ISSN 2587-1366

TUJE

TURKISH JOURNAL OF ENGINEERING

VOLUME 7 ISSUE 1 JANUARY 2023

Dr. Ghazi DROUBI
Robert Gordon University Engineering Faculty
Scotland, UK

JOURNAL SECRETARY

Aydin Alptekin
aydinalptekin@mersin.edu.tr

TURKISH JOURNAL OF ENGINEERING (TUJE)

Turkish Journal of Engineering (TUJE) is a multi-disciplinary journal. The Turkish Journal of Engineering (TUJE) publishes the articles in English and is being published 4 times (January, April, July and October) a year. The Journal is a multidisciplinary journal and covers all fields of basic science and engineering. It is the main purpose of the Journal that to convey the latest development on the science and technology towards the related scientists and to the readers. The Journal is also involved in both experimental and theoretical studies on the subject area of basic science and engineering. Submission of an article implies that the work described has not been published previously and it is not under consideration for publication elsewhere. The copyright release form must be signed by the corresponding author on behalf of all authors. All the responsibilities for the article belongs to the authors. The publications of papers are selected through double peer reviewed to ensure originality, relevance and readability.

AIM AND SCOPE

The Journal publishes both experimental and theoretical studies which are reviewed by at least two scientists and researchers for the subject area of basic science and engineering in the fields listed below:

Aerospace Engineering
Environmental Engineering
Civil Engineering
Geomatic Engineering
Mechanical Engineering
Geology Science and Engineering
Mining Engineering
Chemical Engineering
Metallurgical and Materials Engineering
Electrical and Electronics Engineering
Mathematical Applications in Engineering
Computer Engineering
Food Engineering

PEER REVIEW PROCESS

All submissions will be scanned by iThenticate® to prevent plagiarism. Author(s) of the present study and the article about the ethical responsibilities that fit PUBLICATION ETHICS agree. Each author is responsible for the content of the article. Articles submitted for publication are priorly controlled via iThenticate ® (Professional Plagiarism Prevention) program. If articles that are controlled by iThenticate® program identified as plagiarism or self-plagiarism with more than 25% manuscript will return to the author for appropriate citation and correction. All submitted manuscripts are read by the editorial staff. To save time for authors and peer-reviewers, only those papers that seem most likely to meet our editorial criteria are sent for formal review. Reviewer selection is critical to the publication process, and we base our choice on many factors, including expertise, reputation, specific recommendations and our own previous experience of a reviewer's characteristics. For instance, we avoid using people who are slow, careless or do not provide reasoning for their views, whether harsh or lenient. All submissions will be double blind peer reviewed. All papers are expected to have original content. They should not have been previously published and it should not be under review. Prior to the sending out to referees, editors check that the paper aim and scope of the journal. The journal seeks minimum three independent referees. All submissions are subject to a double blind peer review; if two of referees gives a negative feedback on a paper, the paper is being rejected. If two of referees gives a positive feedback on a paper and one referee negative, the editor can decide whether accept or reject. All submitted papers and referee reports are archived by journal Submissions whether they are published or not are not returned. Authors who want to give up publishing their paper in TUJE after the submission have to apply to the editorial board in written. Authors are responsible from the writing quality of their papers. TUJE journal will not pay any copyright fee to authors. A signed Copyright Assignment Form has to be submitted together with the paper.

PUBLICATION ETHICS

Our publication ethics and publication malpractice statement is mainly based on the Code of Conduct and Best-Practice Guidelines for Journal Editors. Committee on Publication Ethics (COPE). (2011, March 7). Code of Conduct and Best-Practice Guidelines for Journal Editors. Retrieved from http://publicationethics.org/files/Code%20of%20Conduct_2.pdf

PUBLICATION FREQUENCY

The TUJE accepts the articles in English and is being published 4 times (January, April, July and October) a year.

CORRESPONDENCE ADDRESS

Journal Contact: tuje@mersin.edu.tr

CONTENTS

Volume 7 – Issue 1

RESEARCH ARTICLES

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Evaulation of renewable energy source alternatives prioritization

Barış Kantoğlu[1*], İrem Düzdar Argun [1]

[1]*Duzce University, Industrial Department, Türkiye*

Keywords
Renewable Energy
Analytical network process
Consistency ratio
Geometric mean

Research Article
DOI: 10.31127/tuje.1001488

Received: 29.09.2021
Accepted: 09.02. 2022
Published: 01.04.2022

Abstract

Today the renewable energy is an alternative energy source being more important. The reasons for its importance are minimizing the risks at environmental, economic and social areas caused by the traditional energy sources and reducing the need for energy importation. Because of these it is important to determine the affecting criteria of the renewable energy sources for the investors of this industry as a guide. In modelling and analysis of the criteria and the sub-criteria MCDM (Multi Criteria Decision Making Model) can be used. In this study, the ANP model is used to propose evaluation of the renewable energy source selection criteria. In this study, an application of the recommended basis for the renewable energy source selection problem criteria and the ANP model for the optimum renewable energy source project is proposed. The expert opinions are employed to gather data in the proposed model. The technical, environmental, economic, and social risk criteria and the importance levels of the sub-criteria are analyzed under the light of the expert opinions. The Hazardous Waste, Effects to Climate Changes, Noise, Reduction in Gas Emission and National Energy Security are found as the superior criteria.

1. Introduction

The energy is one of the most important elements that trigger the development and growth attempts through the human history. The technologic improvements in industries are enlarging the continuity of energy requirements. The traditional energy sources are not enough for producing goods and services because of the technical, environmental and socio-political affairs [1-2]. Besides, these traditional sources, like coal, petroleum, natural gas and nuclear energy, are damaging the environment and causing climate changes, air and water pollution, and radiation. In the last years the majority of investments to the renewable energy sources are detected to reduce these damages, when the traditional sources are using [3-4]. The renewable energy sources, hydropower, solar, wind, biomass and geothermal power constitute approximately one third of the total energy sources of the World, with 289 billion $ investment in 2018 [5].

The evaluation of renewable energy sources is a complex problem because of the interconnection of criteria and having many of sub-criteria. To solve these types of problems systematically classifying the criteria hierarchically and determining their importance levels, the MCDM (multi criteria decision models), are used [6-8]. Many of researches can be found performed in literature evaluating the technical practicability, feasibility of costs, and minority of negative effects for environment, social benefits, and the suppressibility of risk criteria in the decision process for hydropower, solar, wind, biomass and geothermal power sources [9-14].

In this study, in evaluation of renewable energy sources supplying a hierarchical structure to determine the importance levels of criteria, the ANP model, is proposed. This model is analyzed by application.

There are many studies performed on the renewable energy evaluation. One of these studies, Wu et al [15], proposed ANP Model to evaluate the risk of the renewable energy investment in China.

Most of the studies in literature are the models used for selection of renewable energy studies proposing employment of AHP, Fuzzy AHP, DEMATEL, ANP and

* **Corresponding Author**

*(bariskantoglu@duzce.edu.tr) ORCID ID 0000-0002-7832-1619
(iremduzdar@duzce.edu.tr) ORCID ID 0000-0002-7642-8121

Cite this article

Kantoğlu, B., & Argun, İ. D. (2023). Evaulation of renewable energy source alternatives prioritization. Turkish Journal of Engineering, 7(1), 01-08

VIKOR models (Table 1). There is no research in literature sorting the important criteria performed by using ANP. In this context, this study can help the managers of the energy evaluation industry to determine the important criteria about the environment, technical, social, economic and risks.

In the second section of this study the structure of the model, ANP is explained. In the third section, the hierarchical structure of the model is presented and the outputs of the performed analyses. In the last section, the results of this study and the proposals will be seen.

Table 1. Methodology of studies in renewable energy evaluation

Authors	Publication year	Methodology
Algarin et al. [6]	2017	AHP
Karakaş and Yildiran [7]	2019	Fuzzy AHP
Budak et al. [8]	2019	AHP
Wang et al. [14]	2020	SWOT-Fuzzy AHP
Solangi et al. [1]	2019	Delphi-AHP and Fuzzy TOPSIS
Toklu and Taşkın [3]	2018	Fuzzy AHP and Fuzzy TOPSIS
Ahmad and Tahar [9]	2014	AHP
Çelikbilek and Tüysüz [10]	2016	Integrated Gray-based MCDM
Büyüközkan and Güleryüz [11]	2016	Fuzzy AHP and Fuzzy TOPSIS
Büyüközkan and Güleryüz[12]	2017	DEMATEL, ANP, TOPSIS
Rani et al. [13]	2020	Fuzzy TOPSIS

2. Method

The main problem of the Multi Criteria Decision Making problems is to determine the priority, importance and weight by evaluating more than one criterion between the choices. The AHP is, one of the MCDM techniques, commonly and effectively using to find solve this problem. The objective and subjective ideas of decision makers' can be included in AHP [16].

ANP, providing realistic and effective solutions to the complex decision-making problems. ANP is one of the multi criteria decision making techniques that involve the qualitative values as much as the quantitative values for building hierarchical models to solve the problem by evaluating the relations and interactions between the criteria formulating the model. ANP is the basic form of AHP. While the hierarchical structure from up to down is used in AHP, the decision-making criteria in hierarchical structure, the sub criteria and the interactions, feedbacks, inner and outer dependencies between choices without looking to their ranks are considered in ANP [17].

ANP, developed by Thomas L. Saaty is a method considering the measurable and experimental values to state priorities for elements affecting the solution of the problem [18]. AHP is a strong and comparative method considering the effects of qualitative and quantitative factors proposed by all of the stakeholders of the problem. The ANP can be used to make decisions by studying the social, governmental and common problems in detail. All of the elements, either tangible or intangible, affecting the optimization of problem are considered in this process. All of the internal (between cluster components) and external (between clusters) relations are evaluated in ANP. In the case of uncertainty and risk, the feedback has the ability to exhibit unexpected interacting causes for population.

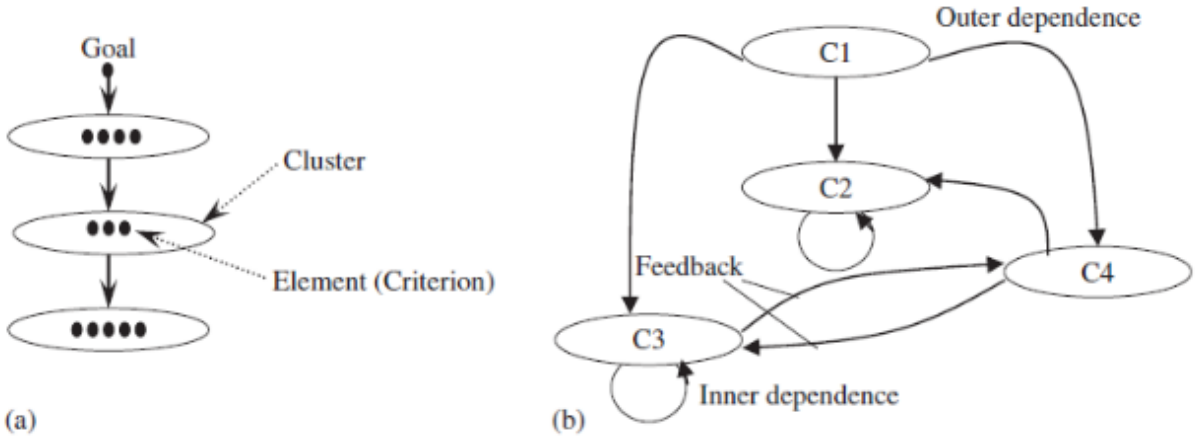


Figure 1. A hierarchy and a network [19].

Fig. 1(a) shows the hierarchical relation of a cluster; where a change occurs at the low-level cluster (affecting node), its influence is seen at the upper-level cluster (affected node). Fig. 1(b) exhibits the network relations. The arc from C4 to C2 defines the outer dependence of C2 on C4 and a loop in C2 indicates the inner dependence. The feedback in the network implies the mutual outer dependencies in a cluster pair [19].

All of the components and the possible relations between them are defined first. Then the pairwise comparisons are performed for all other components

affecting a component to analyze priorities of them in AHP and ANP.

The pairwise comparisons and matrix algebra are used to derive the weights of criteria. The final decision depends on these derived weights of the evaluative criteria [20]. In this way, the criteria are evaluated among themselves, and the importance of criteria on the choices is transformed to numbers.

The ANP contains four main steps [21]:

The first step is building of model and problem construction: The problem must be defined in detail and divided into components in a rational format as a network.

The second step is pairwise comparisons and priority vectors: The decision factors of each cluster are compared pairwise on the importance basis against their control criteria. Another pairwise examination must be performed on the interdependencies between the criteria of a cluster. Eigenvectors represent the effect of an element on others. The relative importance is defined by using the Saaty's Scale.

The third step is construction of the supermatrix: Supermatrix has the same meaning as the Markov Chains. The local priority vectors are put into the suitable columns of the matrix for determining the global priorities with interdependent effects. This is a sectioned matrix named as the supermatrix; its each element exhibits the relation between clusters.

The fourth step is synthesizing the criteria, choices, priorities to determine the best alternative: The normalized supermatrix include the priority weights for criteria and alternatives.

At the end of the pairwise comparison matrices are achieved, the aij element of this matrix is equal to 1/aij and if, i=j then aij=1.The range of wi is in between 1 and 9. 1/1 represents both criteria have equal importance where 9/1 represents significant or absolute importance [22]. The scale for this pairwise comparison can be seen at Table 2. Whatever the real value of (i;i) pair, the result is 1 and the value of (i;j) is equal to the reciprocal of the value of (j;i) or vice versa [19].

Table 2. Saaty Evaluation Scale [23]

Num value	Verbal scale
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong importance
9	Extreme or absolute importance
2,4,6,8	Intermediate values

$$\mathbf{A}=\begin{bmatrix} \mathbf{1} & \mathbf{a12} & \cdots & \mathbf{a1n} \\ \mathbf{a21}=\mathbf{1/a12} & \mathbf{1} & \vdots & \mathbf{a2n} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{an1}=\mathbf{1/a1n} & \mathbf{an2} & \cdots & \mathbf{1} \end{bmatrix} \quad (1)$$

The pairwise comparison is expressed as matrix A at Equality (1) [24].

aij is the pairwise comparison value of criterion i and criterion j. It implies that where aij is the matrix element representing the relationship between components I and j,

$$\text{aij } S=1/\text{aji} \quad (2)$$

Here, the relative importance values are calculated for all decision factors. The importance of the factor may be defined as the effect on the final decision to achieve the objective of decision maker. The required calculations, for achievement to the goal by synthesizing the arguments, are performed by employing the Super Decision software including the computation the eigenvectors of each pairwise comparison matrix, derivation of supermatrix and limit matrix that is the convergence of supermatrix and, if it is required, weighted supermatrix. The priorities of sub-nodes are according to their parent node are covered in the eigenvector [9]. To calculate the eigenvector easily, each element of the matrix is divided by the sum of the column, normalizing the pairwise comparison matrix, the sum of each column will be equal to 1 [8].

