • Scholarly paper impact decaying over time • Early citations • Early citers' impact
Beydokhti et al [28]	• Medicine	• Basic statistical methods	• Journals' impact factor • Level of evidence • Number of references • Number of authors • Number of title words • Length of article • Subject • Type of study design • Geographical area of corresponding author • Journal and publisher
Su [29]	• Physiology	• SVM • Decision Tree • Random Forest • Neural Network	• The sum of citing countries • The number of citing organizations • Total number of citing journals • The amount of citing subjects • The sum of citing languages • Average citation counts obtained • Average increment of citation counts obtained • The sum of funding organizations

Table A2. Estimated parameters for PR, NB, ZIP, ZINB, HP, and HNB Models.Significant factors (p<0.05) are highlighted bold

	PR		NB		ZIP				ZINB				HP				HNB			
	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate/Pr(>|z|) Count Part		Estimate/Pr(>|z|) Binary Part		Estimate/Pr(>|z|) Count Part		Estimate/Pr(>|z|) Binary Part		Estimate/Pr(>|z|) Count Part		Estimate/Pr(>|z|) Binary Part		Estimate/Pr(>|z|) Count Part		Estimate/Pr(>|z|) Binary Part	
Int	**-1.639**	0.000	**-2.288**	0.000	**-1.144**	0.000	1.166	0.088	**-1.353**	0.000	**7.046**	0.000	**-1.195**	0.000	**-2.214**	0.000	**-2.136**	0.000	**-2.214**	0.000
Age	**0.248**	< 2e-16	**0.323**	< 2e-16	**0.189**	< 2e-16	**-0.388**	< 2e-16	**0.247**	< 2e-16	**-2.378**	0.000	**0.188**	< 2e-16	**0.409**	< 2e-16	**0.268**	< 2e-16	**0.409**	< 2e-16
Ref	**0.017**	< 2e-16	**0.018**	0.000	**0.015**	< 2e-16	**-0.019**	0.000	**0.017**	0.000	-0.030	0.079	**0.015**	< 2e-16	**0.022**	0.000	**0.016**	0.000	**0.022**	0.000
Ath	**0.009**	0.000	0.011	0.085	**0.004**	0.003	-0.034	0.132	0.007	0.257	-0.034	0.279	**0.004**	0.003	0.033	0.126	0.005	0.454	0.033	0.126
FCA	**-0.062**	< 2e-16	**0.089**	0.000	**-0.014**	0.000	0.007	0.608	-0.014	0.066	0.029	0.458	**-0.014**	0.000	-0.011	0.432	-0.016	0.085	-0.011	0.432
Pg	**-0.015**	0.000	**-0.019**	0.017	**1.982**	0.000	-0.869	0.271	**2.062**	0.000	-3.485	0.066	**2.039**	0.000	**1.782**	0.002	**2.458**	0.000	**1.782**	0.002
TL	**-0.002**	0.000	**-0.002**	0.034	0.000	0.292	**0.007**	0.000	**-0.002**	0.035	0.004	0.538	0.000	0.315	**-0.007**	0.000	0.000	0.981	**-0.007**	0.000
AI	**0.944**	0.001	**2.627**	0.000	**1.963**	0.000	0.673	0.500	**1.911**	0.000	-19.700	0.988	**2.011**	0.000	0.241	0.779	**3.010**	0.000	0.241	0.779
Bio	**1.164**	0.000	**1.167**	0.003	**0.951**	0.000	-0.595	0.469	**1.050**	0.010	-0.494	0.811	**1.013**	0.001	**1.250**	0.030	**1.253**	0.019	**1.250**	0.030
Chm	**1.291**	0.000	1.218	0.052	**1.724**	0.000	0.947	0.371	1.221	0.067	-0.387	0.876	**1.778**	0.000	-0.062	0.947	**1.907**	0.028	-0.062	0.947
Comp	**2.128**	< 2e-16	**2.266**	0.000	**1.982**	0.000	-0.869	0.271	**2.062**	0.000	-3.485	0.066	**2.039**	0.000	**1.782**	0.002	**2.458**	0.000	**1.782**	0.002
Eco	**1.214**	0.000	**1.202**	0.002	**1.213**	0.000	0.307	0.686	**1.043**	0.010	-2.134	0.401	**1.271**	0.000	0.513	0.345	**1.552**	0.004	0.513	0.345
Edu	**1.839**	0.000	**1.567**	0.000	**1.617**	0.000	-0.898	0.279	**1.314**	0.003	-3.297	0.201	**1.675**	0.000	**1.762**	0.005	**1.627**	0.004	**1.762**	0.005
Engy	**2.776**	< 2e-16	**2.721**	0.000	**2.551**	< 2e-16	-1.492	0.070	**2.402**	0.000	**-4.884**	0.024	**2.608**	< 2e-16	**2.415**	0.000	**2.828**	0.000	**2.415**	0.000
Eng	**2.360**	< 2e-16	**2.384**	0.000	**2.180**	< 2e-16	-1.153	0.093	**2.114**	0.000	**-4.367**	0.014	**2.238**	0.000	**2.075**	0.000	**2.535**	0.000	**2.075**	0.000
Env	**1.712**	0.000	**1.440**	0.000	**1.546**	0.000	-0.269	0.713	**1.316**	0.001	-3.480	0.086	**1.605**	0.000	**1.113**	0.028	**1.701**	0.001	**1.113**	0.028
Fuz	**2.071**	< 2e-16	**1.939**	0.000	**1.980**	0.000	-0.205	0.795	**1.829**	0.000	-2.113	0.251	**2.038**	0.000	1.128	0.055	**2.268**	0.000	1.128	0.055
Man	**0.727**	0.021	**1.178**	0.032	**0.940**	0.010	-0.483	0.623	0.827	0.138	-3.861	0.147	**0.921**	0.022	1.253	0.104	1.086	0.131	1.253	0.104
Math	**1.695**	0.000	**1.518**	0.000	**1.668**	0.000	-0.096	0.885	**1.334**	0.000	-3.362	0.054	**1.725**	0.000	**0.986**	0.016	**1.819**	0.000	**0.986**	0.016
Med	**1.972**	< 2e-16	**1.868**	0.000	**1.814**	0.000	-0.639	0.339	**1.676**	0.000	**-3.408**	0.045	**1.872**	0.000	**1.523**	0.000	**2.065**	0.000	**1.523**	0.000
OR	**2.751**	< 2e-16	**2.532**	0.000	**2.616**	< 2e-16	-0.504	0.600	**2.276**	0.000	-1.648	0.489	**2.672**	< 2e-16	1.429	0.077	**2.817**	0.000	1.429	0.077
Oth	**1.668**	0.000	**1.496**	0.001	**1.658**	0.000	0.353	0.664	**1.242**	0.004	-4.107	0.094	**1.716**	0.000	0.523	0.398	**1.871**	0.001	0.523	0.398
Phy	**1.460**	0.000	**1.131**	0.037	**1.247**	0.000	-0.419	0.665	1.062	0.060	-0.954	0.697	**1.313**	0.000	1.149	0.121	1.223	0.073	1.149	0.121
Soc	**1.274**	0.000	**1.232**	0.001	**1.164**	0.000	-0.287	0.700	**0.989**	0.012	-3.957	0.052	**1.219**	0.000	**1.058**	0.038	**1.365**	0.009	**1.058**	0.038
Stat	**1.563**	0.000	**1.397**	0.000	**1.563**	0.000	-0.009	0.989	**1.274**	0.000	-2.903	0.086	**1.621**	0.000	**0.879**	0.028	**1.727**	0.000	**0.879**	0.028

Research Paper
Academic Platform Journal of Engineering and Smart Systems (APJESS) 10(3), 168-173, 2022
ISSN: 2822-2385
Received: 26-Jun-2022 Accepted: 12-Sep-2022
https://doi.org/10.21541/apjess.1135939

Modelling Genotoxic Effects of Metal Oxide Nanoparticles using QSAR Approach

*[1]Ceyda ÖKSEL KARAKUŞ

[1]Department of Bioengineering, Izmir Institute of Technology, İzmir, Turkey, ceydaoksel@gmail.com

Abstract

We investigated the application of structure-activity relationship approaches to underpin structural properties that potentially control the genotoxic potential of 9 different metal oxide nanoparticles (CuO, ZnO, NiO, SiO_2, TiO_2, CeO_2, Fe_2O_3, Fe_3O_4 and Co_3O_4). In particular, we compiled a pool of quantum-mechanical, experimental and periodic table-driven descriptors and explored their distinctive contribution to the measured activity (genotoxicity). We first employed a clustered heatmap and parallel coordinates plot for visual exploration of the clusters and outliers of the data and finding corresponding responsible physicochemical descriptors. We then investigated the strength (and direction) of the relationship among descriptors and between descriptors and genotoxicity using similarity metrics. By using orthogonal projections to latent structures (OPLS), we were able to quantify the relative contribution of each descriptor to the genotoxicity of metal oxide nanoparticles. Our results suggested that zeta potential, the ratio of core electrons to valence electrons, Fermi energy and electronegativity were significant predictors of genotoxicity. Such computer-assisted approaches hold considerable promise for maximizing the use of accumulated data in nanotoxicology, prioritizing nanoparticles for further testing and filling data gaps required for hazard assessment processes.

Keywords: Genotoxicity; Nanoparticles; Nanotoxicology; Computational toxicology; QSAR

1. INTRODUCTION

Ongoing globalization of nano-sciences has not only made exposure to nano-enabled materials inevitable, but also raised new challenges in the safety and regulatory domains. In particular, unique physicochemical properties of metal oxide (MO) nanoparticles (NPs) have led to an ongoing increase in their manufacture and commercial use. Studies to date have shown that MONPs are capable of inducing cytotoxic [1], genotoxic/oxidative [2], neurotoxic [3, 4] and inflammatory effects [5]. Despite recent collaborative research efforts to tackle safety issues linked to MONPs, a consensus on how to properly assess their health risks has not been reached yet. Advances in molecular biology and information science have led to a paradigm shift in hazard assessment of both traditional and advanced chemicals, from heavily relying on animal-based tests, towards developing non-testing alternatives. The term "non-testing" in the context of hazard assessment refers to the use of computational technologies and to make sound risk management decisions regarding the safe manufacturing and use of products at lower costs and in less time along the path from design to commercial manufacturing and use. Non-testing approaches hold considerable promise for making better use of accumulated hazard data and contributing to the needs of the safety assessment of new chemicals at early design stages.

Non-testing methods are generally placed into three broad categories: 1) grouping and read-across approaches, 2) structure-activity relationship (SAR) methods and 3) expert systems. While each of these methods differs in advanced settings with regard to the type of input data, methodology and level of complexity involved, they all depend on the same assumption that structurally similar compounds behave similarly in biological systems. In the context of chemical safety assessment, machine learning and statistical approaches correlating activity with the structure are widely used for categorization of chemicals based on hazard potential and prioritization for further testing, identification of hazard and hazard-related physicochemical properties and minimization of hazard by modifying toxicity-related properties (so-called *safe-by-design*). While the use of computational approaches in the safety assessment of conventional (drug-like) compounds is common practice, especially in pharmaceutical sciences, their application to NPs is still in its infancy and requires further investigation.

* Corresponding Author

Here, we related the genotoxicity of MONPs that are commonly used in commercial applications to their physicochemical properties using SAR approaches. In particular, the relationship between descriptors encoding physicochemical properties and their contribution to the genotoxicity of MONPs was visually explored by high-dimensional data visualization tools and modelled by a modified version of the partial least squares regression technique. A Pearson's correlation matrix is employed to quantify the direction and strength of the relationship among descriptors and between descriptors and genotoxicity.

2. MATERIALS AND METHODS

2.1. Toxicity Data

Data on the genotoxic potential of 9 different MONPs (CuO, ZnO, NiO, SiO_2, TiO_2, CeO_2, Fe_2O_3, Fe_3O_4 and Co_3O_4) have been collected from the literature [6-8]. The collected toxicity data involve NP-induced DNA damage in lung epithelial cells (A549) measured by the alkaline version of the Comet assay following 4-hr exposure to 20 µg/cm^2 and 40 µg/cm^2 of MONPs.

2.2. Descriptors

Descriptor data used in SAR analysis are given in Table 1.

Table 1. Experimentally-measured, quantum mechanically-computed and periodic table-based descriptors used in this study

	No	Descriptor	Explanation
Experimental	EX1	Size_TEM	Mean particle size
	EX2	Size_DLS	Hydrodynamic particle size
	EX3	Zeta	Zeta potential
	EX4	ROS	Reactive Oxygen Species
Quantum Mechanical	QM1	F_energy	Fermi energy
	QM2	HOMO	Highest occupied molecular
	QM3	LUMO	Lowest unoccupied energy
	QM4	B_gap	HOMO-LUMO band gap
	QM5	Hardness	Chemical hardness
	QM6	C_pot	Chemical potential
	QM7	ΔH_f	Enthalpy of formation
	QM8	χ	Electronegativity
Periodic Table-based	PT1	MW	Molecular weight, g/mol
	PT2	No_Me	Number of metals
	PT3	No_Ox	Number of oxygen atoms
	PT4	EN_Me	Electronegativity of metal
	PT5	T_EN_Me	Total metal electronegativity
	PT6	T_EN/No_Ox	Electronegativity per oxygen
	PT7	Ox_St	Oxidation state
	PT8	Z_Me	Atomic number
	PT9	Zv_Me	Valence electrons of metal
	PT10	PT_Me	Periodic table number of metal
	PT11	(Z_Zv)/Zv	Core electrons/valence
	PT12	V_Me	Valency of metal

2.2.1. Experimental Descriptors

A set of four experimental descriptors including mean particle size measured by Transmission Electron Microscopy (TEM), hydrodynamic diameter and zeta potential measured by Dynamic Light Scattering (DLS) and reactive oxygen species (ROS) formation capacity was collected from literature [6-8] and used as input parameters in SAR analysis.

2.2.2. Quantum-Mechanical Descriptors

A set of 8 quantum mechanical descriptors derived based on density functional theory (DFT) including Fermi energy, highest occupied molecular orbitals (HOMO), lowest unoccupied energy orbitals (LUMO), band gap, hardness, chemical potential, enthalpy of formation, and electronegativity was collected from the literature [9].

2.2.3. Periodic Table-Based Descriptors

A set of 12 descriptors including molecular weight, number of metals and oxygen atoms, electronegativity of metal, total metal electronegativity per oxygen atoms, oxidation state, atomic number, periodic table number of metal, the ratio of the number of core electrons to the number of valence electrons and valency of metal was calculated based on the molecular formula and the periodic table of the elements.

2.3. Data Scaling

Prior to data exploration and modelling, the z-score transformation was carried out for normalizing each value in the dataset using the following formula:

$$z - score = \frac{actual\ value - \mu}{\sigma}$$

where µ is the mean and σ is the standard deviation.

2.4. Exploratory Analysis

For the visual exploration of multivariate descriptor data in relation to toxicity potential, two different techniques, parallel coordinates and heatmaps, were employed. Pearson correlation matrix was constructed to determine inter-correlated descriptors and how strongly each of these descriptors is related to toxicity. All analyses were carried out in R Version 4.2.0 [10] using the packages "pheatmap", "Hmisc", "corrplot", and "PerformanceAnalytics".

2.5. Modelling

Partial least squares (PLS) is a modelling technique that combines dimensionality reduction through principal component analysis (PCA) and regression [11]. While unsupervised PCA is mainly focusing on finding the compressed representation of predictor variables only, in supervised PLS regression, the correlation between the predictor and outcome variables is taken into account to compute principal components which are then used to fit a regression model. One of the most recent extensions of PLS is called orthogonal projections to latent structures (OPLS) which decomposes the PLS solution into predictive (correlated with the response) and orthogonal (uncorrelated with the response) components [12]. To model the relationship between descriptors and the genotoxicity potential of MONPS, OPLS regression was performed on the scaled data using the SIMCA software version 17.

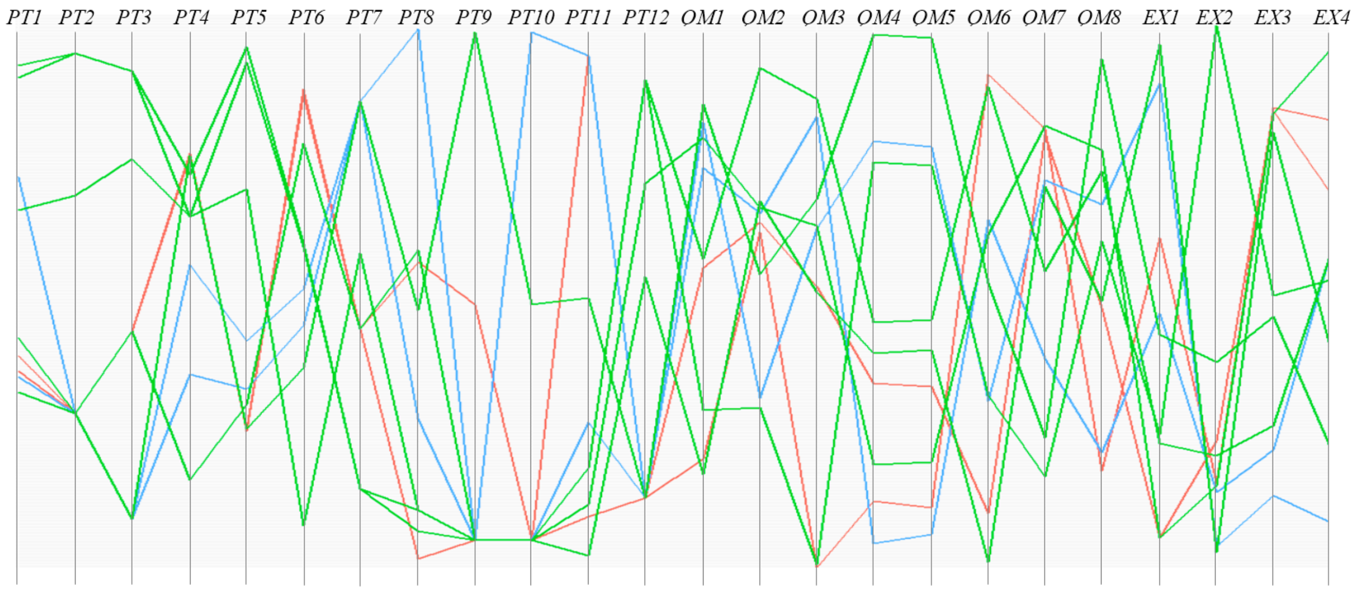


Figure 1. Periodic table (PT) descriptors (PT1-12), Quantum Mechanical (QM) descriptors (QM1-8) and Experimental (EX) descriptors (EX1-4) of 9 MoNPs coloured according to high (red), medium (blue) and low (green) values of genotoxicity at 40 μg/cm^2

3. RESULTS

For the purpose of data exploration, three different techniques including parallel coordinates, heat maps and Pearson's correlation matrix were used. In parallel coordinates, the points used in Euclidean space are represented as a series of lines passing through parallel axes where each variable is represented by one parallel axis. Parallel coordinate visualization of the descriptor data is given in Figure 1. The results associated with high, medium and low genotoxicity are highlighted in red, blue and green, respectively. The parallel coordinate plot shows that electronegativity (QM8) is inversely related to genotoxicity while higher values of zeta potential (EX3) lead to a higher genotoxicity level. Similar observations can be made from Pearson's correlation matrix given in Figure 2.

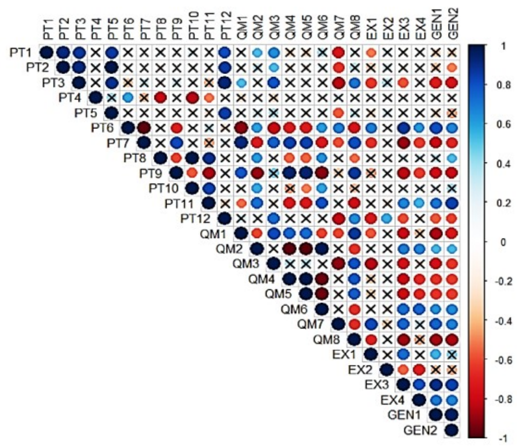


Figure 2. Pearson's correlation matrix displaying correlations between descriptors and genotoxicity [from –1 (blue) to +1 (red)]. Color intensity and circle size are proportional to the strength of correlation.

In Figure 2, positive and negative correlations are displayed in blue and red, respectively, while x indicates statistically non-significant correlations based on p-value of 0.05. GEN1 and GEN2 refer to NP-induced DNA damage in A549 following 4-hr exposure to 20 μg/cm^2 and 40 μg/cm^2 of MONPs. It is clear from Figure 2 that number of Oxygen atoms (PT3), valence electrons of metal (PT9), Fermi energy (QM1) and electronegativity (QM8) are negatively correlated with DNA damage while the ratio of core electrons to valence electrons (PT11) and zeta potential (EXP3) are positively correlated with DNA damage. We then combined heat map with clustering to prioritize MONPs based on genotoxic potential and to demonstrate the physicochemical differences between the ones belonging to different genotoxicity clusters (Figure 3).

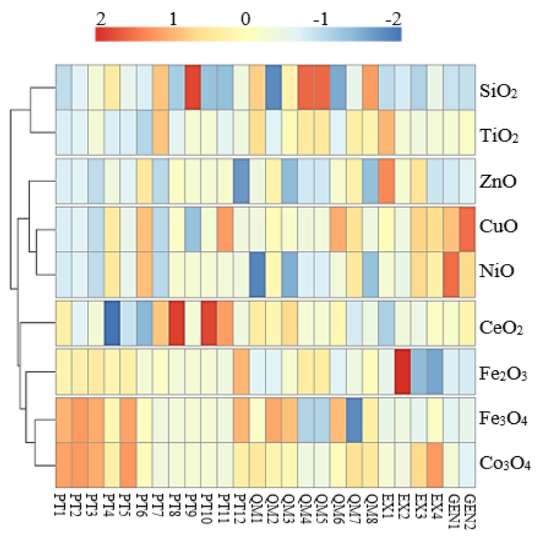


Figure 3. Clustered heat map displaying auto-scaled descriptor and genotoxicity values of 9 MONPs

Several observations can be made from the cluster heat map given in Figure 3. In particular, two MONPs, CuO and NiO are clearly distinguished from the rest due to their relatively high genotoxicity potential. Another interesting finding is that SiO_2 which has the lowest genotoxicity at both measured doses also has the most extreme (e.g. the lowest or the highest) values in 14 descriptors (out of 24), suggesting the great influence of physicochemical properties on toxicity.