To obtain the normalized matrix the Equation (3) is used [25].

$$\mathbf{aij}=\frac{\mathbf{aij}}{\sum_{\mathbf{i=1}}^{\mathbf{n}}\mathbf{aij}} \text{ where i, j= 1,2, ..., n} \quad (3)$$

For normalizing the pairwise comparison matrix the Equation (3) is employed, the normalized matrix is found. The priority vector is calculated by using the values obtained from the calculation of average of the sum of the row of normalized matrix. To verify the result, the consistency of comparison matrices is calculated. The Consistency Index Coefficient must be calculated.

$$\mathbf{CI}=\frac{\boldsymbol{\lambda}\mathbf{maks}-\mathbf{n}}{\mathbf{n}-\mathbf{1}} \quad (4)$$

λmax is calculated by using the Equation (5)

$$\boldsymbol{\lambda}\mathbf{maks}=\left(\frac{\mathbf{1}}{\mathbf{n}}\right).\sum_{\mathbf{i=1}}^{\mathbf{n}}\left[\frac{\sum_{\mathbf{j=1}}^{\mathbf{n}}\mathbf{aij}.\mathbf{wj}}{\mathbf{wi}}\right] \quad (5)$$

Where

$$\mathbf{wi}=\sum_{\mathbf{i=1}}^{\mathbf{n}}\mathbf{aij} \quad \text{i,j=1,2, ..., n} \quad (6)$$

The largest eigenfactor of n-sized matrix A is λmax, its correct eigenfactor is q and the n-sized identity matrix is I.The estimation of relative priorities is generated by correct eigenfactor q. To derive the priorities, the eigenfactors must be scaled to as their sum is equal to 1. Saaty proved that where λmax=n is enough and necessary condition to achieve the consistency. In pairwise comparisons become inconsistent when λmax differs from n. The consistency of matrix A must be tested by using CI index, mentioned above [26].

At the end, the Consistency Ratio is calculated by using Equation (7).

$$\text{CR}=\frac{\text{CI}}{\text{RI}} \quad (7)$$

Where CI is the Consistency Index, RI is the Random Index obtained n-sized matrix and CR is the Consistency Ratio [27]. If CR<0.10, it means the matrix is consistent. If CR is greater than or equal to 0.10, to procure the

consistency some or all of comparisons have to be iterated [7].

The probable faults made in evaluations can be identified by analyzing the inconsistencies as in the real life. The logical conflicts of evaluations can be measured by inconsistencies. For example, events A, B and C are sorted according to their priorities respectively, if the judgement is C is more prior than A; this is the inconsistency [28].

The following three actions must be done when the inconsistency is greater than expected:

The most inconsistent evaluation is determined in the matrix. The interval of values is defined in which the inconsistent evaluation can be improved by changing in this interval. The decision maker must be informed, if it is possible, to change the evaluation with an acceptable value in this interval. If it is rejected by decision maker, the next evaluation is tried. If all of proposed changes are rejected the decision must be suspended to find more explicit criteria. The three acceptable ways are to correct the inconsistencies by changing the judgements [20]. The expert advices have to be combined; the reciprocal of the combined advices is equal to the combination of reciprocals of these advices. The single proven method to achieve this is using the geometric mean. The advice owners may not accept to combine their advices; the final result can be derived by using their hierarchies. In this position the geometric mean of the final outcomes is taken. The advices of the experts will be raised to power of their priorities to form the geometric mean, when their importance priorities are different [29].

The geometric mean is theoretically defined by Feng et.al.[30].

$$\mathrm{GM} = \exp\left[\int_0^\infty \ln(\mathrm{X})\mathrm{f}(\mathrm{X})\mathrm{dX}\right] \quad \text{for X} > 0 \; only \tag{8}$$

f(X) is the probability distribution of X.

3. Results

All of the factors affecting the decision are defined and grouped into clusters to prepare the network in model construction.

The clusters of nature are created after renewable energy description from the literature classification. It is shown in Fig. 2.

The ANP procedures are used to list and take the optimal criteria for this study, where the decision variables are quantitative or qualitative.

The criteria of this study are determined by a group of three experts. The application of geometric mean is performed to their judgements in accordance with the Saaty Scale.

The detailed submatrix of initial supermatrix of the constructed model (Fig. 3) is shown in Table 3. As an example, C defines the influences among the sub-criteria and criteria.

	Energy Alternatives							Criteria																																									
								Technical												Economical								Social								Environmental									Risk				
Methodology	Solar	Wind	Hydro	Biomass	Nuclear	Geothermal	Natural gaz	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	9	1	2	3	4	5
AHP	x	x	x	x				x	x	x	x	x								x	x	x	x					x	x	x	x					x	x	x	x						x	x	x	x	
Fuzzy AHP	x	x	x	x		x		x					x							x	x							x	x							x	x												
AHP	x	x	x	x	x	x			x			x		x						x	x		x					x	x		x						x		x						x	x	x		
SWOT-Fuzzy AHP	x	x		x				x	x		x			x	x					x	x			x	x	x		x	x			x	x			x		x		x									
Delphi-AHP and Fuzzy TOPSIS	x	x	x	x		x		x	x		x				x	x	x			x	x		x		x	x		x	x			x	x	x		x		x		x	x								
Fuzzy AHP and Fuzzy TOPSIS	x	x	x	x	x	x		x	x			x								x	x	x						x	x			x				x		x			x								
Fuzzy TOPSIS	x	x	x			x																																											
AHP	x	x	x	x				x	x					x						x	x				x	x		x	x							x		x			x								
Grey based MCDM	x	x	x	x		x																																											
Fuzzy AHP and Fuzzy TOPSIS	x	x	x	x	x	x		x				x	x							x	x	x						x	x						x	x			x		x								
DEMATEL-ANP	x	x	x	x		x		x	x			x			x			x	x	x	x	x			x	x		x	x		x	x	x	x	x	x	x	x											
AHP and VIKOR	x	x	x				x													x	x				x	x		x	x	x					x	x		x				x	x						
DEMATEL, ANP, TOPSIS	x	x	x	x		x		x	x	x		x						x		x	x			x	x			x	x	x	x	x	x		x	x	x	x											
Fuzzy TOPSIS	x	x	x	x	x	x																																											

Figure 2. Methodology – Renewable Energy – Criteria matrix from literature

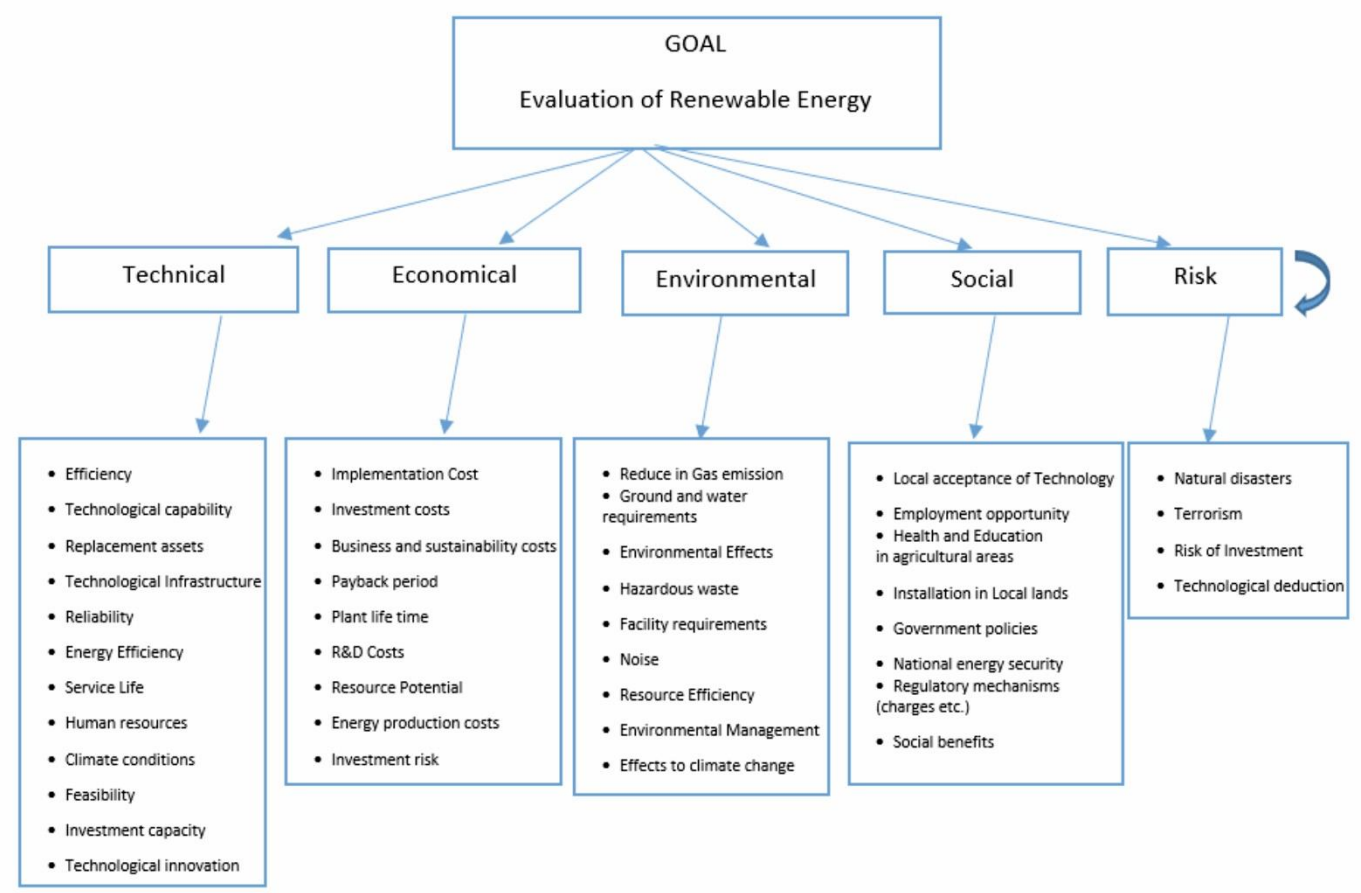


Figure 3. ANP model

Table 3. Submatrix of initial supermatrix

Sub matrix	GOAL	Criteria	Attributes
GOAL	0	0	0
Criteria	A	B	0
Attributes	0	C	0

Formulation of relations and pairwise comparisons of the clusters and elements: In order to compose the eigenvectors and then the supermatrix in the network, the links between the elements must be formulated and the required pairwise comparisons must be performed.

Cluster Comparisons: The pairwise comparisons to state the effects between the clusters are done in accordance with the control criterion of the network. The derived decisions must have consistency with the direction of efficacy. The column elements of the supermatrix are weighted by using the resulting weights of this operation.

Comparisons of Elements: The pairwise comparisons are done on the elements of clusters. The comparison is required to see the effect of an element of a cluster on the element of same or another cluster.

Comparing the Alternatives: Comparison is done for all elements according to their relations. It is necessary for construction of pairwise comparison tables processing the expert judgements on the SuperDecision software. The SuperDecision computes the weights with this operation. The question asked to experts, who have publications on collaborative innovation, is to assess the nature of five main criteria. The Saaty Scale is used to evaluate, and the geometric means are calculated for these behaviors. If the number of variables is in series n must be greater than zero; n is equal to 3 in this study. The decisions of experts can be expressed in numerical form, between 1 and 9 or in verbal meaningful words, as moderately, strongly, etc. [31].

If the point of view is the goal, the result of the pairwise comparison, the environmental criterion is the most important of the other main criteria. The weights and distribution of attributes in goal relations with main criteria is displayed in Fig. 4.

After all iterations are performed, the weights of attributes are calculated. The results have five criteria having weights greater than 10% (Table 4), which will be used as threshold to rank the criteria in this study. All 41 criteria are directed by hazardous waste criteria. It is followed by the effects on the climate changes, reduction in the gas emission and national energy security criteria. The inconsistency level is stated as acceptable according to the results of internal criteria analysis.

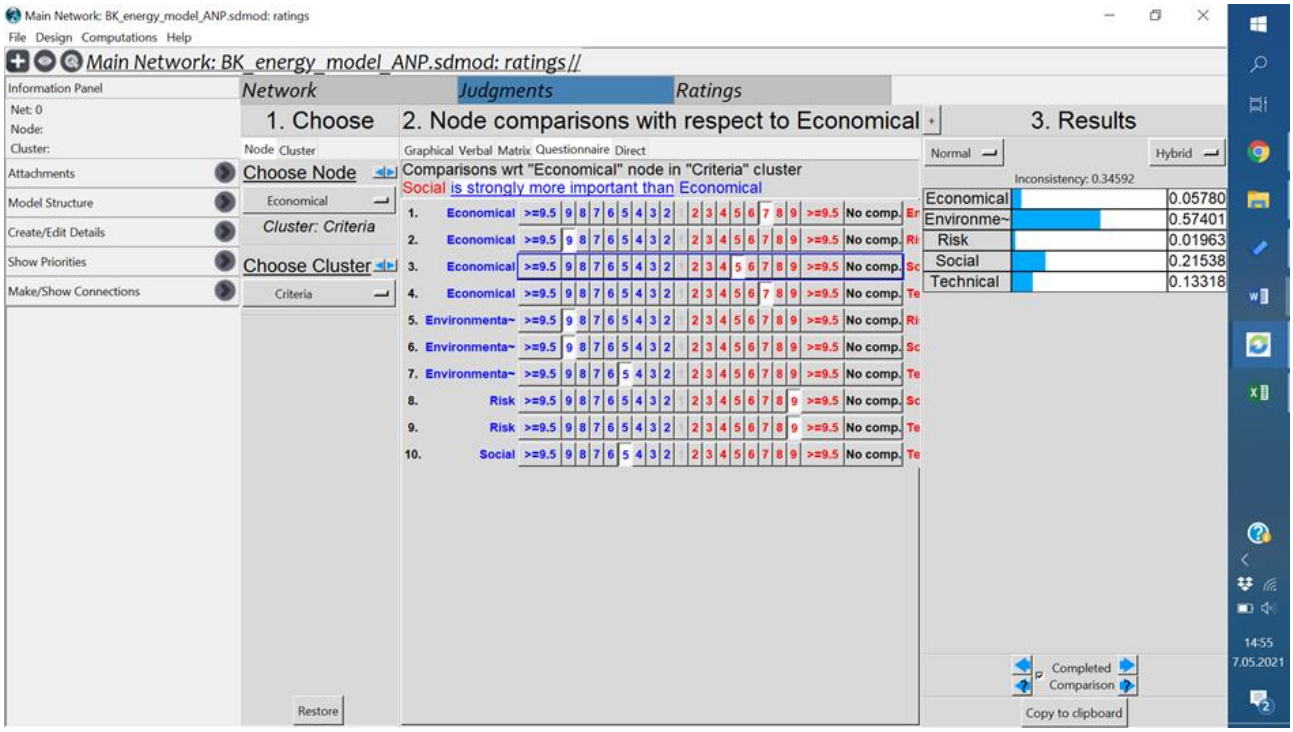


Figure 4. The weights and distribution of attributes.

Table 4. Local weights of attributes

Attributes	Local Weights
Hazardous waste	0,2
Effects to climate changes	0,1
Noise	0,1
Reduction in gas emission	0,1
National energy security	0,1

4. Discussion and Conclusion

Based on the results, some targeted investment recommendations for critical risk factors are given as follows, which could provide a reference for investors in practice. However, there are some limitations in this study. On the one hand, further researches and improvement are required for practical overseas investment on the account of constantly changing international conditions and limited available information.

In this study, an application of the recommended basis for the renewable energy source selection problem criteria and the ANP model for the optimum renewable energy source is proposed. The expert opinions are employed to gather data in the proposed model. The evaluation of the model is improved as the result of the detailed literature survey and gathering the expert opinions. To validate the performance of the proposed model an experimental search is conducted. It is decided that this model will be helpful to develop the decision-making studies having many of criteria affecting their decisions for the local administrations, investors, and practitioners after validation of the model.

More dependable decisions can be performed with ANP method, where the processing time is slightly long. ANP is a useful method to search the interrelations between the decision criteria of the problem if there are tight dependences between the criteria and the choices. The parameters can be compared with their numerical values and classified according their importance in comparison with each other on the base of data obtained from the scientific studies.

The ANP model is proposed to perform an optimum renewable energy source project, together with the application for building the framework of the criteria selection for the project in this study. Energy sources are evaluated realistically and the sight of the decision-making staff to undefinedness of the presented choices also.

The basic elements in definition of selection criteria for renewable energy sources are examined. In measurements of correlations between the adequate elements five dimensions are employed. Totally 42 assessment features are associated with these elements.

The outputs of the application are analyzed and give positive comments to the energy experts by decision making staff. The building of pairwise comparisons must be performed; derivation of data from the expert opinions and network construction is defined as an important work. It must not be forgotten that in the presented model all of the possible decision criteria and alternatives are not covered about the renewable energy projects.

Sorting the criteria on the base of their effects on studied conditions before is achieved. The environmental criteria are seen having greater values than the preset acceptable value where the values are normalized.

The Hazardous Waste criterion is found for all dimensions having the highest weight among the results. When selecting project criteria, it is not only important to consider the impact. Based on the ranking of all the

criteria, the four criteria as assessed by the experts and academicians are Effects to climate changes, Noise, and Reduction in gas emission, National energy security. This shows that the decision makers focus more on the safety and efficiency of renewable energy projects.

The results of this study emphasis some critical risk factors for usage of the investors. It is also known that this study has some restrictions. Because of the scarcity of data sources and continuous changes on the global conditions this study needs further searches and improvements.

Author contributions

Barış Kantoğlu: Literature Review, Writing-Reviewing and Editing. **Irem Duzdar Argun:** Conceptualization, Methodology, Application.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Solangi, Y. A., Tan, Q., Mirjat, N. H., Valasai, G. D., Khan, M. W. A., & Ikram, M. (2019). An integrated Delphi-AHP and fuzzy TOPSIS approach toward ranking and selection of renewable energy resources in Pakistan. *Processes*, *7*(2), 118.
2. Arto, I., Capellán-Pérez, I., Lago, R., Bueno, G., & Bermejo, R. (2016). The energy requirements of a developed world. *Energy for Sustainable Development*, *33*, 1-13.
3. Toklu, M. C., & Taşkın, H. (2018). A fuzzy hybrid decision model for renewable energy sources selection. *International Journal of Computational and Experimental Science and Engineering (IJCESEN)*, *4*(1), 6-10.
4. Boran, F. E., Boran, K., & Menlik, T. (2012). The evaluation of renewable energy technologies for electricity generation in Turkey using intuitionistic fuzzy TOPSIS. *Energy Sources, Part B: Economics, Planning, and Policy*, *7*(1), 81-90.
5. Raturi, A. K. (2019). Renewables 2019 global status report. (Paris: REN21 Secretariat). ISBN 978-3-9818911-7-1.
6. Algarín, C. R. (2017). An analytic hierarchy process based approach for evaluating renewable energy sources.
7. Karakaş, E. (2019). Evaluation of renewable energy alternatives for Turkey via modified fuzzy ahp.
8. Budak, G., Chen, X., Celik, S., & Ozturk, B. (2019). A systematic approach for assessment of renewable energy using analytic hierarchy process. *Energy, Sustainability and Society*, *9*(1), 1-14.
9. Ahmad, S., & Tahar, R. M. (2014). Selection of renewable energy sources for sustainable development of electricity generation system using analytic hierarchy process: A case of Malaysia. *Renewable energy*, *63*, 458-466.
10. Çelikbilek, Y., & Tüysüz, F. (2016). An integrated grey based multi-criteria decision making approach for the evaluation of renewable energy sources. *Energy*, *115*, 1246-1258.
11. Büyüközkan, G., & Güleryüz, S. (2016). An integrated DEMATEL-ANP approach for renewable energy resources selection in Turkey. *International Journal of Production Economics*, *182*, 435-448.
12. Büyüközkan, G., & Karabulut, Y. (2017). Energy project performance evaluation with sustainability perspective. *Energy*, *119*, 549-560.
13. Rani, P., Mishra, A. R., Mardani, A., Cavallaro, F., Alrasheedi, M., & Alrashidi, A. (2020). A novel approach to extended fuzzy TOPSIS based on new divergence measures for renewable energy sources selection. *Journal of Cleaner Production*, *257*, 120352.
14. Wang, Y., Xu, L., & Solangi, Y. A. (2020). Strategic renewable energy resources selection for Pakistan: Based on SWOT-Fuzzy AHP approach. *Sustainable Cities and Society*, *52*, 101861.
15. Wu, Y., Wang, J., Ji, S., & Song, Z. (2020). Renewable energy investment risk assessment for nations along China's Belt & Road Initiative: An ANP-cloud model method. *Energy*, *190*, 116381.
16. Gülenç, İ. F., & Bilgin, G. A. (2010). Yatırım kararları için bir model önerisi: AHP Yöntemi-A model proposal for investment decisions: AHP Method. *Öneri Dergisi*, *9*(34), 97-107.
17. Ömürbek, N. (2013). Analitik hiyerarşi süreci ve analitik ağ süreci yöntemlerinde grup karari verilmesi aşamasina ilişkin bir örnek uygulama. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *18*(3), 47-70.
18. Saaty, T. L. (2005). *Theory and applications of the analytic network process: decision making with benefits, opportunities, costs, and risks*. RWS publications.
19. Burnaz, S., & Topcu, Y. I. (2006). A multiple-criteria decision-making approach for the evaluation of retail location. *Journal of Multi-Criteria Decision Analysis*, *14*(1-3), 67-76.
20. Saaty, T. L. (2004). Decision making—the analytic hierarchy and network processes (AHP/ANP). *Journal of systems science and systems engineering*, *13*(1), 1-35.
21. Görener, A. (2012). Comparing AHP and ANP: an application of strategic decisions making in a manufacturing company. *International Journal of Business and Social Science*, *3*(11).
22. Kahraman, C., Birgün, S., & Yenen, V. Z. (2008). Fuzzy multi-attribute scoring methods with applications. In *Fuzzy Multi-Criteria Decision Making* (pp. 187-208). Springer, Boston, MA.
23. Saaty, T. (1980). The analytic hierarchy process (AHP) for decision making. In *Kobe, Japan*.
24. Çakıl, E. Y. (2017). *Bir mermer fabrikasındaki makinenin çok kriterli karar verme metotları kullanılarak seçilmesi.* Master's thesis, İzmir Kâtip Çelebi University.
25. Maya, R., & Eren, T. (2018). Türk gıda sektörünün finansal performans analizinin çok kriterli karar verme yöntemleri ile yapılması. *Verimlilik Dergisi*, *3*(3), 31-60.

26. Kurttila, M., Pesonen, M., Kangas, J., & Kajanus, M. (2000). Utilizing the analytic hierarchy process (AHP) in SWOT analysis—a hybrid method and its application to a forest-certification case. *Forest Policy and Economics*. *1*(1), 41-52.
27. Saaty, T. L. (1993). What is relative measurement? The ratio scale phantom. *Mathematical and Computer Modelling*, *17*(4-5), 1-12
28. Saaty, T. L. (2003). Decision-making with the AHP: Why is the principal eigenvector necessary. *European journal of operational research*, *145*(1), 85-91.
29. Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), 83-98.
30. Feng, C., H. Wang, & X. M. Tu. (2013). Geometric mean of nonnegative random variable. Communications in Statistics - Theory and Methods 42 (15):2714–2717.
31. Karpak, B., & Topcu, I. (2010). Small medium manufacturing enterprises in Turkey: An analytic network process framework for prioritizing factors affecting success. *International Journal of Production Economics*, *125*(1), 60-70.

Turkish Journal of Engineering

https://dergipark.org.tr/en/pub/tuje

e-ISSN 2587-1366

Comparison of online and face-to-face exams conducted in Physics I course in higher education

Mustafa Nuri Ural[1], Zeynep Başkan Takaoğlu*[2]

[1]*University of Gumushane, Department of Software Engineering, Türkiye*
[2]*University of Gümüşhane, Department of Physics Engineering, Türkiye*

Keywords
Physics I Course
Online Exam
Face-to-Face Exam

Research Article
DOI: 10.31127/tuje.1004417

Received: 25.10.2021
Accepted: 17.01.2022
Published: 01.04.2022

Abstract

The coronavirus pandemic has affected most fields, including education, and most lessons have had to be conducted as distance education. In this study, it is aimed to compare student success in face-to-face exams with that in distance exams in the Physics I course. The sample consisted of 167 students who took a multiple-choice exam in distance education in the 2020-2021 academic year and 155 students who took the multiple-choice exam in face-to-face education in the 2019-2020 academic year. In the first phase of the study, the questions asked in both types of exams were classified according to Bloom's taxonomy, and then descriptive and comparative statistics were used. As a result, according to Bloom's taxonomy, the majority of the questions asked were at the application level. In the statistical analysis, it was determined that the averages of both groups were close to 45 out of 100, and there was no significant difference between the group averages as a result of the t-test. Accordingly, it has been concluded that the use of online exams during the pandemic as an alternative to face-to-face exams in normal processes will not make a significant difference.

1. Introduction

Today's society faces different innovations in the field of technology with each passing day. These innovations are placed in every area of our lives and have become an irreplaceable part of them. Educational activities gain new and different perspectives as technology advances, and courses are delivered via the internet [1]. Especially recently, a pandemic all over the world has supported these developments, and the transition has been made rapidly from face-to-face training to online training in all areas. As a result, internet-based learning and teaching activities have almost entirely replaced face-to-face educational activities, not only in schools but also in all other sectors.

Due to the COVID-19 pandemic in our country, universities did not carry out face-to-face training for a long time and instead carried all their courses on the internet. When the applications carried out in this regard were examined, it was announced that the universities were suspended for 3 weeks on March 16, 2020, and then on March 23, 2020, they were asked to start distance education according to their capacities [2]. As a result, universities sought serious infrastructure, and the courses began to continue their educational activities for a long time as synchronous or asynchronous.

As a result, some faculty members started the distance education process unprepared and inexperienced [2]. Course materials were inadequate, and this affected the activities to be carried out negatively [3]. Especially in quantitative courses such as mathematics and physics, students mentioned the problem of communication, as well as the fact that the course contents were not fully reflected in these areas and the documents were inadequate [4-5].

Due to the COVID-19 pandemic, all educational institutions have made some innovations in their assessment and evaluation activities, with radical changes in their online educational activities due to the COVID-19 pandemic. In this regard, each institution has sought solutions to the assessment and evaluation within its own means. While education was carried out online in primary and secondary education institutions, it was

* **Corresponding Author**

(mnu23@yahoo.com) ORCID ID 0000-0001-7011-401X
*(zeynepbaskan@hotmail.com) ORCID ID 0000-0003-1706-7933

Cite this article

Ural, M. N., & Takaoğlu, Z. B. (2023). Comparison of online and face-to-face exams conducted in Physics I course in higher education. Turkish Journal of Engineering, 7(1), 09-16

decided to conduct the exams face-to-face in these institutions during certain periods. However, all assessment and evaluation activities have been moved to online environments except for certain application-based areas in universities. As a result, the instructors used different types of exams, such as multiple choice, short answer, project, portfolio, educational games, etc. in online exams [6]. Two important factors to be considered in the conduct of such exams are to take care of validity and reliability just as attention is paid to face-to-face exams [7]. Cheating is one of the most important factors that are thought to affect reliability in the studies to be carried out here [4]. In addition, asking different questions can be another topic of discussion. However, asking questions that measure similar aspects of learning in online and face-to-face exams may seem like a solution to prevent this.