As a next step, OPLS has been carried out to model the relationship between 24 descriptors and 2 genotoxicity results simultaneously. The R^2 values reflecting the model fit are found to be 0.83 and 0.85 for genotoxicity measured at 20 and 40 µg/cm^2, respectively. The resulting score scatter, loadings and variable importance plots are given in Figures 4–6.

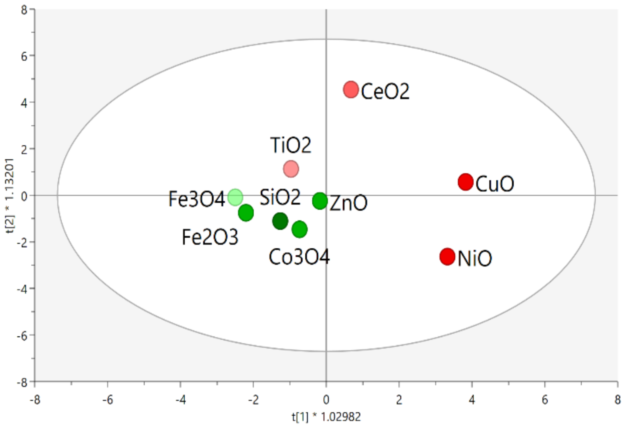


Figure 4. OPLS t1/t2 score plot showing the relationship between MONPs (red and green indicate higher and lower values of genotoxicity, respectively)

As can be seen from Figure 4, three particular MONPs with relatively higher genotoxicity potential, CuO, NiO and CeO_2, are clearly separated from the main cluster formed of MONPs with lower-genotoxicity values. The OPLS loadings given in Figure 5 display how the x-variables are combined to form the OPLS-score plot given in Figure 4.

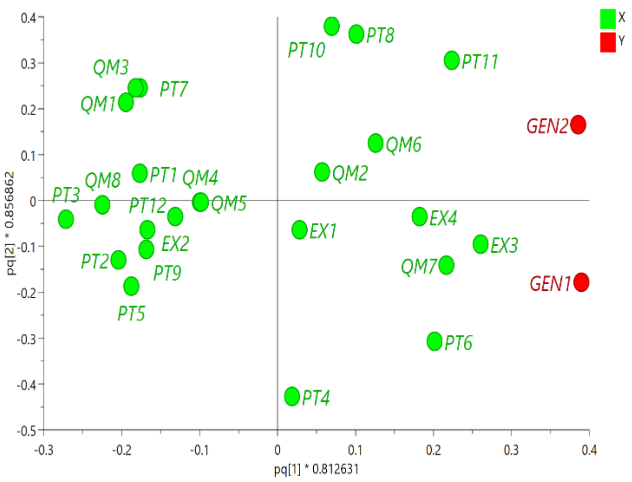


Figure 5. OPLS loadings plot showing the relationship between 24 descriptors (x variables) and 2 genotoxicity responses (y variables)

By looking at the loading plots given in Figure 5, we can identify the descriptors contributing to the positioning and separation of MONPs shown in Figure 4. For the ease of interpretation, one can imagine a straight line passing through the response variable (GEN1 or GEN2) and the origin of the plot. When the descriptors (x variables) are perpendicularly projected onto this imaginary line, their distance from the origin signifies their contribution to the response variable. Based on this, it can be concluded that mean particle size measured by TEM (EX1) has a near-zero contribution to genotoxicity while zeta potential (EX3), number of oxygen atoms (PT3), the ratio of core electrons to valence electrons (PT11) and electronegativity (QM8) have the largest impact on genotoxic potential.

The variable importance in projection (VIP) plot showing the relative contribution of each descriptor to genotoxicity is given in Figure 6. The variables with the highest VIP scores are the most influential parameters for genotoxicity.

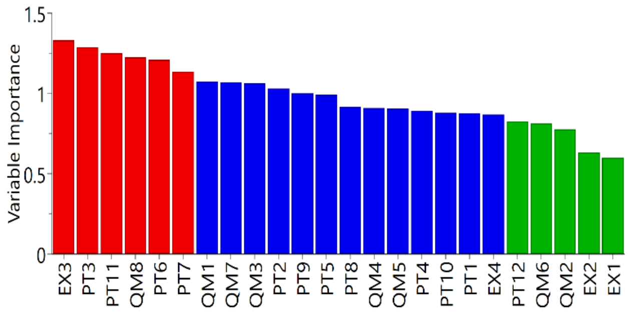


Figure 6. Variable importance in projection (VIP) plot summarizing each variable's contribution to the OPLS model

4. DISCUSSION

Using a combination of multidimensional data visualization techniques, similarity metrics, clustering and regression-based models, we identified intrinsic physicochemical properties that are seemingly related to the genotoxicity potential of MONPs. To ensure the collection of homogenous genotoxicity data, we restricted our selection to a specific cell line (lung epithelial cells), assay (alkaline Comet assay), NP class (metal oxides), exposure duration (4 hours) and concentration (20 and 40 µg/cm^2). We combined literature collected experimental [6-8] and theoretical [9] descriptor data with simple periodic-table driven descriptors to encode the physicochemical properties of MONPs.

Our results suggested that NiO and CuO, induced the highest level of DNA damage in the human lung epithelial cells, followed by CeO_2 and TiO_2. Indeed, the high genotoxic potential of both NPs has been frequently reported in the literature [13-18]. The next step was to explore the physicochemical drivers of the observed high and low genotoxicity. Out of 24 experimentally-measured and theoretically calculated descriptors, four particular ones, zeta potential, number of oxygen atoms, the ratio of core electrons to valence electrons and electronegativity were repeatedly found to be correlated with the observed genotoxicity. In particular, highly positive (>30 mV) values of low molecular weight (<80 g/mol) MONPs led to high genotoxicity (>20% DNA damage compared to the control). The close correlation between positive surface charge and high toxicity is generally attributed to the increased capacity of positively-charged MONPs to interact with the negatively-charged cell membrane leading to higher levels of cellular uptake [19]. Despite its high zeta potential, Co_3O_4 did not induce any DNA damage (potentially) due to its high molecular weight (241 g/mol) and/or the high number of oxygen atoms.

A reverse relationship between electronegativity and genotoxic potential was observed. This could be attributed to the metal atoms with low electronegativity triggering an increase in the ionic concentration inside the endosome membrane (so-called proton sponge effect) and consequently leading to a swelling structure due to osmotic pressure, lysosome damage and cell death [20]. Another significant parameter contributing to the genotoxicity of MONPs was the ratio of the number of core electrons to the number of valence electrons. This trend was also observed by Kar, et al. [21] who reported that a higher value of the core and valence electron ratio resulted in higher levels of toxicity to E.coli.

5. CONCLUSION

The alternative non-testing methods and approaches under EU's REACH (Registration, Evaluation, Authorization and Restriction of Chemicals) regulation refer to the use of data-driven models for the purpose of NP labelling, categorization and prediction of toxicity, environmental fate and ecotoxicity. Ultimately, they are expected to support decision-making processes in the context of safety assessment by exploring data-driven estimates for occupational, consumer and environmental risks associated with the manufacture and use of NPs. In particular, non-testing methods hold great promise for maximizing the use of experimental toxicity data and contributing to the needs of the safety assessment of new materials at early design stages. While these techniques are not yet able to replace animal testing in nanotoxicology, they are capable of supporting 3R (Replacement, Reduction and Refinement) principles by minimizing the number of animals used in scientific procedures. Despite the strong industrial need and growing scientific and regulatory interest in the use of non-testing methods for prioritizing NPs according to their toxic potential and understanding structure-related root causes of their observed toxicity, the practical application and regulatory acceptance of computer-enabled material categorization and toxicity predictions have been limited so far. Given the diverse advantages that non-testing approaches offer to nanotoxicology, their use in regulatory context is expected to grow in the coming years.

Author contributions: Concept – C.O.K.; Data Collection &/or Processing – C.O.K.; Literature Search – C.O.K.; Writing – C.O.K.
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.

6. REFERENCES

[1] X. Hu, S. Cook, P. Wang, and H.-m. Hwang, "In vitro evaluation of cytotoxicity of engineered metal oxide nanoparticles," Science of the Total Environment, vol. 407, no. 8, pp. 3070-3072, 2009.

[2] N. Golbamaki et al., "Genotoxicity of metal oxide nanomaterials: review of recent data and discussion of possible mechanisms," Nanoscale, vol. 7, no. 6, pp. 2154-2198, 2015.

[3] S. Singh, "Zinc oxide nanoparticles impacts: cytotoxicity, genotoxicity, developmental toxicity, and neurotoxicity," Toxicology mechanisms and methods, vol. 29, no. 4, pp. 300-311, 2019.

[4] B. Song, Y. Zhang, J. Liu, X. Feng, T. Zhou, and L. Shao, "Is neurotoxicity of metallic nanoparticles the cascades of oxidative stress?," Nanoscale research letters, vol. 11, no. 1, pp. 1-11, 2016.

[5] W.-S. Cho et al., "Metal oxide nanoparticles induce unique inflammatory footprints in the lung: important implications for nanoparticle testing," Environmental health perspectives, vol. 118, no. 12, pp. 1699-1706, 2010.

[6] H. L. Karlsson, J. Gustafsson, P. Cronholm, and L. Möller, "Size-dependent toxicity of metal oxide particles—a comparison between nano-and micrometer size," Toxicology letters, vol. 188, no. 2, pp. 112-118, 2009.

[7] J. Kain, H. Karlsson, and L. Möller, "DNA damage induced by micro-and nanoparticles—interaction with FPG influences the detection of DNA oxidation in the comet assay," Mutagenesis, vol. 27, no. 4, pp. 491-500, 2012.

[8] H. L. Karlsson, P. Cronholm, J. Gustafsson, and L. Moller, "Copper oxide nanoparticles are highly toxic: a comparison between metal oxide nanoparticles and carbon nanotubes," Chemical research in toxicology, vol. 21, no. 9, pp. 1726-1732, 2008.

[9] M. Thwala et al., "Using the Isalos platform to develop a (Q) SAR model that predicts metal oxide toxicity utilizing facet-based electronic, image analysis-based, and periodic table derived properties as descriptors," Structural Chemistry, vol. 33, no. 2, pp. 527-538, 2022.

[10] R. C. Team. "R: A Language and Environment for Statistical Computing." (accessed.

[11] H. Wold, "Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach," Journal of Applied Probability, vol. 12, no. S1, pp. 117-142, 1975.

[12] J. Trygg and S. Wold, "Orthogonal projections to latent structures (O-PLS)," Journal of Chemometrics: A Journal of the Chemometrics Society, vol. 16, no. 3, pp. 119-128, 2002.

[13] M. Ahamed, M. A. Siddiqui, M. J. Akhtar, I. Ahmad, A. B. Pant, and H. A. Alhadlaq, "Genotoxic potential of copper oxide nanoparticles in human lung epithelial

cells," Biochemical and biophysical research communications, vol. 396, no. 2, pp. 578-583, 2010.