Various methods have been tried in order to carry out a consistent teaching activity in education. However, dividing the objectives and measurement tools into certain categories and planning the teaching according to these categories has been one of the most accepted approaches [8]. In this way, certain levels were created by ensuring standardization in education, and it was ensured that each student was evaluated equally at these learning levels. Although different approaches are used for this purpose, the most popular application is the taxonomy developed by Bloom [9]. With the help of this taxonomy, a standardization has been established in both teaching and measurement tools. In this way, it is aimed at making sure that each and every one of the students studying in different schools and classes learns and is evaluated equally. With Bloom's taxonomy, the knowledge and skill level of the attainment determined for the students is clarified and associated with a systematic [10]. As a result, the same behavior can be measured to an equal level by using different question types in assessment and evaluation.

Concern about cheating in distance education exams and the idea that undeserved grades may be taken have been cited as the most important reasons for the formation of anxiety in students [5]. Comparing the success of face-to-face exams in real environments with the success scores obtained from online exams, on the other hand, may provide an idea of the specific opinions to be given in this regard. In the literature examined in this regard, it was seen that there are mostly studies aimed at developing online exams and the focus is on the problems carried out in online courses. For this reason, it is necessary to examine student success by comparing face-to-face exams with online exams. Therefore, the main purpose of the study is to conduct a comparison of the end-of-term exams of the Physics I course conducted face-to-face and online in terms of success. The sub-problems of working in this direction are:

1. What is the classification of the questions asked in the face-to-face paper-pen and online end-of-term exams according to Bloom taxonomy in the Physics I course?

2. Is there a statistically significant difference in the end-of-term exam results of Physics I courses between face-to-face pen and paper exams and online exams?

2. Method

2.1. Research Design

Comparative research is used to examine the relationship between at least two variables and the situations that occur between them [11]. In such studies, two different events can be involved, and comparisons can be made between these events. In the current study, the end-of-term grades of the students who took the Faculty of Engineering and Natural Sciences Physics I course between the 2019-2020 and 2020-2021 academic years were compared. In this respect, the study is a comparative research study.

2.2. Sample

The sample of this study consists of students who took the Physics I course at Gümüşhane University, Faculty of Engineering and Natural Sciences, in the fall semesters of the 2019-2020 and 2020-2021 academic years. The study is based on a comparison of these two groups. One of these groups consisted of 155 students who took this course face-to-face and participated in the Physics I end-of-term exam in the fall semester of the 2019-2020 academic year. The exam for this group was administered as a paper and pencil test. The other group consisted of 167 students who took the course online and also participated in the end-of-term exam conducted online in the fall semester of the 2020-2021 academic year. In both years, the students taking the Physics I course were from various departments. These departments were civil engineering, geomatics engineering, genetics and bioengineering, food engineering, electrical and electronics engineering, and mechanical engineering. The total sample of the study consists of 322 people. Both exams were conducted as multiple choice.

2.3. Data Collection

Within the scope of the courses given in two different periods, it was carried out by two different faculty members over the same curriculum, achievements, and textbook. The questions prepared in both terms were created separately by the instructors of the courses, but care was taken to ensure that the gains, number of content items, and quality were the same. To show that online exams are replaceable with face-to-face exams, they have to be equivalent in most respects. One of them is the level of knowledge they are assessing. Bloom's Taxonomy shows us a standardized method for defining the level of cognitive achievement. Here, different tests are applied to different students, but the number of items, their quality, and the competencies they measure are similar to each other [11]. In both exams, one held at the end of the semester in the 2019-2020 and the other held in the 2020-2021 academic years, tests consisting of 20 multiple-choice questions were administered to the students. Five choices were included in both test questions. In the face-to-face exam, students from different departments were gathered in the same lecture hall, and the exam was conducted in one session with the

same question order for all. For the online exams, an exam hour was determined, and all students were asked to take the Physics I exam at this time via the online system. In the online exam, 20 questions and 5 options per question were both mixed and presented to the students, and so the question-and-answer options order was different for each student. Unlike in the face-to-face exam, in the online exam, after completing a question, it was not possible to return and change your answer to the same question again. In Table 1, one can see examples for all three stages of Bloom's taxonomy: comprehension, application, and analysis stages, from both exams.

Table 1. Examples of Bloom's taxonomy stages for both exams

Face To Face Exam	Online Exam
Comprehension Level	
Which of the following is not a vectoral quantity? a) Velocity b) Acceleration c) Force d) Displacement e) Kinetic energy	Which of the following is a scalar quantity? a) Angular acceleration b) Velocity c) Potential energy d) torque e) displacement
Application Level	
If A=2i+3j-4k, B= I+j+7k, where A and B are vectors, find the angle between vectors A and B. a) 126.7 b) 136.7 c) 146.7 d) 156.7 e)166.7	The vectors A=i+2j-3k and B=4i-5j-6k are given. Find the sine of the angle resulting from the cross product of AxB. a) 69 b) 85 c) 60 d) 50 e) 21
Analysis Level	
Which variable or variables affect the acceleration of an object due to the change in the direction of its velocity? a) Velocity and radius b) Mass and force c) Gravitational acceleration and angular velocity d) Linear velocity and mass e) Linear acceleration and displacement	Which of the following is among the variables that affect angular acceleration? a) travel distance and linear velocity b) radius and angular velocity c) Linear velocity and mass d) Kinetic energy and mass e) Mass and force

In the study, a second researcher was asked to classify both online and face-to-face exams according to Bloom's taxonomy. Huberman and Miles [12] proposed the formula [Agreement/ (Agreement + Disagreement) x 100] to compare the secondary data obtained with the original data. As a result of this calculation, the agreement of the coders was determined as 82%. Since 70% or more of this value is accepted as a perfect fit in the calculations, it has been concluded that the coding is reliable.

In addition, the exam questions asked of both groups were analyzed on a subject-by-subject basis, and care was taken to ensure that they were close to each other. Table 2 below shows the distribution of topics for these exams.

Table 2. The distribution of the numbers of questions in online and face-to-face exams by subject

Subject	Face-to-face	Online
Physics and measurement	2	2
Vectors	1	1
One-dimensional movement	2	2
movement in two dimensions	2	2
Newton's laws of motion	2	2
Work and energy	2	2
Potential energy and conservation of energy	2	2
Linear momentum and collisions	2	2
Rotation of rigid bodies about a fixed axis	3	3
Rolling motion, angular momentum, and torque	3	3

2.4 Data Analysis

The data obtained from the study was analyzed in two stages. First, the questions in both exams were examined in terms of content. For this purpose, the questions were analyzed, taking into account the Bloom taxonomy. In this analysis, the questions were examined in accordance with the taxonomy, taking into account their contents, and notes were taken to the edge of each one. Then, within the specified grades, questions for both exam types were transferred to the tables. Finally, with the help of the tables created, the data was converted to writing and the necessary results were created.

In the study, only total scores were taken into account when examining quantitative data, and each question was examined separately. The data obtained was first graphed, and histograms of both face-to-face and online exams were presented separately. In the analysis of quantitative data, both descriptive statistics and comparative statistics were used among statistical techniques, and the results obtained were compared with each other. The general mean, median, standard deviation, and variance, skewness, and kurtosis values were found in the descriptive statistics, while the t-test analysis values were obtained for Levene and independent samples for comparative statistics. These

operations have been confirmed by calculating with SPSS.

If the test of the equality of variances is significant, Welch's t-test should be used instead of Student's t-test because the assumption of equal variances is violated. Since Welch's t-test has practically the same power as Student's t-test [13]. SPSS provides Welch's t-test statistics in the second row of the t-test statistics with the title "equal variances not assumed" section.

Finally, both test results were interpreted for the sample, and the data was displayed.

3. Results

Depending on the data obtained, the findings are presented in two sub-categories. In the first stage, the questions were examined in terms of content and presented by classifying them according to Bloom's taxonomy. In the second stage, quantitative data was examined and the findings from this part were included.

3.1. Findings on the classification of questions according to Bloom's Taxonomy

In this section, the data from the test, consisting of 20 questions for each group, totaling 40 questions, asked in the Physics I course in the 2019-2020 and 2020-2021 academic fall semesters are presented by grouping them according to Bloom's taxonomy.

Table 3. Classification of questions asked in different years according to Bloom's taxonomy

	Comprehension	Application	Analysis
Face-to-Face Exam	5	14	1
Online Exam	3	16	1

When Table 3 is examined, it is seen that the most frequently asked question level in both years was the "apply" step. While the application-level questions are 14 questions in the 2019-2020 academic year, there are 16 questions in the 2020-2021 academic year. In both years, only 1 question was asked at the analysis level. At the comprehension level, while 5 questions were asked in the 2019-2020 academic year, 3 questions were asked in the 2020-2021 academic year. As seen in Table 3, the distribution of the questions belonging to both years according to cognitive learning levels is not very different from each other.

3.2. Findings from the data obtained from the exam gradings

Descriptive statistical techniques were used in the analysis of quantitative data. The sample sizes of the two groups to be compared are quite close to each other (n=155 and n=167). The comparison of some descriptive statistics of these students can be seen in Table 4.

Table 4. The results of the Physics I course's descriptive statistics

	Online	Face to face
n	167	155
Avg	45.96	44.13
S. Dev	14.28	25.37
Variance	203.90	643.88
min	0	0
max	85	100
Kurtosis	0.66	-0.42
Skewness	0.17	-0.22
mode	45	50
median	45	50
Quartile1	40	30
Quartile2	45	50
Quartile3	50	60

Looking at Table 4, it can be seen that the numbers of members of the groups are quite close. The end-of-term exam averages of both groups are very close to each other. Whether this difference was statistically significant or not was re-evaluated with the t-test in the next section.

A remarkable situation is the difference between the standard deviation, variance, kurtosis, and skewness of

the groups. The standard deviation of the online exam is about 14.3, while the standard deviation of the face-to-face exam is about 25.4. In light of these findings, it can be said that the distribution of students who participated in the face-to-face exam is more heterogeneous. The same situation can be observed when the variance of the online end-of-term exam, which is about 204, is compared with the variance of the face-to-face exam, which is about 644.

The kurtosis of the online exam was 0.66, and the kurtosis of the paper-pencil exam was -0.42. The online exam is sharper, and the face-to-face exam is quite flat. Since one of these two values is positive and the other is negative, the difference is quite obvious. When the quartiles are examined, it is understood that the paper-pencil exam has a wider distribution.

Histogram graphs of students who participated in both face-to-face and online exams were examined separately. Figure 1-a shows the data of the students who participated in the face-to-face exams, and Figure 1-b shows the students who participated in the online exams.

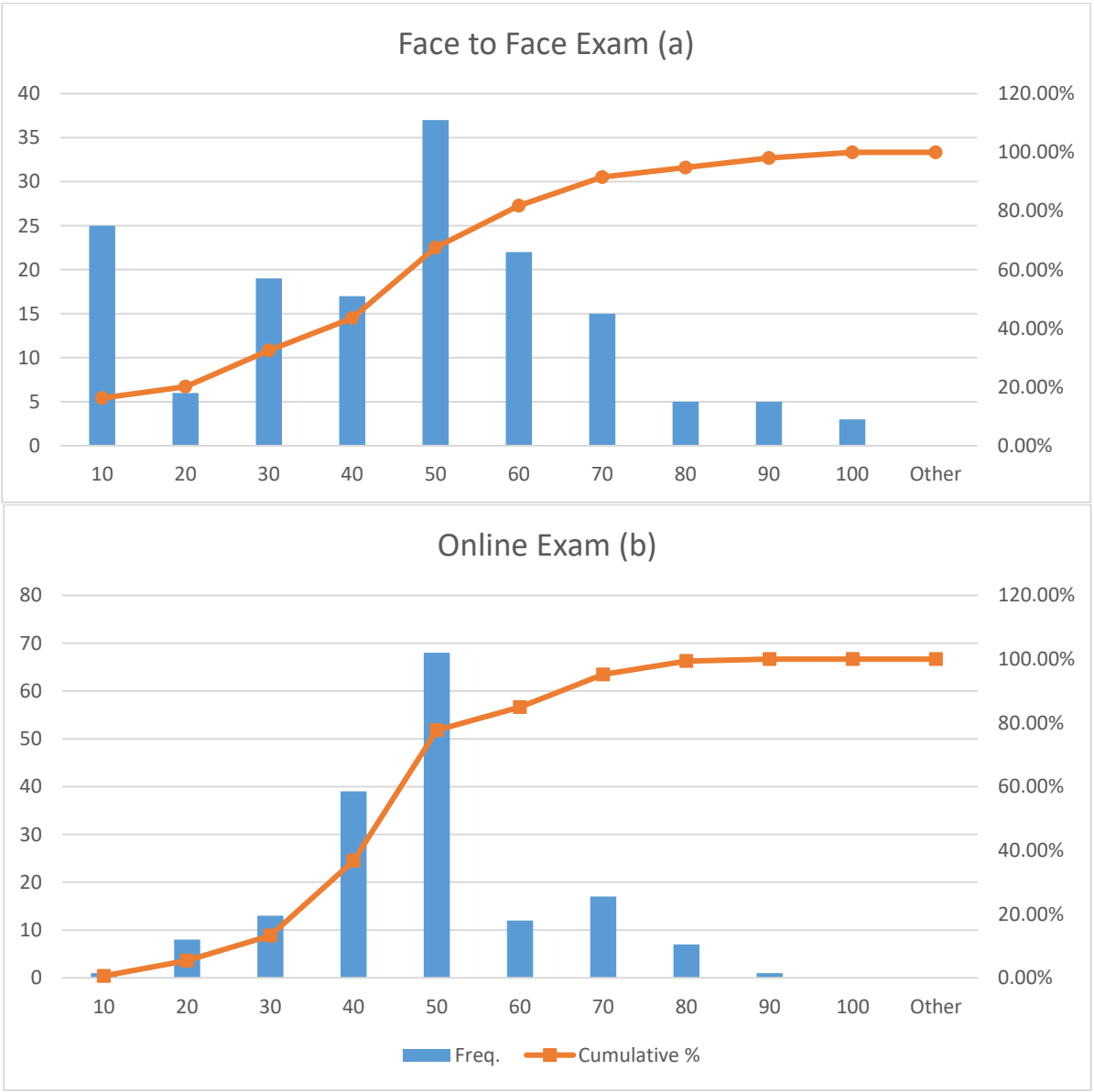


Figure 1. Face-to-face and online exam histogram

Figure 1-a shows histograms and the cumulative percentage curve for face-to-face exams. As can be seen in Figure 1-a, below 30 and over 70, the curve has a slightly less steep slope. From the histogram, it can be observed that the highest frequency of the scores is on scores 10 and 50.