[14] M. J. Akhtar, S. Kumar, H. A. Alhadlaq, S. A. Alrokayan, K. M. Abu-Salah, and M. Ahamed, "Dose-dependent genotoxicity of copper oxide nanoparticles stimulated by reactive oxygen species in human lung epithelial cells," Toxicology and industrial health, vol. 32, no. 5, pp. 809-821, 2016.

[15] A. Semisch, J. Ohle, B. Witt, and A. Hartwig, "Cytotoxicity and genotoxicity of nano-and microparticulate copper oxide: role of solubility and intracellular bioavailability," Particle and fibre toxicology, vol. 11, no. 1, pp. 1-16, 2014.

[16] L. Capasso, M. Camatini, and M. Gualtieri, "Nickel oxide nanoparticles induce inflammation and genotoxic effect in lung epithelial cells," Toxicology letters, vol. 226, no. 1, pp. 28-34, 2014.

[17] M. Ahamed, D. Ali, H. A. Alhadlaq, and M. J. Akhtar, "Nickel oxide nanoparticles exert cytotoxicity via oxidative stress and induce apoptotic response in human liver cells (HepG2)," Chemosphere, vol. 93, no. 10, pp. 2514-2522, 2013.

[18] R. F. De Carli et al., "Evaluation of the genotoxic properties of nickel oxide nanoparticles in vitro and in vivo," Mutation Research/Genetic Toxicology and Environmental Mutagenesis, vol. 836, pp. 47-53, 2018.

[19] M. Weiss et al., "Density of surface charge is a more predictive factor of the toxicity of cationic carbon nanoparticles than zeta potential," Journal of nanobiotechnology, vol. 19, no. 1, pp. 1-19, 2021.

[20] Y. Huang et al., "Use of dissociation degree in lysosomes to predict metal oxide nanoparticle toxicity in immune cells: Machine learning boosts nano-safety assessment," Environment International, vol. 164, p. 107258, 2022.

[21] S. Kar, K. Pathakoti, P. B. Tchounwou, D. Leszczynska, and J. Leszczynski, "Evaluating the cytotoxicity of a large pool of metal oxide nanoparticles to Escherichia coli: Mechanistic understanding through In Vitro and In Silico studies," Chemosphere, vol. 264, p. 128428, 2021.

Research Paper
Academic Platform Journal of Engineering and Smart Systems (APJESS) 10(3), 174-183, 2022
ISSN: 2822-2385
Received: 27-Jul-2022 Accepted: 16-Sep-2022
https://doi.org/10.21541/apjess.1149455

Optimal Control of Automatic Voltage Regulator System with Coronavirus Herd Immunity Optimizer Algorithm-Based PID plus Second Order Derivative Controller

*1Selçuk EMİROĞLU, 2Talha Enes GÜMÜŞ

[1]Department of Electrical and Electronics Engineering, Sakarya University, Sakarya, Turkey, selcukemiroglu@sakarya.edu.tr
[2]Department of Electrical and Electronics Engineering, Sakarya University, Sakarya, Turkey, tgumus@sakarya.edu.tr

Abstract

Optimal control of the Automatic Voltage Regulator (AVR) system can improve the system behavior with the optimal parameters obtained based on optimization. The optimal design of proposed Proportional Integral Derivative (PID) and Proportional Integral Derivative plus Second Order Derivative ($PIDD^2$) controllers are modeled as an optimization problem including objective function and constraints. The optimization problem is solved by using the Coronavirus Herd Immunity Optimizer (CHIO) algorithm to find the best controller parameters. The optimal design of PID and $PIDD^2$ controllers for the AVR system is presented considering different objective functions. CHIO is inspired by herd immunity against COVID-19 disease by social distancing. The performances of CHIO-based controllers in the AVR system are compared with those of some modern well-known algorithms such as Atom Search Optimization (ASO), Opposition-Based Atom Search Optimization (OBASO), Particle Swarm Optimization (PSO), Improved Whale Optimization (IWO), Artificial Bee Colony (ABC), Kidney-inspired algorithm (KA), Differential Evolution (DE), Ziegler- Nichols (ZN), Local Unimodal Sampling (LUS), Biogeography-Based Optimization (BBO) and Pattern Search (PS) algorithms. Also, the obtained results demonstrate that the CHIO algorithm yields the least objective value in comparison with the other algorithms, and the superiority of the proposed approach is demonstrated.

Keywords: AVR; CHIO; Optimization; PIDD plus Second Order Derivative Controller

1. INTRODUCTION

Several voltage controllers automatically maintain the reactive power and voltage profile during power system operation and control, one of which is the automatic voltage regulator (AVR) [1].

A terminal voltage of a synchronous generator can be automatically controlled, adjusted, or maintained with the help of an AVR. The primary function of the AVR is to keep the voltage at generator terminals within a certain level or limits [2]. Therefore, the stability of the AVR system has a significant impact on the reliability of the power system [1]. So, the AVR system faces some issues with an insufficient oscillating transient response, maximum overshoot, more settling time, and steady-state errors.

The AVR system can maintain the terminal voltage at the desired level under closed-loop control using a Proportional Integral Derivate (PID) controller. Due to ease of implementation, simple structure, and robust performance, PID controllers are widely employed in process control [3]. Three parameters of PID controllers are generally adjusted with the conventional methods such as Ziegler/Nichols method, pole placement, etc. [4]. However, many intelligent algorithms are applied to adjust and obtain the parameters of the controller, such as Artificial bee colony (ABC) [3], Pattern Search (PS) [3], Particle Swarm Optimization (PSO) [5], Chaotic Ant Swarm (CAS) [5], Improved Whale Optimization (IWO) [6], Kidney-inspired algorithm (KA) [7], atom search optimization (ASO) [8], opposition-based atom search optimization (OBASO) [8], hybrid simulated annealing – Manta ray foraging optimization (SA-MRFO) [9], Ziegler- Nichols (ZN) [10], grasshopper optimization algorithm (GOA) [10], local unimodal sampling (LUS) [11], biogeography-based optimization (BBO) [12], Nonlinear Sine Cosine (NSCA) algorithms, etc. due to some disadvantages of the conventional methods. Also, tuning of fractional-order PID (FOPID) controllers has also recently gained importance. FOPID differs from traditional PID controllers in that it has fractional values as the order of the derivative and integral terms [13], [14]. The PID controller has seven independent parameters are optimally tuned based on sine-cosine algorithm (SCA) by doing time-domain, frequency and robustness analysis [13]. Also, a robust FOPID controller is designed by using Cuckoo

* Corresponding Author

Search (CS) algorithm [14]. There are also studies using fuzzy rules for traditional PID structures called Fuzzy Logic PID (FLPID). [15]. NSCA based sigmoid PID are proposed for AVR system considering objective function with steady state error and overshoot [16]. However, a Feedback Error Learning (FEL) controller, consisting of a classical controller (PD controller) and an intelligent controller (MLP neural network controller), has been proposed for the control of the AVR system for the control of AVR system with an uncertain plant model [17]. Moreover, different type of control method, such as Sliding Mode Control (SMC) is applied for AVR system [18].

In addition, several research has recently looked into a variation of the PID regulator known as the PID plus second-order derivative term generally called PIDD^2 [5], [6], [8]. This controller has one more parameter than a traditional PID, which is a second-order derivative gain (K_{dd}). In this paper, the optimal design of the PID and PIDD^2 controllers are presented by using a recently proposed optimization algorithm in the literature namely CHIO [19]. CHIO is a population-based metaheuristic algorithm inspired by herd immunity against COVID-19 [19]. CHIO algorithm has been applied for capacited vehicle routing problem [20]. The feature selection problem in the medical diagnosis domain has been also solved with the implementation of CHIO [21]. Moreover, CHIO algorithm has been successfully used in many areas such as high renewable penetration microgrid [22] and economic dispatch problems [23] etc.

The objective of the work is the design and implementation of an efficient CHIO-based PID and PIDD^2 controllers for the AVR system. In the study, the design of controllers formulated as an optimization problem is achieved by using CHIO algorithm to obtain the optimal controller parameters. The choice of an appropriate objective function is critical for the controller design to get a better and desired response. So, six different objective functions widely employed in the literature are used and compared in terms of the system performance with each other considering settling time, rise time, maximum overshoot, and peak time.

2. MATERIALS AND METHODS

The optimal parameters of the AVR system are obtained by using CHIO.

2.1. Coronavirus Herd Immunity Optimizer (CHIO) Algorithm

CHIO, an efficient search algorithm for global optimization problems, is a population-based metaheuristic algorithm inspired by herd immunity against COVID-19 disease by social distancing [19].

The spread of viruses among humans is very rapid. The method of fighting viruses is to create immunity against the virus. The vaccine is used to create immunity, but a certain time is needed for vaccine production. During this period, there are two ways to fight viruses.

• Infected people are isolated from society.

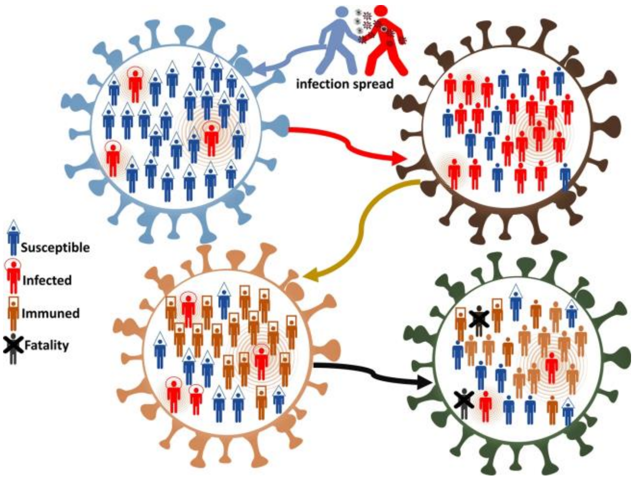


Figure 1. Herd immunity scenario [19]

• When the majority of the population is infected, herd immunity develops against the disease, and other individuals are indirectly protected from the disease. (If a large part of the population is infected, a collective immunity is formed against the disease, and other individuals are indirectly protected from the disease as shown in Fig. 1.)

The development of herd immunity has been carried out with the COVID-19 optimization algorithm. The algorithm consists of a total of 6 steps [19]. Also, the flowchart of the algorithm is depicted in Fig. 2.

Step 1: In the first step, the objective function is created and the optimization problem is defined. CHIO includes four algorithm variables and two control parameters. The number of infected individuals (initially one), the maximum number of iterations, the population size, and the size of the problem are the parameters of the algorithm. The control variables are the rate of spread of the virus and the maximum age of the infected [19].

Step 2: Creating herd immunity population; the number and size of the population and herd immunity control variables are determined.

Step 3: Disease spread; some of the infected people here die according to the coronavirus mortality rate. These individuals cannot infect new individuals. Individuals who recover after becoming coronavirus cause the virus to spread in two ways. The first will infect according to the rate of virus spread. The latter will infect the virus according to its super spread rate [19].

Step 4: Population update: 3 populations are updated for each generation.

- Dead population: if the infected individual dies, it is added to this population and is not used again.
- Healing population: In each iteration, the infected individuals are sent to the recovered population.
- Newly infected population: In each iteration, all infected individuals are added to this population.

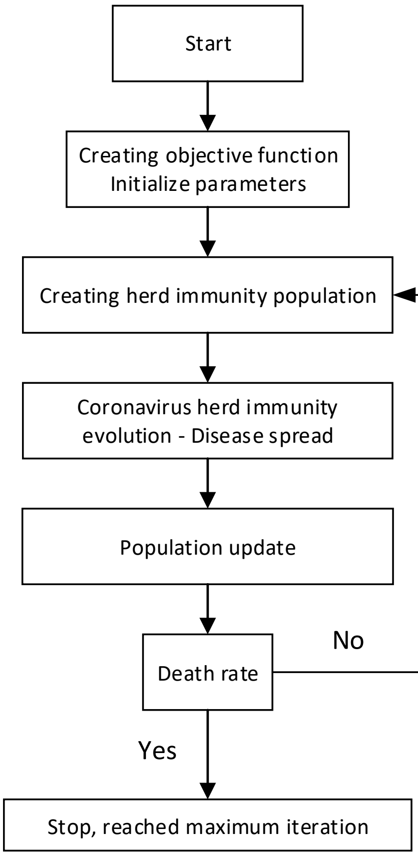


Figure 2. Flowchart of CHIO

Step 5: Mortality: If the immune rate of the infected case is not recovered for the number of iterations determined by the Maximum age parameter, this case is considered dead. This increases the existing population diversity, allowing the local outcome to move away from local minimums [19].

Step 6: Stopping criterion: The most important feature of the COVID-19 optimization algorithm is that it terminates without the need to check any parameters. This is because the recovering and dead population is constantly increasing, and newly infected individuals cannot transmit the disease to other individuals. The algorithm terminates when it reaches the maximum number of refreshes [19].

2.2. Automatic Voltage Regulator System

In power systems, the system voltage needs to be kept within the defined limit and constant. Otherwise, the system suffers stability problems, and the loads are fed by lower voltages. So, voltage drops should be prevented in power systems. AVR is designed to control and keep the terminal voltage of the generator constant. AVR controls the terminal voltage of the generator with the field current. It maintains the output voltage and keeps the voltage at the desired value. The AVR system brings the terminal voltage to the desired value under closed-loop control. AVR system contains some parts, these are an amplifier, exciter, generator, and sensor [4], [6].

a) Amplifier

The amplifier's transfer function has a gain and a time constant.

$$TF_A = \frac{K_a}{1+sT_a} \tag{1}$$

Where T_a and K_a are the time constant for the amplifier and gain, respectively. Time constant T_a is between 0.02 s and 0.1 s. The values of K_a are generally between 10 and 40 [4], [6].

b) Exciter

The transfer function for the Exciter is given as the below function. It has a time constant T_e and gain K_e. K_e and T_e values are generally between 1 and 10, and between 0.4 s and 1 s, respectively [4], [6].

$$TF_E = \frac{K_e}{1+sT_e} \tag{2}$$

c) Generator

The generator's transfer function is given below. The load affects the generator's gain and time constant. According to the load, the time constant T_g and the gain K_g are changing from 1 s to 2 s and from 0.7 to 1, respectively [4], [6].

$$TF_G = \frac{K_g}{1+sT_g} \tag{3}$$

d) Sensor

The sensor transfer function is also given as in Eq 4. T_s is between 0.001 s and 0.06 s and K_s is about 1 [4], [6].

$$TF_S = \frac{K_S}{1+sT_S} \tag{4}$$

Table 1. The parameters of the AVR system [24]

	Parameter Limits for *K* and *T*	**Used Parameter Values**
Amplifier	$10 \le K_a \le 40$ $0.02 \le T_a \le 0.1$	K_a=10, T_a=0.1 s
Exciter	$1 \le K_e \le 10$ $0.4 \le T_e \le 1$	K_e=1, T_e=0.4 s
Generator	$0.7 \le K_g \le 1$ $1 \le T_g \le 2$	K_g=1, T_g=1 s
Sensor	$1 \le K_s \le 2$ $0.001 \le T_s \le 0.06$	K_s=1, T_s=0.01 s

The parameters and parameter limits of the AVR system are taken as in [24] and given in Table 1.

When there is no controller, the transfer function of the AVR system as shown in Fig. 3 is expressed as follows.

$$\frac{\Delta V_t(s)}{\Delta V_{ref}(s)} = \frac{0.1s+10}{0.0004s^4+0.0454s^3+0.555s^2+1.51s+11} \tag{5}$$

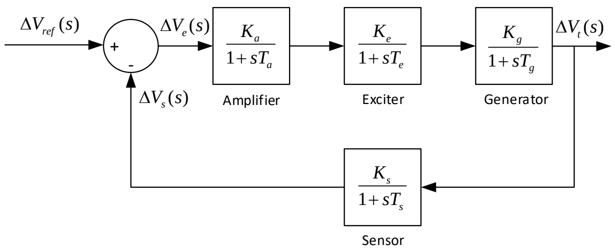


Figure 3. Block diagram of the AVR system

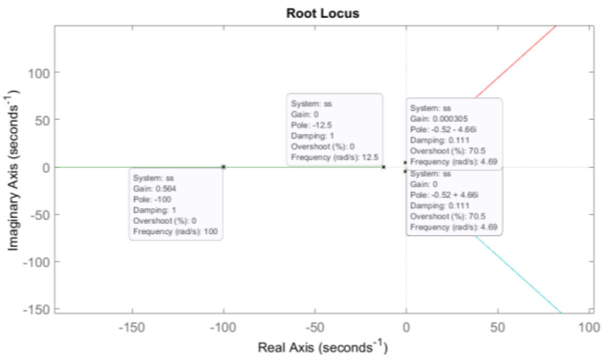


Figure 4. Root locus of the AVR system without PID controller

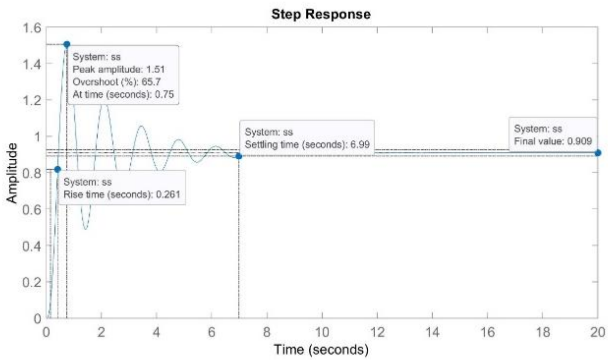


Figure 5. Step response of the AVR system

As can be seen from the root locus in Fig. 4, the open loop poles of the AVR system are s_1=-99.97, s_2=-12.4892, $s_{3,4}$=-0.5198±4.6642i, respectively, and their respective damping ratios are 1.000, 1.000, 0.111 and 0.111 respectively. As shown in Fig. 5, the steady-state value, overshoot, rise time, and settling time are 0.909, 65.7%, 0.261 s, and 6.99 s (2% bant), respectively. The steady-state error of the system is obtained as 0.091 pu. So, the system needs to have a controller to eliminate the steady state error and improve the transient response of the system.

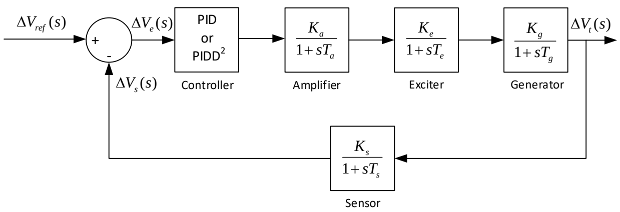


Figure 6. Block diagram of the AVR system with PID or PIDD^2 controller

The PID and PIDD^2 controllers are designed to improve the response of the AVR system. The AVR system block diagram with PID and PIDD^2 controllers is depicted in Fig. 6.

The transfer function for the AVR system with PID and PIDD^2 controllers depicted in Fig. 6 is defined in Eq. 6 and 7, respectively.

$$\frac{\Delta V_t(s)}{\Delta V_{ref}(s)}=\frac{0.1K_d s^2+(0.1K_p+10K_d)s^2+(0.1K_i+10K_p)s+10K_i}{0.0004s^5+0.045s^4+0.555s^3+(1.51+10K_d)s^2+(1+10K_p)s+10K_i} \quad (6)$$

$$\frac{\Delta V_t(s)}{\Delta V_{ref}(s)}=\frac{0.1K_{dd} s^4+(0.1K_d+10K_{dd})s^3+(10K_d+0.1K_p)s^2+(0.1K_i+10K_p)s+10K_i}{0.0004s^5+0.0454s^4+(10K_{dd}+0.555)s^3+(10K_d+1.51)s^2+(10K_p+1)s+10K_i} \quad (7)$$

2.3. Determination of PID and PIDD^2 parameters, and objective functions

Three optimal control parameters are necessary for the design of the PID controller. The objective function is minimized to estimate these optimum parameters. Several objective functions are considered such as integral of time multiplied absolute error (ITAE), integral of time multiplied square error (ITSE), integral square error (ISE), integral absolute error (IAE), integral of square time multiplied by square error (ISTSE), integral square time multiplied square error (ISTES). The aforementioned objective functions are given in the following equations.

- ITAE

$$J=\int_0^T t\left|e(t)\right|dt \quad (8)$$

- ITSE

$$J=\int_0^T t\left|e^2(t)\right|dt \quad (9)$$

- ISE

$$J=\int_0^T \left|e^2(t)\right|dt \quad (10)$$

- IAE

$$J=\int_0^T \left|e(t)\right|dt \quad (11)$$

- ISTSE

$$J=\int_0^T t^2 e^2(t)dt \quad (12)$$

- ISTES

$$J=\int_0^T [t^2 e(t)]^2 dt \quad (13)$$

Where $e=V_r - V_t$, V_t and V_r are the terminal and reference voltage, respectively.

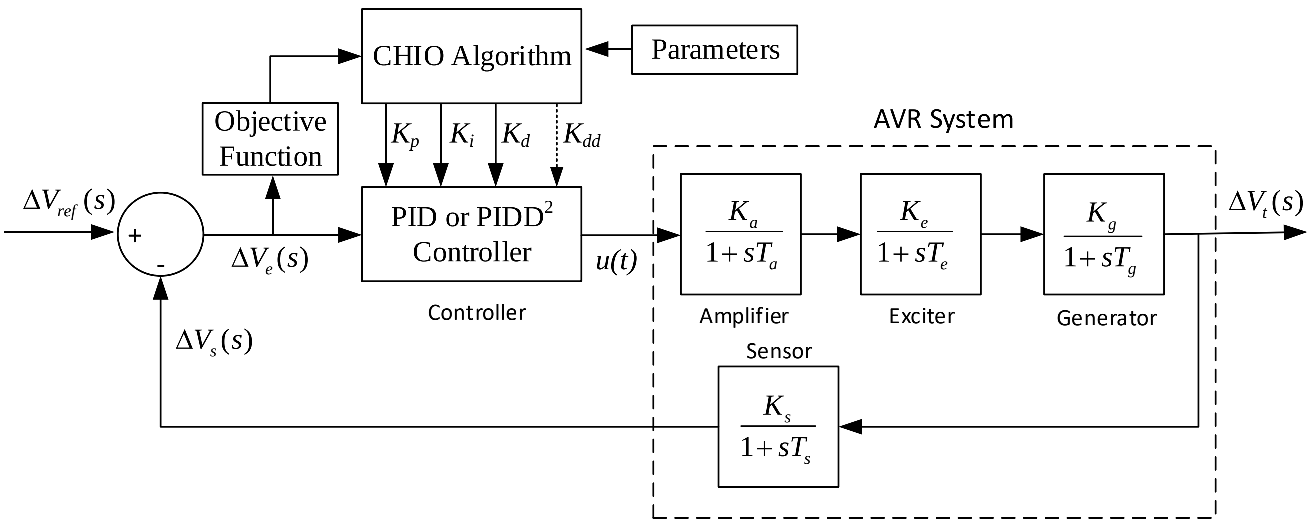


Figure 7. PID or PIDD^2 controlled AVR system with CHIO algorithm

To find out the optimal parameters of the PID or PIDD^2 controllers, the formulation of the optimization problem can be given below.

$$\begin{aligned} &\min J \\ &st.\,x = f(t,u) \\ &K_{min} < K < K_{max} \end{aligned} \tag{14}$$

where J and x are the objective function of optimization problem and system model, respectively. K_{min}-K_{max} are the limits of the controller parameters used as an inequality constraint in optimization problem. K are the controller parameters K_p, K_i, K_d, and K_{dd} in short form.