Figure 1-b shows histograms and a cumulative percentage curve for online exam results. As can be seen in Figure 1-b, below 30 and over 70, the curve has a slightly less steep slope. From the histogram, it can be observed that the highest frequencies for the scores are on scores 40 and 50, while all other scores have relatively lower frequencies.

When the skewness values in Fig. 3-a and Fig. 3-b are examined, it can be said that there is a separation in terms of this parameter since the skewness of the online exam is about 0.17 and the skewness of the paper-pencil exam is -0.22, although not as much as kurtosis. The online exam is more skewed to the right than the paper-and-pencil exam is more skewed to the left. Although the difference is not huge, it is still striking.

Table 5. Data for cumulative frequencies for online and face-to-face exams

	Face-to-face Exam		Online Exam	
Score	Freq.	Cumulative %	Freq.	Cumulative %
10	25	16.23%	1	0.60%
20	6	20.13%	8	5.42%
30	19	32.47%	13	13.25%
40	17	43.51%	39	36.75%
50	37	67.53%	68	77.71%
60	22	81.82%	12	84.94%
70	15	91.56%	17	95.18%
80	5	94.81%	7	99.40%
90	5	98.05%	1	100.00%
100	3	100.00%	0	100.00%
Other	0	100.00%	0	100.00%

Table 5 represents the data for the cumulative frequencies and percentages for face-to-face exam results and for online exam results. As seen in Table 5, the scores vary between 10 and 100. There are quite a lot of students with a score of under 10 on the face-to-face exam. Other than that, the majority of the scores on the face-to-face exam are between 30 and 70.

In Table 5, it is observable that the scores for online exams vary between 10 and 90. The distribution seems quite normal at first glance. Other than that, the majority of the scores are between 30 and 70.

When Table 6 is examined, it is seen that the the average of the online group is about 46, and the average of the face-to-face exam is about 44. The standard deviation of the online group is approximately 14.28, while the standard deviation of the face-to-face exam is approximately 25.37. The mean of the online group's standard error is about 1.1, and the face-to-face exam is about 2.0.

Table 6. Distribution of end-of-term grades by groups

Group	N	Avg	S. Dev	Std. Err. Mean
Online	167	45.96	14.28	1.10
Face-to-Face	155	44.13	25.37	2.04

Table 7. Comparative statistical results of the Physics I course end-of-term exam

	F	Sig	t	df	Sig.	Mean Diff.	Std. Err. Diff.	Lower	Upper
Equal variances assumed	48.96	.000	.80	320	.42	1.83	2.27	-2.64	6.30
Equal variances are not assumed			.79	238.70	.43	1.83	2.32	-2.74	6.40

When the Levene test results are examined in Table 7, it is seen that the F value is approximately 49 and the significance is zero. According to these results, it can be said that the variances and homogeneity of the two groups are different. At the beginning of the findings section, this difference, kurtosis, skewness, etc., was also demonstrated in terms of other parameters.

When the results of the independent samples t-test are examined, it is seen that the t value is approximately 0.79. However, the significance (bipolar) was 0.43, which shows us that there is no statistically significant difference between the end-of-term exam grades of the two groups. Other data for the t-test can be examined from Table 7.

4. Discussion and Conclusion

Bloom's taxonomy consists of six steps: knowledge, comprehension, application, analysis, synthesis, and evaluation [14-17]. However, considering only three steps (comprehension, application, and analysis) in both exams and preparing questions for these steps can be expressed as a deficiency. What is expected from such exams is an assessment for all levels. What will provide this evaluation is to prepare questions for students at all levels.

In the study, the structure of the Physic I exam questions was examined according to Bloom's taxonomy. Here it is seen that both exams mostly contain application-level questions. However, in the face-to-face

exam, it was understood that several questions were at the level of comprehension. In this respect, it can be stated that there is not much difference in difficulty between online exams and face-to-face exams, and that they are exams of similar difficulty. The questions in the knowledge, comprehension, and application steps are the questions prepared to measure the knowledge at the basic level [18]. Here, the fact that both exams have questions at close levels according to Bloom's taxonomy shows that the two exams can be easily compared.

As a result of the examination of quantitative data, the descriptive statistical values of the groups, excluding skewness and kurtosis, are parallel to each other. Although the variances were different, when the t test results were examined, there was no statistically significant difference between the end-of-term exam results taken from the face-to-face exam and the results of the end-of-term exam carried out online in distance education. However, the study conducted by Yağcı, Ekiz and Gelbal [19] shows differences in the scores of students for different types of exams (face-to-face and online). It is thought that this difference in comparison with the current literature may be due to a lack of computer usage skills. Because students who use the internet too much have a positive view of online exams and do not worry too much about them [20-21]. However, recently, education has mostly moved to digital platforms, and students frequently use such platforms in their lessons, enabling them to become more familiar with subjects such as the internet, online exams, and computer knowledge. In this respect, it is believed that the validity of online exams may increase. In this context, it was seen that there was no difference in the online application of the same type of exam as an alternative to face-to-face exams, and it was concluded that there was no difference in the results of the exams in this way. It can also be said that the instructor can administer the exams online or face-to-face depending on his or her own preference. There is no difference that will affect the results of assessment and evaluation in both types of exams. The absence of such a difference will provide the instructor with the opportunity to administer exams with broader possibilities. In particular, higher education students studying in different cities from the city they live in often have problems when they have to take a make-up exam for any course. The student has to make a long journey just to take a single exam. However, such exams can be administered online, which is extremely convenient for the student. So, it is recommended that alternative assessment and evaluation tools can be used in addition to traditional assessment and evaluation tools in online exams used in the field. In addition, instructors do not rely on online exams in some cases. In such cases, although the exam is still done online, it can be supported by alternative measurement tools to make online exams more secure. In this study, only multiple-choice question types were compared. However, in assessment and evaluation, many different types of questions can be used, both face-to-face and online. It is recommended to compare other measurement tools and question types in future studies.

Author contributions

Mustafa Nuri Ural: Data Analysis, Writing, Reviewing and Editing, **Zeynep Başkan Takaoğlu:** Reviewing, Methodology, Educational Analysis, Literature Reviewing, Writing, Reviewing and Editing

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Saban, A., Özer, H. Ġ., & Tümer, A. E. (2010). Çevrimiçi ders materyalleri ve çevrimiçi sınav sistemi ile ilgili öğrenci görüşleri [Students' opinions about online course materials and online examination system]. E-Journal of New World Sciences Academy, 5(4), 2238-2244.
2. Durak, G., Çankaya, S., & İzmirli, S. (2020). Covid–19 pandemi döneminde Türkiye'deki üniversitelerin uzaktan eğitim sistemlerinin incelenmesi [Examining the Turkish universities' distance education systems during the COVID-19 pandemic]. Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education, 14(1), 787–809.
3. Çakın, M., & Akyavuz, E. K. (2020). Covid–19 süreci ve eğitime yansıması: öğretmen görüşlerinin incelenmesi [The Covid-19 process and its reflection on education: An analysis on teachers' opinions]. International Journal of Social Sciences and Education Research, 6(2), 165–186.
4. Akdemir, A., & Kılıç, A. (2020). Yükseköğretim öğrencilerinin uzaktan eğitim uygulamalarına bakışının belirlenmesi [Higher education students'views on distance education practices]. Milli Eğitim Dergisi, Salgın Sürecinde Türkiye'de ve Dünyada Eğitim [Özel Sayı], 49(1), 685-712.
5. Akıncı, M., & Pişkin, T. M. (2021). The Problems Encountered by Pre-Service Mathematics Teachers in Distance Education Practices and Solution Suggestions. Ekev Academy Journal, 25 (85), 359 – 376.
6. Baran, H. (2020). Açık ve uzaktan eğitimde ölçme ve değerlendirme [Assessment and Evaluation in open and distence education]. Öğretmen, Açıköğretim Uygulamaları ve Araştırmaları Dergisi AUAd, 6(1), 28-40.
7. Altan, T., & Seferoğlu, S. S. (2009). Uzaktan eğitimde değerlendirme süreci: Öğrenci görüşlerinin sistemin gelişimine katkıları [Evaluation process in distance education: Contribution of Students views to the development of the system]. 3 rd International Computer & Instructional Technologies Symposium (ICITS 2009). 7-9 October, Karadeniz Technical University, Trabzon
8. Birgin, O. (2016). Bloom taksonomisi. E. Bingölbali, S. Arslan, & İ.Ö. Zembat (Ed.). Matematik eğitiminde teoriler içinde (ss. 839-860). Ankara: Pegem Akademi.
9. Baştürk, S., & Taştepe, M. (2018). Bilişsel Davranışların ölçülmesi. Baştürk S. (Ed.). Eğitimde

Ölçme ve Değerlendirme içinnde (ss. 307-325). 2. Basım, Ankara: Nobel Yayın.
10. İlhan, A. & Gülersoy, A. E. (2019). Evaluation of the achievements of 10th grade geography course curriculum according to the revised bloom taxonomy. International Journal of Geography and Geography Education, 39, 10-28.
11. Çepni, S. (2007). Araştırma ve Proje Çalışmalarına Giriş, [Introduction research and Project research] (Genişletilmiş 3. Baskı.), Trabzon: Celepler Yayıncılık.
12. Hubberman, M. & Miles, M. B. (2002). The qualitative researcher's companion. California: Sage.
13. Delacre, M., Lakens, D., & Leys, C. (2017). Why Psychologists Should by Default Use Welch's t-test Instead of Student's t-test. International Review of Social Psychology, 30(1), 92–101. https://doi.org/10.5334/irsp.82
14. Bloom B S (Ed.). Englehard, M. D., Furst, E. J., Hill, W. H. & Krathwohl, D. R. (1956). Taxonomy q/'educational objectives. "The classification of educational goals. Handbook L" Cognitive domain, New York: David McKay.
15. Küçükahmet, L. (2003). Öğretimde planlama ve değerlendirme [Planning and evaluation in teaching], 13. Baskı (13th edition), Ankara: Nobel Yayıncılık.
16. Köğce, D., Aydın, M. & Yıldız, C. (2009). Bloom Taksonomisinin Revizyonu: Genel Bir Bakış [A Revision of Bloom's Taxonomy: An Overview]. İlköğretim Online, 8. 3, 1-7
17. Tutkun, Ö. F. (2012). Bloom'un Yenilenmiş Taksonomisi Üzerine Genel Bir Bakış. Sakarya University Journal of Education, 14, 14-21.
18. Erginer, E. (2006). Öğretimi planlama uygulama ve değerlendirme. Ankara: Pegem akademi yayıncılık.
19. Yağcı, M., Ekiz, H. & Gelbal, S. (2015). Yeni bir çevrimiçi sınav modeli geliştirilmesi ve uygulanması [Designing and Applying a New Online Examination Model]. Ahi Evran University Journal of Kırşehir Education Faculty. 16(1), 269-288.
20. Çiğdem, H. & Tan, Ş. (2014). Matematik dersinde isteğe bağlı çevrimiçi kısa sınav uygulanması hakkında öğrenci görüşleri [Students' Opinions on Administering Optional Online Quizzes in a two-year College Mathematics Course]. Bilgisayar ve Eğitim Araştırmaları Dergisi, 2(4), 51-73.
21. Sırakaya, M, Alsancak, S. D., & Kılıç, Ç. E. (2015). Uzaktan eğitim öğrencilerinin çevrimiçi sınava yönelik tutum düzeylerinin incelenmesi. Kastamonu Journal of Education, 23 (1), 87-104.

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Firefly-Based feature selection algorithm method for air pollution analysis for Zonguldak region in Turkey

Esra Saraç Eşsiz [*1], **Vahide Nida Kılıç** [1], **Murat Oturakçı** [1]

[1]*Adana Alparslan Türkeş Science and University, Computer Engineering Department, Türkiye*

Keywords
Machine Learning
Feature Selection
Air Pollution Analysis
Covid19

Research Article
DOI: 10.31127/tuje.1005514

Received: 11.10.2021
Accepted: 20.12.2021
Published: 01.04.2022

Abstract

Air pollution in cities is a serious environmental issue. In Turkey, the air quality index values of the measurement stations are calculated according to European Union standards. There are many kinds of measurement parameters (features) and 6 different kinds of air quality classes according to measurement stations in Turkey. Non-valuable features can be eliminated effectively with feature selection methods without any performance loss in classification. This study aims to investigate, analyze and implement a feature selection method using the FireFly Optimization Algorithm (FOA) approach. In the study, data from measurement stations for the Zonguldak region, which is known as the most polluted region in Turkey, are obtained and analyzed. Along with the acquired data, new features have been added such as day type day slots and the Covid19 feature since it is thought that curfew restrictions have an impact on air quality. The results were compared with a filter-based feature selection algorithm namely ReliefF. Experimental results show that FOA based feature selection method outperforms the ReliefF method at classification using the Random Forest classifier for air pollution even if with a fewer number of features. The Macro averaged F-score of the data set is increased from 0.685 to 0.988 using the FOA-based feature selection method.

1. Introduction

The release of pollutants in the air that significantly affects the life of living creatures is defined as the air pollution [1]. Today, the unpredictable progress of technology and industrialization causes an increase in air pollution and therefore a negative impact on human health. Some of the diseases caused by air pollution can be listed as lung diseases, respiratory disabilities, cardiovascular problems, cancer, eye disorders, and skin irritations [2-7].

In order to control air pollution, air quality should be measured with stations established in regions or cities and according to the results, necessary cautions should be taken. According to the European Union Framework Directive, it was stated that "Ozone (O_3), nitrogen dioxide (NO_2), sulfur dioxide (SO_2) and particulate matter (PM10) with particle diameters less than 10 μm" are the main parameters to be measured in the assessment [8]

In Turkey, measuring stations in all cities for air quality index calculation have been established. Sulfur dioxide (SO_2) and particulate matter (PM10) parameters, as well as nitrogen oxides (NO_x), carbon monoxide (CO) and ozone (O_3) concentrations, are also measured at the stations. For the air quality index calculation; the air quality index values of the measurement stations are calculated according to EU standards (target index), taking into account the PM10, SO_2 and NO_2 concentrations, and the results are interpreted according to six classes such as Excellent, Good, Lightly polluted; Moderately polluted; Heavily polluted and Severely polluted [9].

The fact that many studies are corresponding to the damages of air pollution and existing air quality index calculations has led to the emergence of new methods. The accuracy of the air quality index calculated with the measurement values and the accurate selection of features is an indication that the results and the measures to be taken accordingly will be more reassuring. Therefore, in this study, the effects of feature selection on air pollution calculation on the air pollution data drawn for the province of Zonguldak, which is

* **Corresponding Author**

*(esarac@atu.edu.tr) ORCID ID 0000-0002-2503-0084
(vnuzel@atu.edu.tr) ORCID ID 0000-0003-2181-9309
(moturakci@atu.edu.tr) ORCID ID 0000-0001-5946-3964

Cite this article

Eşsiz, E. S., Kılıç, V. N. & Oturakçı, M. (2023). Firefly-Based feature selection algorithm method for air pollution analysis for Zonguldak region in Turkey. Turkish Journal of Engineering, 7(1), 17-24

known as the most polluted city in Turkey, were investigated. In the study, Firefly Optimization Algorithm (FOA), which is one of the nature-inspired meta-heuristic algorithms used in many fields in recent years, was used and the FF-Based Feature Selection Algorithm Method for Air pollution Analysis was developed and applied to the data obtained. The importance and timeliness of the air quality calculation subject and the application of the algorithm used to this subject are predicted as the original contributions of this study to the literature. In addition, based on the results of the developed method, this study reveals the benefit and flexibility in terms of the applicability of the method to bigger data.

In this study, our aim is to investigate the effects of FOA based feature selection method on classification performance for air pollution analysis. For this purpose, we compared our FOA-based feature selection method with a well-known feature selection method called ReliefF. This paper is organized as follows: in the next section, we introduce the relevant studies in the literature. In the third section, we describe the methods that we use for selecting features, present the information about the data, performance evaluation for our experiments. In the fourth section, we present the experimental results and finally, we present the main conclusion and outlines for future work.

2. Related work

In the literature, methods such as Statistical, Determinative Models, Physical, Photochemical Models and Machine Learning have been used in air quality studies [10]. It has been examined that the biggest disadvantage of the methods proposed in the literature is that they require high operational performance, and it has been detected that machine learning methods do not have this disadvantage and give accurate results in studies on air pollution problems [10-12].

In the literature, there are machine learning-based air quality studies such as artificial neural network (ANN), Genetic Algorithm-ANN Model, Random Forest Model, Decision Tree Model, Deep belief network and LSSVM [13-16] proposed a new feature selection method called "Causality Based Linear Method" to select the appropriate parameters that affect air pollution and the application of the method was carried for the air quality dataset of Delhi and results were compared with existing machine learning techniques. Li et. al. [17] developed a novel forecast–analysis system for air quality index calculation novel analysis-forecast system is proposed for forecasting of air quality index by using modified Least Square Support Vector Machines (LSSVM) based on multi-objective optimization and applied in eight major cities in China.

Past studies include many applications related to algorithms inspired by nature. In recent years, developed methods or hybrid methods are used to increase the performance of algorithms inspired by nature. Ant colony optimization (ACO), particle swarm optimization (PSO), bat algorithm (BA), firefly algorithm (FA), cuckoo search (CS) and others are some of the algorithms inspired by nature. This type of algorithm tends to be global optimizers that use multiple interactive tools to create search movements in the search space (Yang, 2020). These nature-inspired algorithms used for feature selection have been the subject of many studies. For example, [18-20] performed feature selection in various areas using ACO. Jeyasingh and Veluchamy [21] and Qasim and Algamal [22] used BA developed for feature selection. Pandey et. al. [23] and Gunavathi and Premalatha [24] carried out feature selection studies in various fields using CS.

The Firefly algorithm, on the other hand, is used in the literature in the fields of engineering, decision sciences, computer sciences, economics and medical due to its effective use [25-31]. When the studies on the Firefly algorithm were examined, no study on air quality was found. The use of this algorithm in this study to fill this gap in the literature is one of the novelties of the study.

3. Methods and Dataset

In this study, we implemented an FOA-based feature selection method and give a comparison of the proposed feature selection method with filter-based feature selection method namely ReliefF at classification using the Random Forest classifier for air pollution. While ReliefF is a statistically based method, Firefly is a heuristic method. Many researchers prefer filter methods insofar as they are easy to use due to their simple algorithmic structure.

3.1. ReliefF

The relief algorithm is one of the well-known filter-based algorithms which is proposed by [32] to feature weighting. This practical and effective algorithm was extended by [33] for multi-class problems. ReliefF estimates W[A] of the quality of attribute A according to the equation in lines 8-9 in Fig. 1. In Fig. 1, n indicates the number of training instances, a indicates the number of features and m indicates the number of random training instances out of n used to update W. We use the ReliefF algorithm from Weka data mining software package [34].

Algorithm ReliefF
Input: for each training instance, a vector of feature values and the class value
1. initialize vector W
2. **for** i= 1 to m **do**
3. randomly select a target instance R_i;
4. find a nearest hits H and nearest miss M.
5. **for** A= 1 to a **do**
6. $W[A] = W[A] - diff(A, R_i, H)/m + diff(A, R_i, M)/m$
7. **end**
8. **end**
9. **return** W

Figure 1. ReliefF feature selection algorithm Pseudo code

When performing weight updates, the difference in the value of attribute A between two instances I1 and I2, where I1 = Ri and I2 are either H or M is calculated by diff

function in Fig. 1. Diff function (Equation 1) is defined as follows for discrete features [35]:

$$\text{diff}(A, I_1, I_2) = \begin{cases} 0 & \text{if value}(A, I_1) = \text{value}(A, I_2) \\ 1 & \text{if otherwise} \end{cases} \quad (1)$$

For continuous features, diff function is defined (Equation 2) as follows:

$$diff(A, I_1, I_2) = \frac{value(A, I_1) - value(A, I_2)}{max(A) - min(A)} \quad (2)$$

The max(A) and min(A) values are determined over the whole set of instances. By this normalization, all weight updates fall between 0 and 1 for all type of features. When updating W[A], to normalize final weights between -1 and 1, the output of diff function is divided by m.

3.2. Firefly Optimization Algorithm

Xin-She introduced the Firefly Optimization Algorithm, which is inspired by the social behavior of fireflies and the phenomenon of bioluminescence communication [36]. Fireflies can produce light and this production process is a type of chemical reaction. The generated light is used for communication. It can be employed for a number of goals, including finding a mate, search for food, alerting purposes to protect themselves from enemy hunters, and successful reproduction. Mostly, there are unique flashes patterns for a particular species of fireflies.

Inverse-square law is used to calculate the light intensity between the light source and a particular distance. The light intensity value decreases while the distance increases. Moreover, the air absorbs the light which becomes weaker via an increase in distance. In Firefly Optimization Algorithm, most fireflies are visible only to a limited distance. And this feature enables fireflies to communicate with each other in a limited distance. The flashing light can be used as a fitness function to be optimized. By this mapping, FOA can be used effectively in different optimization problems as well. The main steps of FOA are described in Fig. 2.

1. Generate initial population of fireflies (x_i)
2. Determine light intensity I_i for each firefly x_i by $f(x_i)$
   where $f(x_i)$ is the objective function value for by x_i
3. Define light absorption coefficient λ
4. While $(t<MaxGeneration)$ do
   4.1 for each firefly x_i do
   for each firefly x_j do
   if $(I_j > I_i)$ then
   Move firefly i towards j in d-dimension
   endif
   Update attractiveness with distance r via $exp^{-\lambda r}$
   Evaluate new solution and light intensity
   end for
   end for
   4.2 Rank the fireflies and find the current best one
   4.3 Increment t
   end while
5. Display the best firefly

Figure 2. Firefly Optimization Algorithm

In this study, an FOA-based [36] algorithm is developed to select valuable features for air pollution analysis to provide more accurate and faster classification with reduced computation cost.

Each feature is represented as a node in the proposed FOA-based feature selection method, and all nodes are independent of one another. And we used the selection probability of features Pk(i), which is the given weights of nodes by the ReliefF algorithm (prf) to select features. Evaluated Pk(i) values were used with a roulette wheel selection algorithm to select the next feature [37]. We calculated the F-measure values of selected subgroups and utilized them as fitness functions f(xi). The steps of the proposed FOA-based feature selection algorithm are given in Fig. 3.

We began experiments with a predefined number of an initial population of fireflies. In the following step, prf values of features were used to determine the initial light intensities of features. prf values were chosen as the features' light intensities because this value is a significant metric for the attractiveness of features such as fireflies' attractiveness. At first, three distinct features were chosen by each firefly randomly. After all of the fireflies accomplished their feature selection process, two .arff files were generated for each firefly's solution, such as train and test files. Over the training dataset, a decision tree classifier model was generated using the Random Forest classifier on the Weka data mining tool. Then, the test dataset was classified. The result of this classification process was evaluated according to the F-measure metric. To determine the best one of these k fireflies, mentioned basic steps were running for all fireflies. Then the other fireflies were encouraged to seem like the defined best firefly. The most appealing firefly was updated, and the light intensities of these firefly's features were updated using the F-measure value of the best firefly's solutions. The light intensity values were updated by using the given formula (Equation 3);

$$pdf(i) = pdf(i) * exp - \lambda * F - measure \quad (3)$$

In the given formula; λ=-1, i Є the best firefly's subset and F-measure value belong to the best firefly's solution. Since our aim is to increase classification performance, we used the F-measure value as a parameter in the light intensity update step.

There are two separate feature insert functions in the proposed algorithm's second phase; if firefly is discovered to be the local best firefly, this firefly chooses a new feature at random from the unselected feature list. If a firefly isn't the best firefly in the area, it chooses a feature at random from the best firefly's feature list. Two .arff files were produced for each firefly xi, namely, train and test files. We evaluated F-measure values of all fireflies and then found the local best one. These processes were repeated until the termination condition was satisfied (i.e., t = MaxGeneration). At the end of the algorithm, each firefly has chosen n number of features (Fig. 3).

1. Generate initial population of fireflies (x_i)
2. Determine light intensity I_i for each feature by their *prf* values
3. While $(t<MaxGeneration)$ do
   3.1 for each firefly x_i do
      for each firefly x_j do
         if $(I_j > I_i)$ then
            Move firefly i towards j in d-dimension
         endif
      Update attractiveness with respect to *F-measure*
      Evaluate new solution and light intensity
      end for
   end for
   3.2 Rank the fireflies and find the current best one
   3.3 Increment t
   end while
4. Display the best firefly

Figure 3. Firefly Feature Selection Algorithm

The above computations were repeated for each firefly and the best feature subset was saved. All these processes continued until the termination condition was satisfied. We defined iteration number as our termination condition. The maximum iteration number is chosen as 40 empirically.

3.3. Classification

In the proposed study, Random Forest [38] classifier was utilized to determine and evaluate the classification performances of selected feature subsets by fireflies. The performances of fireflies' solutions were compared according to F-measure values and 10-folds cross-validation was applied during experiments. All experiments were performed in Weka [34] environment.

The Random Forest classifier creates a series of decision trees from a randomly selected part of the training data. And then, to decide the final class of the test instance, the decisions of different decision trees are gathered. Furthermore, Random Forest classifiers differ from many other well-known classifiers such as discriminant analysis, support vector machines, and neural networks because they use random selections to split nodes. And by this strategy, RF can deal with over-fitting problems.

3.4. Dataset

We conduct experiments on the Turkish Air pollution dataset [39] belonging to Zonguldak province. In Fig. 4, the Turkish National Air Quality Monitoring Network map is presented. Also, a closer look at the selected province, Zonguldak, is shown.

Measurement values have been taken into consideration between 01.01.2019 and 15.04.2020 The dataset contains "PM10 (µg/m^3), PM10 Flow (m^3/hr), SO2 (µg/m^3), CO (µg/m^3), NO_2 (µg/m^3), NOX (µg/m^3), NO (µg/m^3), O_3 (µg/m^3), Temperature (°C), Relative Humidity (%), Wind Speed (km/hr), Air Pressure (mbar), Cabin Temperature (°C), PM 2.5 Flow (m^3/saat), PM 2.5 (µg/m^3), Hour, Month, Year, Day Slots, Day Type and Covid19 feature". The data was enriched by adding day slots, day type and features related to covid19 on the data extracted from the system. Day slots; for the 24-hour day zone, it is coded as 0 between 08:00 and 16:00, as 1 between 16:00 and 24:00, and as 2 between 24:00 and 08:00. For the day type; weekdays are coded as 1, weekends as 2, and national holidays as 3. The feature of Covid19 is coded based on the first case seen in China, which is November 1, 2019. While it was coded as 0 before 1 November 2019, the days until the first case in Europe were coded as 1. The days until the first case in Turkey were coded as 2 and the days after the first case is revealed were coded as 3. Features are presented in Table 1 and features that have been added in this study are shown in bold. The missing data among the extracted data was completed by taking the average of the 5 nearest neighbor values to the missing data.

In this study, the effects of feature selection on air pollution calculation on air pollution data obtained from measurement stations for the Zonguldak region were investigated. The reason for choosing Zonguldak province in the study is that the air pollution values are very high in this province and the province is known as the most air-polluted city in Turkey. In addition, the reason for adding the covid19 feature in the dataset is to see the effect of curfew restrictions on air quality.

The labeled data was created by calculating the AQI values of each data row for air pollution labels. The Air Quality Index (AQI) is a system that is scale-designed that displays air quality status to inform the public. It is also a health protection practice designed to help people maintain their health by minimizing short-term exposure to pollution and regulating activity levels during increased air pollution levels. Six separate indications are included in AQI systems, each of which is measured using different parameters. The indicators "SO_2, NO_2, CO, PM10, and PM2.5" are standardized at 1 hour and 8-hour intervals, whereas the O3 indicator is measured as a daily average. Indicators are measured separately and defined as the "Individual Air Quality Index (IAQI)". The pollutant with the highest IAQI score is defined as "Primary Pollutant". Among these six indicators, any IAQI above 100 is defined as "non-attainment Pollutant".

$$I = \frac{I_{high} - I_{low}}{C_{high} - C_{low}} (C - C_{low}) + I_{low} \quad (5)$$

In Equation 5, where I is the index for pollutant p, C is the rounded concentration of pollutant p, C_high is the breakpoint that is greater than or equal to C, C_low is the breakpoint that is less than or equal to C, I_high is the AQI value corresponding to C_high, I_low is the AQI value corresponding to C_low. The final AQI is equal to Max (I1, I2, I3, I4…In) where n is the number of pollutants [40]. According to AQI calculation, Air pollution level and Air pollution categories are decided. Table 2 presents the AQI and its corresponding categories [40].