The CHIO algorithm is used to solve the optimization problem given above and find the optimal controller parameters. Fig. 7 shows the block diagram of the AVR system with the proposed CHIO algorithm tuned PID or PIDD^2 controller.

3. FINDINGS

In the paper, CHIO technique is utilized to obtain the optimal PID controller parameters K_p, K_i, K_d and PIDD^2 controller parameters K_p, K_i, K_d, K_{dd} as shown in Fig. 7. To make a comparison in terms of transient response, different objective functions are used. The optimal parameters of the controllers determined by minimizing several objective functions are given in Table 2.

The implementation of the CHIO algorithm, its adaptation to the AVR system and all analysis have been carried out using MATLAB [25]. Steady-state error, overshoot, settling time, rise time and peak time showing the response and behavior of the AVR system with PID controller designed according to the different objective functions are given in Table 2.

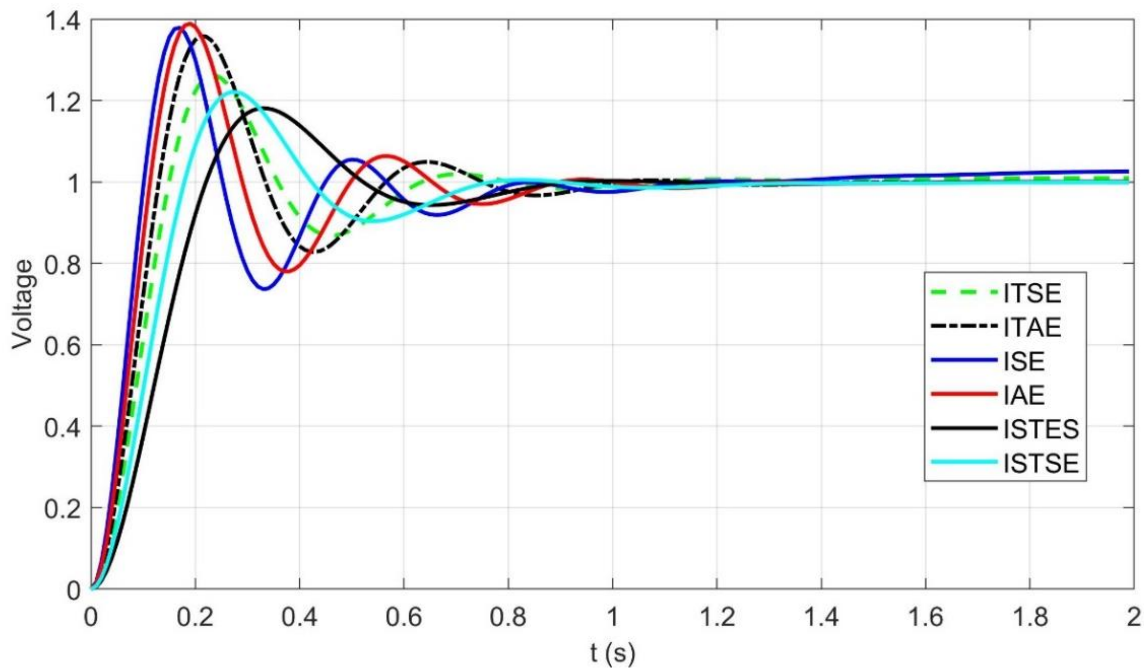


Figure 8. Terminal voltage of the AVR system with CHIO-PID controller for different objective functions.

Table 2. The optimal parameters of CHIO-based PID controllers according to the different objectives

Objective Function \ Parameters	K_p	K_i	K_d	Objective Value	M_p (pu)	E_{ss} (pu) (1.8 s)	t_s (s) (2% bant)	t_r (s)	t_p (s)
ITSE	1.5627	1.4835	0.7864	0.00558	1.2617	0.0093	**0.6055**	0.1012	0.23
ITAE	2.3384	1.5386	0.9246	0.0386	1.358	0.0012	0.9307	0.0875	0.21
ISE	1.1141	2.3534	1.4503	0.0652	1.3787	0.022	3.0251	**0.0683**	**0.17**
IAE	2.2532	1.6406	1.1687	0.1619	1.3888	**0.0004**	0.8464	0.0761	0.19
ISTES	1.3269	0.8857	0.4418	0.00024	**1.1811**	0.0013	0.8184	0.1494	0.33
ISTSE	1.5106	1.0075	0.5982	0.00122	1.2208	0.0013	0.7090	0.1210	0.27

Table 3. The optimal parameters of PID controller and transient response results obtained by ITSE with the different approaches.

Controllers \ Parameters	K_p	K_i	K_d	M_p (pu)	E_{ss} (pu)	t_s (s) (2% bant)	t_r (s)	t_p (s)	ITSE (t_{sim}=20 s)
CHIO-PID (Proposed)	1.5627	1.4835	0.7864	1.2617	0.0093	**0.6065**	**0.101**	**0.23**	**0.00558**
KA-PID [7]	1.0685	1.0018	0.5103	**1.136**	0.0129	0.771	0.143	0.31	0.0061
ZN-PID [10]	1.021	1.8743	0.139	1.515	0.00553	3.052	0.237	0.644	0.107
ABC-PID [24]	1.6524	0.4083	0.3654	1.25	0.02684	3.094	0.156	0.36	0.018
LUS-PID [11]	1.2012	0.9096	0.4593	1.156	0.00358	0.800	0.149	0.322	0.0064
PS-PID [3]	1.2771	0.8471	0.4775	1.169	**0.00174**	0.804	0.144	0.316	0.0064
BBO-PID [12]	1.2464	0.5893	0.4596	1.16	0.01564	1.446	0.149	0.317	0.0077

Table 4. The results for CHIO-based PIDD^2 controller with ITAE objective function under different parameter limits.

Constraint \ Parameters	K_p	K_i	K_d	K_{dd}	Objective Value	M_p (pu)	E_{ss} (pu) (1.8 s)	t_s (s) (2% bant)	t_r (s)	t_p (s)
0.001 < *K* < 3	2.9945	1.9947	1.0797	0.079754	0.0015744	1	0	0.1468	0.0839	1.05
0.001 < *K* < 5	4.9796	3.2962	1.8223	0.1492	0.0008023	1.0002	0	0.0561	0.036	0.73
0.001 < *K* < 10	9.9334	6.6262	3.5827	0.26644	0.00026937	1.144	0	0.0620	0.0154	0.035

Table 5. The results for CHIO-based PIDD^2 controller with ITSE objective function under different parameter limits.

Constraint \ Parameters	K_p	K_i	K_d	K_{dd}	Objective Value	M_p (pu)	E_{ss} (pu) (1.8 s)	t_s (s) (2% bant)	t_r (s)	t_p (s)
0.001 < *K* < 3	3	2.1396	1.3947	0.12497	0.00027538	1.0021	0.002	0.3777	0.0529	2
0.001 < *K* < 5	5	3.2073	2.7619	0.26957	0.00011654	1.0983	0.0003	0.2851	0.0159	0.033
0.001 < *K* < 10	8.804	5.6957	3.0964	0.28613	3.5221e-05	1.1298	0.000177	0.0997	0.0146	0.032

Table 6. The results for CHIO-based PIDD^2 controller with ISE objective function under different parameter limits.

Constraint \ Parameters	K_p	K_i	K_d	K_{dd}	Objective Value	M_p (pu)	E_{ss} (pu) (1.8 s)	t_s (s) (2% bant)	t_r (s)	t_p (s)
0.001 < *K* < 3	3	3	3	0.42594	0.0054334	1.2431	1.7702e-04	0.8645	0.0092	0.022
0.001 < *K* < 5	5	5	3.3196	0.4366	0.004787	1.2627	0.0053	0.3832	0.0089	0.022
0.001 < *K* < 10	10	5.301	2.93705	0.444665	0.0044699	1.2604	0.0015	0.1939	0.0087	0.022

Table 7. The results for CHIO-based PIDD^2 controller with IAE objective function under different parameter limits.

Constraint \ Parameters	K_p	K_i	K_d	K_{dd}	Objective Value	M_p (pu)	E_{ss} (pu) (1.8 s)	t_s (s) (2% bant)	t_r (s)	t_p (s)
0.001 < *K* < 3	3	2.0037	1.0903	0.079727	0.0401	1.0001	0	0.1431	0.083	1.91
0.001 < *K* < 5	5	3.31	1.8359	0.14542	0.0202	1	0	0.0562	0.0367	2
0.001 < *K* < 10	10	6.659	3.60947	0.27295	0.0128	1.1485	0	0.0604	0015	0.034

Table 8. The results for CHIO-based PIDD^2 controller with ISTES objective function under different parameter limits.

Parameters / Constraint	K_p	K_i	K_d	K_{dd}	Objective Value	M_p (pu)	E_{ss} (pu) (1.8 s)	t_s (s) (2% bant)	t_r (s)	t_p (s)
0.001 < *K* < 3	2.1485	1.3948	0.8376	0.078934	4.83e-06	1.0042	2.95e-05	0.3939	0.132	2
0.001 < *K* < 5	5	3.3473	1.7741	0.12287	1.47e-08	1.0178	7.73e-06	0.0629	0.0432	0.101
0.001 < *K* < 10	7.2514	4.7424	2.7353	0.24268	4.77e-08	1.0821	1.27e-05	0.1091	0.0181	0.037

Table 9. The results for CHIO-based PIDD^2 controller with ISTSE objective function under different parameter limits.