In Fig. 5, the class distribution of the dataset is presented.

As it can be seen in Fig. 5 dataset which is used in the proposed study is an unbalanced dataset. The instance numbers for the 6 classes determined as a result of the calculation are as follows; Severely Polluted: 1, Heavily Polluted: 21, Moderately Polluted: 96, Lightly Polluted: 699, Good: 6321, Excellent: 4406. In order to ensure the class balance in the data, the first 4 classes with very little

data were combined to create the Polluted class and the class value has been reduced to 3.

In Fig. 6, the overall workflow of this study is presented.

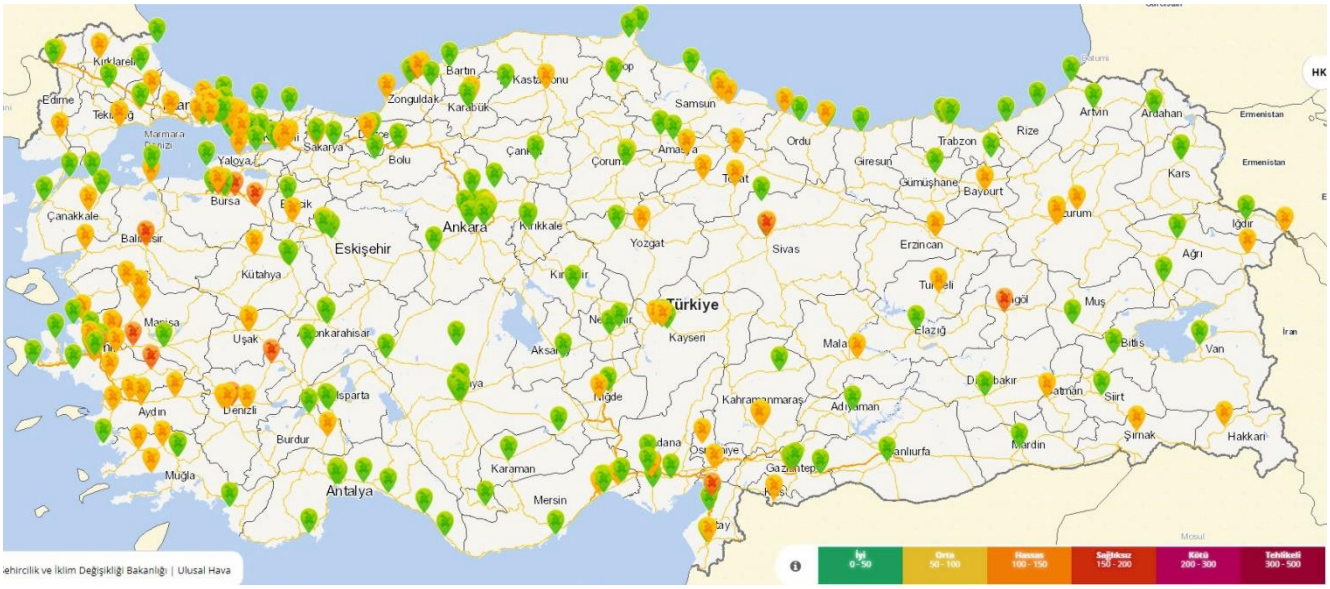


Figure 4. Geographical Location of Zonguldak and Turkish National Air Quality Monitoring Map [9]

Table 1. Features of the study

Feature	Feature Type
Day Slots	Nominal {0, 1, 2}
Hour	numeric
Day	numeric
Month	numeric
Year	numeric
Covid19	Nominal {0, 1, 2, 3}
Day type	Nominal {1, 2, 3}
PM_{10}	numeric
PM_{10} Flow	numeric
SO_2	numeric
CO	numeric
NO_2	numeric
NO_X	numeric
NO	numeric
O_3	numeric
Temperature	numeric
Wind Speed	numeric
Relative Humidity	numeric
Air Pressure	numeric
Cabin Temperature	numeric
PM_{25} Flow	numeric
PM_{25}	numeric
Wind Direction	numeric
Class	Ordinal {Good, Moderate, Unhealthy for Sensitive Groups, Unhealthy, Very Unhealthy, Hazardous}

Table 2. Air pollution categories

AQI	Air Pollution Level	Air Pollution Category
"0 to 50"	"Level 1"	"Excellent"
"51 to 100"	"Level 2"	"Good"
"101 to 150"	"Level 3"	"Lightly Polluted"
"151 to 200"	"Level 4"	"Moderately Polluted"
"201 to 300"	"Level 5"	"Heavily Polluted"
"Above 300"	"Level 6"	"Severely Polluted"

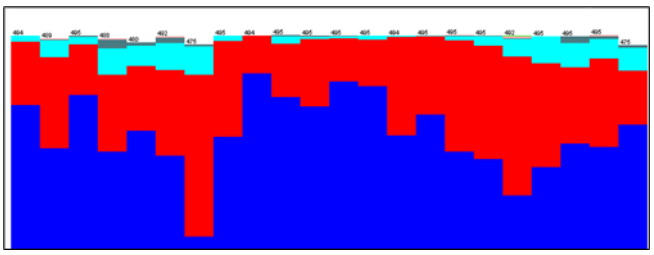


Figure 5. Class distribution of dataset

3.5. Evaluation Metrics

The classification performance of experiments is evaluated using F-score. The F score is based on the measurement terms of precision and recall. Precision (P) is the percentage of classified instances among all instances that are classified to a class correctly. Recall (R) is the percentage of instances that are classified to that class. F-score (F) is defined as the harmonic mean of recall and precision (Equation 6) [41]:

$$F = 2\frac{P * R}{P + R} \tag{6}$$

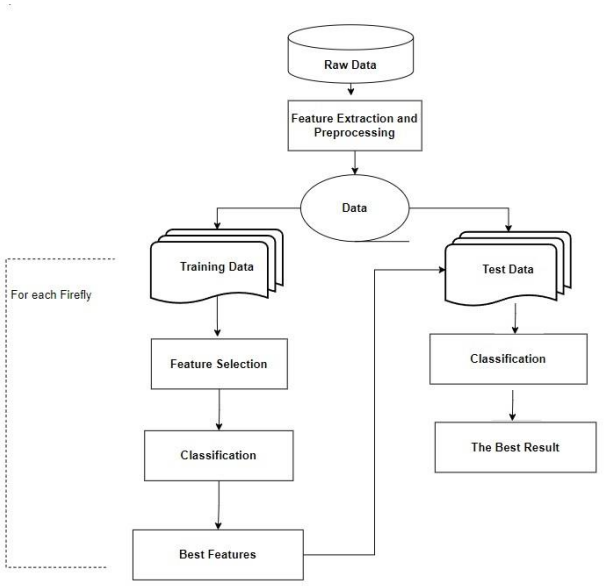


Figure 6. Overall workflow of the study

4. Experiments and Results

"The suggested FOA-based feature selection approach was compared to well-known traditional feature selection methods for air pollution analysis in this work, and the impacts of the proposed FOA-based feature selection method were studied". Using ten-fold cross-validation, the Random Forest classifier is applied. The foundation for studying the effects of feature selection strategies using all features is built in the first step of our research. Then, each feature selection method gave a vote for features and those features are ranked. Since we attempted to select the most valuable features, top-ranked 5, 10, 15, and 19 features for measuring classification performance are chosen. FOA parameters are initialized according to [36] as $\lambda = -1$. According to our previous experiments, the number of fireflies was determined as 20, iteration number (i.e., MaxGeneration) was set as 40 for the proposed method. It has been observed that fireflies can find the optimum feature subset for air pollution analysis with given parameters. And to ensure global search, every 5 iterations, 2 fireflies with the poorest performances are removed from the population and new ones are selected from the roulette wheel.

FOA-based feature selection algorithm selects a number of features that are predefined for each firefly. Afterward, the Random Forest classifier is used to evaluate the performance of each firefly's choices. The performance of the selected features was used to update the selection probability of the features. At the end of this iterative process, the best features were selected. After the selection of the best features, the test dataset was classified by using the selected feature subset. Using all of the features in the training set (as shown in Table 1), we conclude at a baseline of 0.965. According to Table 3, which compares the findings, it has been discovered that when the features are selected using our proposed FOA-based feature selection technique, classification performance improves in terms of F-score. Using the FOA-based feature selection method, the F-measure is increased from 0.965 to 0.988. Additionally, the FOA-based method outperformed the ReliefF filter-based method.

Table 3. Results Of the Feature Selection Methods with Reduced Feature Sizes

Feature Sizes	Feature Selection Methods	
	ReliefF	FOA
5	0,685	0,988
10	0,923	0,983
15	0,918	0,977
19	0,918	0,974
All Features	0,965	

While reducing feature sizes, the time required to classify test datasets were also reduced sharply without loss of accuracy in classification (Fig. 7).

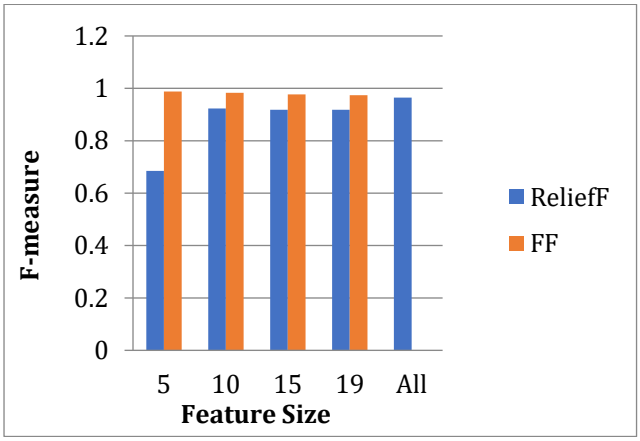


Figure 7. Results of the feature selection methods with reduced feature sizes

5. Conclusion

Air pollution is one of the leading environmental problems in the world that negatively affects living things. Measurement of air pollution levels is carried out at measurement stations depending on many parameters. Correct measurement and evaluation of measurement results are of great importance in terms of labor and costs to be allocated for the measures to be taken. Obtaining as many measurements features as required in measurement stations has also become important in terms of efficiency. For this reason, feature selection was performed for one of the measurement stations in this study. In the study, data from measurement stations for the Zonguldak region, which is known as the most polluted region in Turkey, are obtained and analyzed. Hence, an FOA-based feature selection method for the air pollution classification is proposed. All experiments were conducted by using the Turkish air pollution dataset with the Random Forest classifier.In comparison to the well-known ReliefF filter-based feature selection approach, the experimental evaluation demonstrates that our FOA-based feature selection method, which is a wrapper-based feature selection method, is able to choose preferable features.

Using the FOA-based feature selection method, the data set's Macro averaged F-score is raised from 0.685 to 0.988. The selected features were determined as PM10, PM10 Flow, NO, Wind Speed and PM2.5. Since the proposed FOA-based feature selection method is efficient in reducing the number of features, it is convenient for the classification of high-dimensional data. By reducing the feature space, our method also reduces the time required to classify test datasets sharply without loss of accuracy in classification. In future studies, different nature-inspired algorithms can be applied for similar and wider air quality datasets.

Author contributions

Esra Saraç Eşsiz: Conception and design of the research, analysis and interpretation of the data, Writing of the manuscript, Critical revision of the manuscript for important intellectual content.
Vahide Nida Kılıç: Conception and design of the research, Analysis and interpretation of the data, Critical revision of the manuscript for important intellectual content.
Murat Oturakci: Conception and design of the research, Acquisition of data, Writing of the manuscript, Critical revision of the manuscript for important intellectual content.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Dagsuyu, C. (2020). Process capability and risk assessment for air quality: An integrated approach. Human and Ecological Risk Assessment: An International Journal, 26(2), 394–405.
2. Vineis, P., & Husgafvel-Pursiainen, K. (2005). Air pollution and cancer: Biomarker studies in human populations †. Carcinogenesis, 26(11), 1846–1855.
3. Brook, R. D., Rajagopalan, S., Pope, C. A., Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., Holguin, F., Hong, Y., Luepker, R. V., Mittleman, M. A., Peters, A., Siscovick, D., Smith, S. C., Whitsel, L., & Kaufman, J. D. (2010). Particulate Matter Air Pollution and Cardiovascular Disease: An Update to the Scientific Statement from the American Heart Association. Circulation, 121(21), 2331–2378.
4. Kelly, F. J., & Fussell, J. C. (2011). Air pollution and airway disease: Air pollution and airway disease. Clinical & Experimental Allergy, 41(8), 1059–1071.
5. Gold, D. R., & Samet, J. M. (2013). Air pollution, climate, and heart disease. Circulation, 128(21).
6. Łatka, P., D. Nowakowska, K. Nowomiejska, and R. Rejdak. 2018. How air pollution affects the eyes—A review. Ophthalmology Journal 3 (2):58–62.
7. Ghorani-Azam, A., Riahi-Zanjani, B., & Balali-Mood, M. (2016). Effects of air pollution on human health and practical measures for prevention in Iran. Journal of Research in Medical Sciences, 21(1), 65.
8. Flemming, J., Stern, R., & Yamartino, R. (2005). A new air quality regime classification scheme for O, NO, SO and PM10 observations sites. Atmospheric Environment, 39(33), 6121–6129.
9. https://sim.csb.gov.tr/
10. Kaur, P., Sharma, M., & Mittal, M. (2018). Big Data and Machine Learning Based Secure Healthcare Framework. Procedia Computer Science, 132, 1049–1059.
11. Philibert, A., Loyce, C., & Makowski, D. (2013). Prediction of N_2O emission from local information with Random Forest. Environmental Pollution, 177, 156–163.
12. Kleine Deters, J., Zalakeviciute, R., Gonzalez, M., & Rybarczyk, Y. (2017). Modeling PM 2.5 Urban Pollution Using Machine Learning and Selected Meteorological Parameters. Journal of Electrical and Computer Engineering, 2017, 1–14.
13. Deleawe, S., Kusznir, J., Lamb, B., & Cook, D. J. (2010). Predicting air quality in smart environments. Journal of Ambient Intelligence and Smart Environments, 2(2), 145–154.
14. Ip, W. F., Vong, C. M., Yang, J. Y., & Wong, P. K. (2010). Least Squares Support Vector Prediction for Daily Atmospheric Pollutant Level. 2010 IEEE/ACIS 9th International Conference on Computer and Information Science, 23–28.
15. Yu, R., Yang, Y., Yang, L., Han, G., & Move, O. (2016). RAQ–A Random Forest Approach for Predicting Air Quality in Urban Sensing Systems. Sensors, 16(1), 86.
16. Sethi, J. K., & Mittal, M. (2019). A new feature selection method based on machine learning technique for air quality dataset. Journal of Statistics and Management Systems, 22(4), 697–705.
17. Li, H., Wang, J., Li, R., & Lu, H. (2019). Novel analysis–forecast system based on multi-objective optimization for air quality index. Journal of Cleaner Production, 208, 1365–1383.
18. Aghdam, M. H., & Kabiri, P. (2016). Feature selection for intrusion detection system using ant colony optimization. IJ Network Security, 18(3), 420-432.
19. Peng, H., Ying, C., Tan, S., Hu, B., & Sun, Z. (2018). An Improved Feature Selection Algorithm Based on Ant Colony Optimization. IEEE Access, 6, 69203–69209.
20. Ghosh, M., Guha, R., Sarkar, R., & Abraham, A. (2020). A wrapper-filter feature selection technique based on ant colony optimization. Neural Computing and Applications, 32(12), 7839–7857.
21. Jeyasingh, S., & Veluchamy, M. (2017). Modified Bat Algorithm for Feature Selection with the Wisconsin Diagnosis Breast Cancer (WDBC) Dataset. Asian Pacific Journal of Cancer Prevention, 18(5).
22. Qasim, O. S., & Algamal, Z. Y. (2020). Feature Selection Using Different Transfer Functions for Binary Bat Algorithm. International Journal of Mathematical, Engineering and Management Sciences, 5(4), 697–706.
23. Pandey, A. C., Rajpoot, D. S., & Saraswat, M. (2020). Feature selection method based on hybrid data transformation and binary binomial cuckoo search. Journal of Ambient Intelligence and Humanized Computing, 11(2), 719–738.