Parameters / Constraint	K_p	K_i	K_d	K_{dd}	Objective Value	M_p (pu)	E_{ss} (pu) (1.8 s)	t_s (s) (2% bant)	t_r (s)	t_p (s)
0.001 < *K* < 3	3	1.995	1.1051	0.084801	1.4682e-05	1.0001	6.67e-05	0.1583	0.0813	1.93
0.001 < *K* < 5	5	3.3085	1.8376	0.14435	1.8623e-06	1	5.12e-06	0.0563	0.037	2.87
0.001 < *K* < 10	7.6209	5.0543	2.7458	0.20872	6.9491e-07	1.007	2.61e-05	0.0706	0.0214	0.045

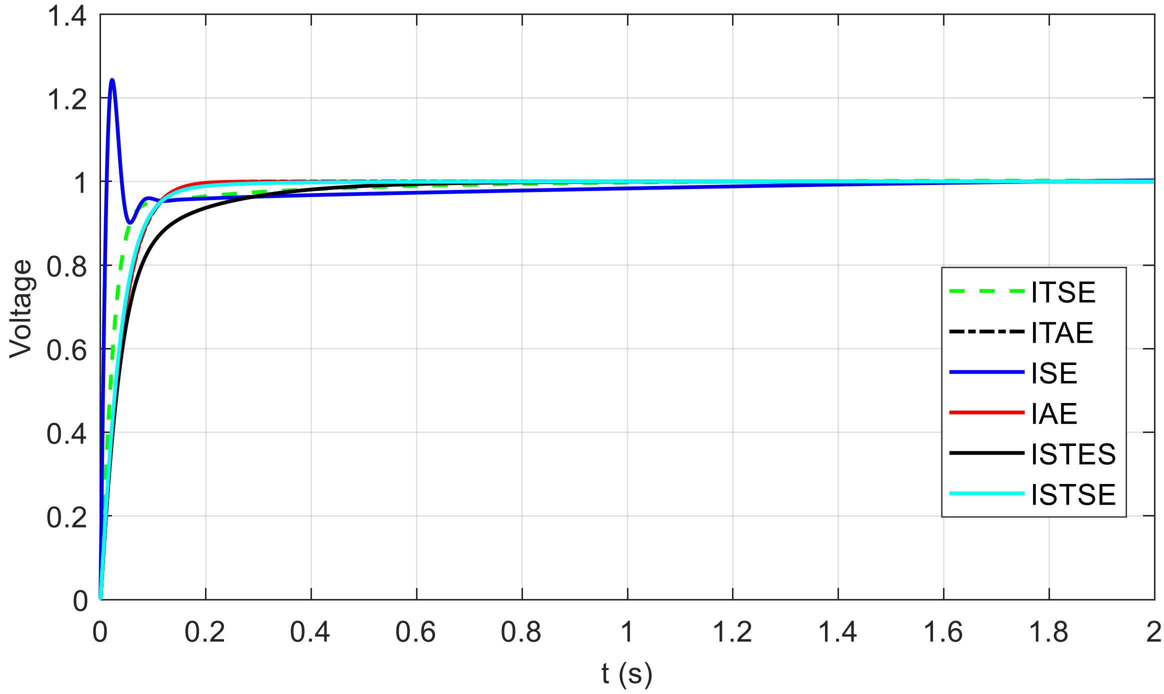


Figure 9. Terminal Voltage of AVR system with CHIO - PIDD^2 controller (0.001<*K*<3)

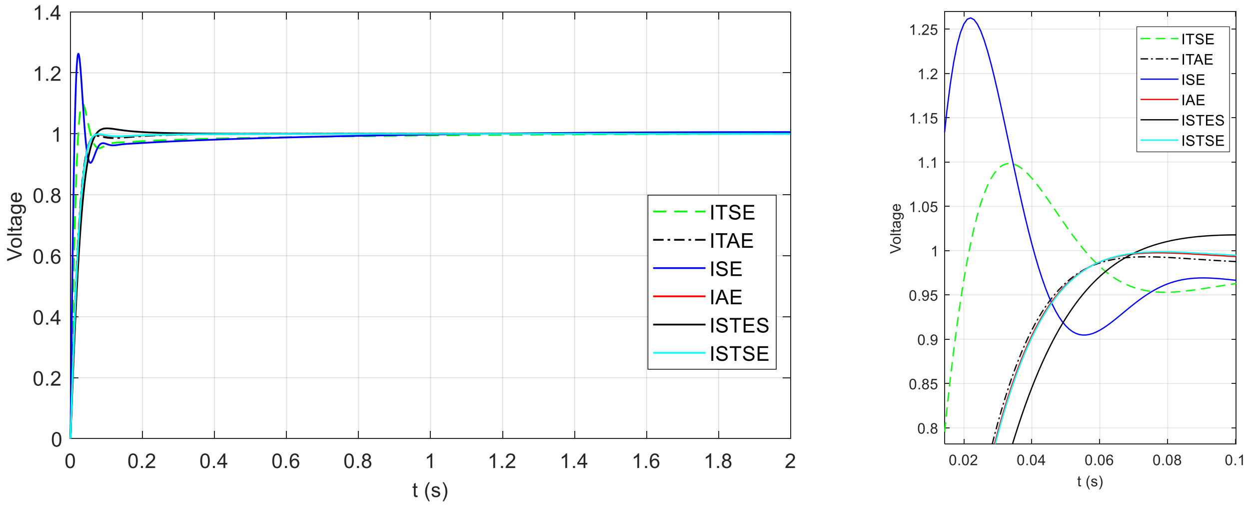


Figure 10. Terminal Voltage of AVR system with CHIO - PIDD^2 controller (0.001<*K*<5)

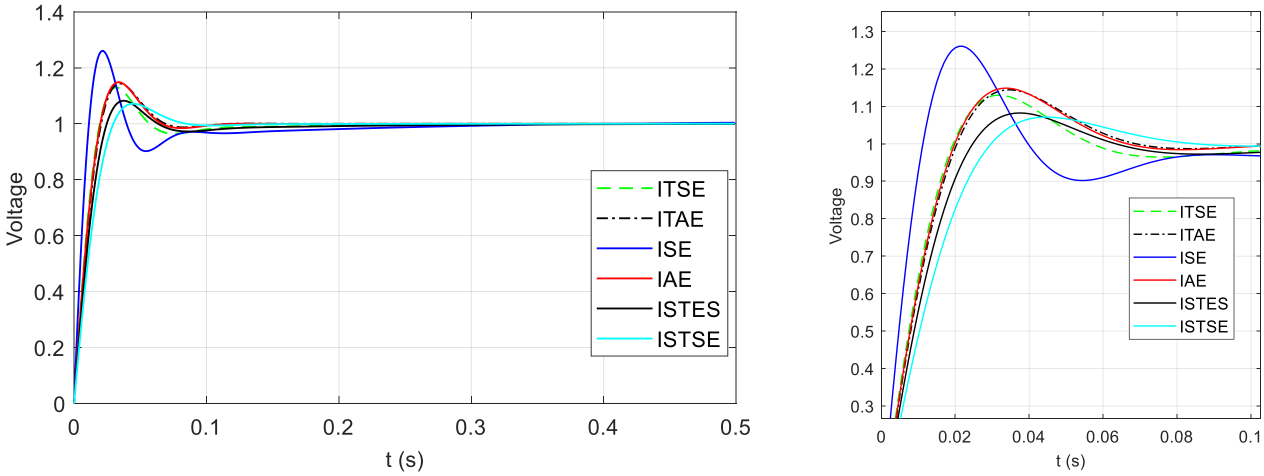


Figure 11. Terminal Voltage of AVR system with CHIO - PIDD^2 controller (0.001<K<10)

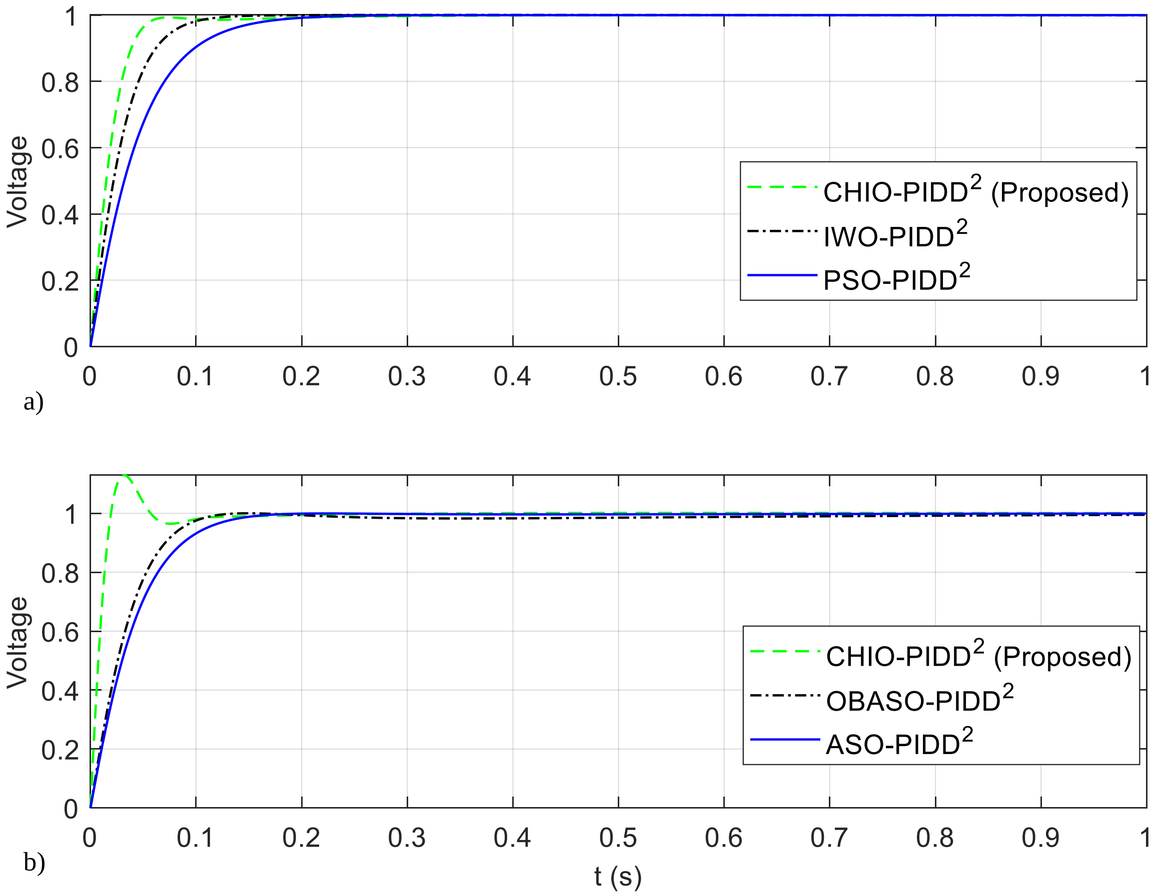


Figure 12. Comparative results of AVR system with ITAE (a) and ITSE (b) objective function

In Tables, the time response characteristic parameters M_p, E_{ss}, t_s, t_r and t_p are the percentage overshoot, steady state error, settling time, rise time and peak time, respectively. The settling time (t_s) is defined as the time within a band of ±2% around the final value of the step response.

To demonstrate the effect of the objective function on the design of controller and controller performance, some objective functions are utilized in the optimization problem.

The optimal parameters obtained with different objective functions are given in Table 2. Objective value, steady state error, maximum overshoot, rise time, settling time, and the peak time for parameters found with each objective function are also given in Table 2. The system responses with these parameters are shown in Fig. 8. Fig. 8 shows the comparative simulation results for the response of the AVR system designed with different objective functions.

Table 10. The comparative results for PIDD² controller based on different approaches

Controllers \ Parameters	K_p	K_i	K_d	K_{dd}	Objective Value	M_p (pu)	E_{ss} (pu) (1.8 s)	t_s (s) (2% bant)	t_r (s)	t_p (s)
CHIO-PIDD² (Proposed) - ITAE	4.9796	3.2962	1.8223	0.1492	**0.0008**	1.0002	**0**	**0.0561**	**0.036**	0.73
IWO - PIDD² [6]	3.9348	2.5753	1.3985	0.10453	0.0019	1.0007	2.47e-04	0.0981	0.0584	**0.3392**
PSO - PIDD² - ITAE [5]	2.7784	1.8521	0.9997	0.07394	0.0018	**1**	4.29e-06	0.1635	0.0930	0.4295
CHIO - PIDD² (Proposed) - ITSE	8.804	5.6957	3.0964	0.28613	**3.52e-05**	1.1298	1.77e-04	**0.0997**	**0.0146**	**0.032**
OBASO - PIDD² - ITSE [8]	2.9209	1.9463	1.3359	0.08813	3.56e-04	1.0009	2.75e-04	0.1036	0.0662	2.48
ASO - PIDD² - ITSE [8]	2.9310	1.9571	1.1033	0.07771	4.16e-04	**1.0002**	**1.74e-04**	0.1363	0.0825	2.11

In terms of settling time, ITSE objective function gives the shortest settling time as 0.6055 s in 2% bant. In terms of rise time, ISE objective function gives the shortest rise time as 0.0683 s among all objective functions. Also, ISTES and IAE objective functions give the smallest maximum overshoot as 1.1811 pu and steady-state error as 0.0004 pu (at 1.8 s), respectively.

Comparisons are made with other approaches in the literature using ITSE objective function to demonstrate the effectiveness and advantage of the proposed CHIO-PID controller. For comparison, the values which show the system behavior with proposed and other approaches are given in Table 3. As can be seen in Table 3, the proposed CHIO based PID controller gives better results among them. It seems that rise time, settling time, and the peak time is the smallest among the approaches given in Table 3. Especially, ITSE objective value of 0.00558 is the smallest, so the proposed approach provides the minimum objective value.

Tables 4-9 show the results for CHIO-based PIDD² controller with different objective functions ITAE, ITSE, IAE, ISE, ISTES and ISTSE under different parameter limits, respectively. It seems that the three different parameter limits are employed for the optimization as an inequality constraint. One can deduce from Tables 4 - 9 that the objective values decrease when the maximum parameters' values are increased for all objective functions. Also, steady state error, settling time, rise time, maximum overshoot, and the peak time of the AVR system with PIDD² controller with optimal parameters obtained by using the CHIO algorithm are given in Tables 4 - 9.

Figs. 9- 11 show the transient response of the AVR system with CHIO-based PIDD² controller using different objective functions and different parameter limits. Three different parameter limits are used to find the optimal parameters in optimization as a constraint like $0.001<K<3$, $0.001<K<5$, $0.001<K<10$ for the parameters K_p, K_i, K_d, K_{dd}. Fig. 10 shows the comparative results for different PIDD² controllers of the AVR system with ITAE and ITSE functions using the parameters given in Table 10. It seems that the system with the proposed approach has a smaller objective value, rise and settling time than the other methods seen in Table 10 and Fig. 12. In addition, the system with the proposed approach used in ITSE objective function has a little bit larger maximum overshoot. However, the objective function value, settling time, and rise time are the smallest among them.

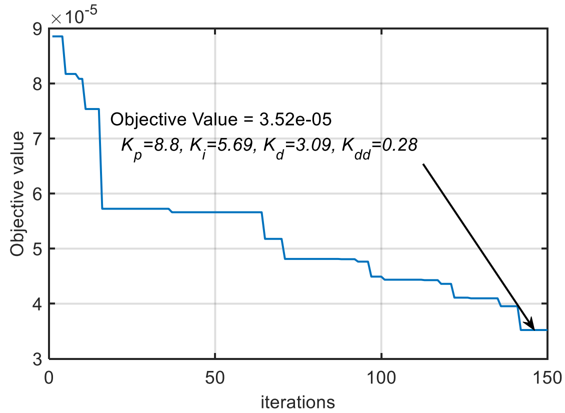


Figure 13. Convergence characteristic for CHIO-PIDD² (Proposed approach) using ITSE.

It is compared with other approaches given in Table 10 to demonstrate the effectiveness and advantage of the proposed CHIO-PIDD² controller, which improves the AVR system's transient response.

Also, objective value versus iteration graph is given in Fig. 13. The convergence behavior of CHIO algorithm for PIDD² controller based on ITSE as in Table 10 can be seen in Fig. 13.

In simulations, the parameters of CHIO are taken as 100 for the population size, 100 for the Max age, with a spreading rate equal to 0.1.

4. DISCUSSION AND CONCLUSION

In this paper, a recently proposed CHIO algorithm is used to obtain the optimal parameters of PID and PID plus second-order derivative (PIDD²) controllers. Different objectives are investigated to demonstrate the effect of objective functions on the design of PID and PIDD² controllers. Furthermore, the proposed CHIO-based PID and PIDD² controllers and other optimization algorithm-based PID and PIDD² controllers such as ASO, OBASO, PSO, IWO, ABC, KA, DE, ZN, LUS, BBO and PS etc. have been compared and shown in figures and tables.

The response of the AVR system is improved with the proposed approach in terms of rise time, settling time, and maximum overshoot compared to approaches cited in the paper. Simulation results showed that proposed PID and PIDD² controllers provide superior response performance.

Author contributions: Concept – S.E, T.E.G; Data Collection &/or Processing – S.E, T.E.G; Literature Search – S.E, T.E.G; Writing – S.E, T.E.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] P. Kundur, Power System Stability and Control, vol. 20073061. McGraw-Hill, 1994.

[2] H. Saadat, Power System Analysis. PSA Publishing LLC, 2011.

[3] B. K. Sahu, P. K. Mohanty, and N. Mishra, “Robust Analysis and Design of PID controlled AVR system using Pattern Search algorithm,” in IEEE International Conference on Power Electronics, Device and Energy System 2012, 2012.

[4] S. Panda, B. K. Sahu, and P. K. Mohanty, “Design and performance analysis of PID controller for an automatic voltage regulator system using simplified particle swarm optimization,” J. Franklin Inst., vol. 349, no. 8, pp. 2609–2625, 2012.

[5] M. A. Sahib, “A novel optimal PID plus second order derivative controller for AVR system,” Eng. Sci. Technol. an Int. J., vol. 18, no. 2, pp. 194–206, 2015.

[6] D. Mokeddem and S. Mirjalili, “Improved Whale Optimization Algorithm applied to design PID plus second-order derivative controller for automatic voltage regulator system,” J. Chinese Inst. Eng. Trans. Chinese Inst. Eng. A, vol. 43, no. 6, pp. 541–552, 2020.

[7] S. Ekinci, A. Demirören, H. L. Zeynelgil, and S. Kaya, “Böbrek-ilhamlı Algoritma ile Otomatik Gerilim Regülatör Sistemi için PID Kontrolör Tasarımı,” Gazi Üniversitesi Fen Bilim. Derg. Part C Tasarım ve Teknol., vol. 7, no. 2, pp. 383–398, 2019.

[8] S. Ekinci, A. Demirören, H. L. Zeynelgil, and B. Hekimoğlu, “An opposition-based atom search optimization algorithm for automatic voltage regulator system,” J. Fac. Eng. Archit. Gazi Univ., vol. 35, no. 3, pp. 1141–1157, 2020.

[9] M. Micev, M. Ćalasan, Z. M. Ali, H. M. Hasanien, and S. H. E. Abdel Aleem, “Optimal design of automatic voltage regulation controller using hybrid simulated annealing – Manta ray foraging optimization algorithm,” Ain Shams Eng. J., vol. 12, no. 1, pp. 641–657, 2021.

[10] B. Hekimoğlu and S. Ekinci, “Grasshopper optimization algorithm for automatic voltage regulator system,” 2018 5th Int. Conf. Electr. Electron. Eng. ICEEE 2018, pp. 152–156, 2018.

[11] P. K. Mohanty, B. K. Sahu, and S. Panda, “Tuning and assessment of proportional-integral-derivative controller for an automatic voltage regulator system employing local unimodal sampling algorithm,” Electr. Power Components Syst., vol. 42, no. 9, pp. 959–969, 2014.

[12] U. Güvenç, T. Yiğit, A. H. Işik, and I. Akkaya, “Performance analysis of biogeography-based optimization for automatic voltage regulator system,” Turkish J. Electr. Eng. Comput. Sci., vol. 24, no. 3, pp. 1150–1162, 2016.

[13] M. S. Ayas and E. Sahin, “FOPID controller with fractional filter for an automatic voltage regulator,” Comput. Electr. Eng., vol. 90, no. April 2020, p. 106895, 2021.

[14] A. Sikander, P. Thakur, R. C. Bansal, and S. Rajasekar, “A novel technique to design cuckoo search based FOPID controller for AVR in power systems,” Comput. Electr. Eng., vol. 70, pp. 261–274, 2018.

[15] A. J. H. Al Gizi, “A particle swarm optimization, fuzzy PID controller with generator automatic voltage regulator,” Soft Comput., vol. 23, no. 18, pp. 8839–8853, 2019.

[16] M. H. Suid and M. A. Ahmad, “Optimal tuning of sigmoid PID controller using Nonlinear Sine Cosine Algorithm for the Automatic Voltage Regulator system,” ISA Trans., vol. 128, pp. 265–286, 2022.

[17] M. A. Labbaf Khaniki, M. Behzad Hadi, and M. Manthouri, “Feedback Error Learning Controller based on RMSprop and Salp Swarm Algorithm for Automatic Voltage Regulator System,” 2020 10h Int. Conf. Comput. Knowl. Eng. ICCKE 2020, pp. 425–430, 2020.

[18] V. Sharma, V. Kumar, R. Naresh, and V. Kumar, “Automatic voltage regulator system with state-feedback and PID based sliding mode control design,” Proc. 2021 1st Int. Conf. Adv. Electr. Comput. Commun. Sustain. Technol. ICAECT 2021, 2021.

[19] M. A. Al-Betar, Z. A. A. Alyasseri, M. A. Awadallah, and I. Abu Doush, “Coronavirus herd immunity optimizer (CHIO),” Neural Comput. Appl., vol. 33, no. 10, pp. 5011–5042, 2021.

[20] L. M. Dalbah, M. A. Al-Betar, M. A. Awadallah, and R. A. Zitar, “A modified coronavirus herd immunity optimizer for capacitated vehicle routing problem,” J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 8, pp. 4782–4795, 2022.

[21] M. Alweshah, S. Alkhalaileh, and M. A. Al-betar, “Coronavirus herd immunity optimizer with greedy crossover for feature selection in medical diagnosis,” Knowledge-Based Syst., vol. 235, p. 107629, 2022.

[22] M. Alqarni, “Sodium sulfur batteries allocation in high renewable penetration microgrids using coronavirus herd immunity optimization,” Ain Shams Eng. J., vol. 13, no. 2, p. 101590, 2022.

[23] S. Amini, S. Ghasemi, H. Golpira, and A. Anvari-moghaddam, “Coronavirus Herd Immunity Optimizer (CHIO) for Transmission Expansion Planning,” in 2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), 2021, no. 18.

[24] H. Gozde and M. C. Taplamacioglu, “Comparative performance analysis of artificial bee colony algorithm for automatic voltage regulator (AVR) system,” J. Franklin Inst., vol. 348, no. 8, pp. 1927–1946, 2011.

[25] “MATLAB.” The Mathworks, Inc., Natick, Massachusetts, United States.