24. Gunavathi, C., & Premalatha, K. (2015). Cuckoo search optimisation for feature selection in cancer classification: A new approach. International Journal of Data Mining and Bioinformatics, 13(3), 248.
25. Pan, F., Ye, C., Wang, K., & Cao, J. (2013). Research on the Vehicle Routing Problem with Time Windows Using Firefly Algorithm. Journal of Computers, 8(9), 2256–2261.
26. Alweshah, M. (2014). Firefly Algorithm with Artificial Neural Network for Time Series Problems. Research Journal of Applied Sciences, Engineering and Technology, 7(19), 3978–3982.
27. Abdelaziz, A. Y., Mekhamer, S. F., Badr, M., Algabalawy, M.A. (2015). The firefly meta-heuristic algorithms: developments and applications. International Electrical Engineering Journal (IEEJ), 6(7),1945–1952
28. Kumar, A., & Khorwal, R. (2017). Firefly Algorithm for Feature Selection in Sentiment Analysis. In H. S. Behera & D. P. Mohapatra (Eds.), Computational Intelligence in Data Mining (Vol. 556, pp. 693–703). Springer Singapore.
29. Wang, H., Wang, W., Cui, Z., Zhou, X., Zhao, J., & Li, Y. (2018). A new dynamic firefly algorithm for demand estimation of water resources. Information Sciences, 438, 95–106.
30. Sawhney, R., Mathur, P., & Shankar, R. (2018). A Firefly Algorithm Based Wrapper-Penalty Feature Selection Method for Cancer Diagnosis. In O. Gervasi, B. Murgante, S. Misra, E. Stankova, C. M. Torre, A. M. A. C. Rocha, D. Taniar, B. O. Apduhan, E. Tarantino, & Y. Ryu (Eds.), Computational Science and Its Applications – ICCSA 2018 (Vol. 10960, pp. 438–449). Springer International Publishing.
31. Dash, S., Thulasiram, R., & Thulasiraman, P. (2019). Modified firefly algorithm with chaos theory for feature selection: A predictive model for medical data. International Journal of Swarm Intelligence Research (IJSIR), 10(2), 1-20.
32. Kira, K., & Rendell, L. A. (1992). A Practical Approach to Feature Selection. In Machine Learning Proceedings 1992 (pp. 249–256). Elsevier.
33. Kononenko, I. (1994). Estimating attributes: Analysis and extensions of RELIEF. In F. Bergadano & L. Raedt (Eds.), Machine Learning: ECML-94 (Vol. 784, pp. 171–182). Springer Berlin Heidelberg.
34. http://www.cs.waikato.ac.nz/ml/weka
35. Robnik-Šikonja, M., & Kononenko, I. (2003). [No title found]. Machine Learning, 53(1/2), 23–69.
36. Yang, X.-S. (2008). Nature-inspired metaheuristic algorithms. Luniver Press.
37. Bäck, T. (1996). Evolutionary algorithms in theory and practice: Evolution strategies, evolutionary programming, genetic algorithms. Oxford University Press.
38. Ho, T.K. (1995) Random Decision Forest. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, 14-16 August 1995, 278-282.
39. https://sim.csb.gov.tr/Services/AirQuality
40. Gao, F. (2013). Evaluation of the Chinese new air quality index (GB3095-2012): based on comparison with the US AQI system and the WHO AQGs.
41. Han, J. and Kamber, M. (2006) Data Mining: Concepts and Techniques. 2nd Edition, Morgan Kaufmann Publishers, San Francisco.

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

A combined approach of base and meta learners for hybrid system

Abdul Ahad Abro*[1], Waqas Ahmed Siddique [2], Mir Sajjad Hussain Talpur [3], Awais Khan Jumani[2], Erkan Yaşar[1]

[1]*Ege University, Department of Computer Engineering, Türkiye*
[2]*Ilma University, Department of Computer Science, Pakistan*
[3]*Sindh Agriculture University, Information Technology Centre, Pakistan*

Keywords
Artificial Intelligence
Machine Learning
Ensemble Learning
Pattern Recognition
Data Mining
Classification

Research Article
DOI: 10.31127/tuje.1007508

Received: 09.10.2021
Accepted: 20.02. 2022
Published: 01.04.2022

Abstract

The ensemble learning method is considered a meaningful yet challenging task. To enhance the performance of binary classification and predictive analysis, this paper proposes an effective ensemble learning approach by applying multiple models to produce efficient and effective outcomes. In these experimental studies, three base learners, J48, Multilayer Perceptron (MP), and Support Vector Machine (SVM) are being utilized. Moreover, two meta-learners, Bagging and Rotation Forest are being used in this analysis. Firstly, to produce effective results and capture productive data, the base learner, the J48 decision tree is aggregated with the rotation forest. Secondly, machine learning and ensemble learning classification algorithms along with the five UCI Datasets are being applied to progress the robustness of the system. Whereas, the recommended mechanism is evaluated by implementing five performance standards concerning the accuracy, AUC (Area Under Curve), precision, recall and F-measure values. In this regard, extensive strategies and various approaches were being studied and applied to obtain improved results from the current literature; however, they were insufficient to provide successful results. We present experimental results which demonstrate the efficiency of our approach to well-known competitive approaches. This method can be applied to image identification and machine learning problems, such as binary classification.

1. Introduction

In the field of data mining, the classification task is to correctly predict the class of a given instance. Several theoretical and empirical studies have been published that demonstrate the advantages of the hybrid model. These approaches are known as multi-classifiers or ensembles. A huge number of research was carried out to produce multiple classifier systems based on the same classifier models trained on different data or feature subsets [1-2].

The primary agenda of the research is to evaluate and compare various techniques (J48, MP, SVM) with Bagging and Rotation Forest for binary classification. In this paper, we provide a technique based on the J48 Machine Learning algorithm integrated with the rotation forest ensemble learning algorithm [3-4].

Decision tree J48 is the execution of an algorithm (Iterative Dichotomiser 3). J48 algorithm is a classification algorithm producing a decision tree focused on information theory. It is one of the best machine learning algorithms for categorising and continuously examining data [5]. To produce accurate classification results, the J48 method is utilised to classify numerous applications.

On the other side, Rotation Forest is a method focused on feature extraction for generating classifier ensembles [6]. It has been broadly used to resolve a variety of tasks relating to medical images, computer vision and machine learning to achieve outstanding performances.

In [7], bagging and classification tree methods were combined to introduced the BAGCT and BAGCT-SVM framework to improve the reliability and robustness. The

* **Corresponding Author**

*(abdulahadabro1@gmail.com) ORCID ID 0000-0002-3591-9231
(eng.waqas@outlook.com) ORCID ID 0000-0001-8206-4451
(mirsajjadhussain@sau.edu.pk) ORCID ID 0000-0001-9897-3916
(awaisjumani@yahoo.com) ORCID ID 0000-0001-9468-0446
(erkan.yasar@ege.edu.tr) ORCID ID 0000-0001-7333-6320

Cite this article

Abro, A. A., Siddique, W. A., Talpur, M. S. H., Jumani, A. K. & Yaşar, E. (2023). A combined approach of base and meta learners for hybrid system. Turkish Journal of Engineering, 7(1), 25-32

outcomes indicate that the BAGCT-SVM contributes improved analytical capability than CT and SVM.

This paper has been structured with several sections. Section 2, discusses the related works about machine learning and ensemble learning algorithms, mainly focusing on J48 and Rotation Forest. Section 3, presents the proposed methodology in detail. Section 4, describes the datasets, experimental methods, and outcomes. Finally, the conclusion and future works are stated in Section 5.

2. Related work

In general, hybrid system base and meta-approaches can enhance the effects and integrate the dynamic approaches in the system. A surge of research efforts has been recently witnessed for the classifications based on J48, MP, SVM and Bagging along with Rotation Forest. In this paper, we have included the classification of datasets concerning the base learners and meta-learners.

We analyzed various research articles to find current state-of-the-art developments in the domain of the Hybrid System. A few of them are discussed as follows:

In [8], proposed a hybrid model for Parkinson's diagnosis using machine learning techniques. The hybrid model includes feature selection methods such as an extra tree and mutual information gain and three classifiers k-nearest-neighbors, random forest and naive bayes. The combination of random forest and the genetic algorithm was performed and 95.58% accuracy was achieved.

In [9], the model is suggested primarily to assist and optimize the movement patterns of aged people. A new classifier named Apriori based Probability Tree Classifier (APTC) is integrated into the bagged J48 machine learning algorithm to yield a better outcome.

In [10], multiple ensemble methods Random SubSpace, Rotation Forest, Bagging, MultiBoost, Dagging and AdaBoost with the base classifier of Multiple Perceptron Neural Networks. The execution of the base classifier of MLP significantly improved concerning AUC. The results of the review are indicated in the current research, and paradigms using machine learning ensemble frameworks have worked properly.

In [11], extreme learning machine (ELM) created hierarchical learning structure was proposed for MP. The architecture of ELM based on feature extraction and random has initialized hidden weights. This method had better learning efficiency than Deep learning. The proposed algorithm achieved better and faster convergence than the existing state-of-the-art hierarchical learning methods.

In [12], robust machine learning SVM-based algorithms has been suggested. It is based on the framework of the double duality strategy of the decision-making process to get the additional constraints for optimization variables incorporated of imprecise information.

In [13], a hybrid ensemble learning method bagging, boosting, random forest and rotation forest along with logistic regression with stacking classifiers were introduced, which resultant occupy more space and consume more time for computations.

In [14], a rotation forest algorithm created on heterogeneous classifiers ensemble is applied to classified the gene expression outline. The local optimum and overfitting were improved through heterogeneous rotation forests. It improves the high stability, classification accuracy and time efficiency.

In [15], proposed a collaborative approach of blockchain and metaheuristic-enabled genetic algorithms. Blockchain technology provides a secure communication channel between stakeholders where a metaheuristic-enabled genetic algorithm, process and analyze the forecast pricing from records by scheduling, managing and monitoring them in real-time from day-to-day agriculture production detail. This approach achieved 95.3% accuracy and maintains transparency, integrity, availability and secure operational control access.

In [16], propsed the state-of-the-art utilization of ML algorithm, which are C4.5 (J48), K-Nearest Neighbor (KNN), Logistic Regression (LR), Naive Bayes (NB), Support Vector Machine (SVM), and One Rule (OneR) along with the five UCI Datasets. A retrospective study that looked at different sizes of training and test sets had a significant impact on the sensitivity and specificity of the same algorithm. The collaborative nature of the proposed system is to improve the efficiency of binary classification.

In [17], rotation forest algorithms are proposed for gene appearance of data classification. Three types of classification named; misclassification, test and rejection cost were integrated into the framework to make it more reliable and efficient. The experimental results have shown that the overall classification accuracy was improved significantly.

In [18], proposed a Parkinson's diagnosis system by using an optimized version of the BAT algorithm. Only 23 features were selected from the UCI Parkinson's disease classification data set and directly feed into the 23 neurons in the input layer of the model. The 96.74% accuracy was achieved by the proposed method with a 3.27% loss.

In [19], address the state-of-the-art utilization of ML in computer vision and image processing. This survey will provide details about the type of tools and applications and datasets. Multiple techniques and various types of supervised and unsupervised ML algorithms, the overview of image processing and the results based on the impact; neural network-enabled models, limitations, tools and application of computer vision have been discussed.

In [20], the metaheuristic optimization procedure along with the whale optimization set of rules and rotation forest algorithm was applied for the selection of email features and categorising the emails as spam and non-spam. The results obtained showed that the suggested technique generated notable improvement as compared with some previous methods.

In [21], compared and investigated state-of-the-art ensemble techniques Bagging, AdaBoost and Rotation Forest with the base classifier of J48 for the susceptibility of the landslide. The performance was assessed through ROC, AUC and statistical indexes. The J48 with the Rotation Forest model presented the highest prediction

capability followed by AdaBoost and Bagging respectively. Moreover, J48 with Rotation Forest has proved the best-enhanced approach and promising one for better accuracy.

In [22], SVM, Naïve Bayes, Logistic Regression and K-Nearest Neighbor classifier had been utilized for binary classification. In supervised ML algorithms, the robustness of the method is progressed accordingly.

In the literature, some features have a negative impact on classification algorithms. The primary goal of classification is to reliably predict the target class for each occurrence in the data. A classification algorithm coordinates between the values of the predictors and the values during the model build training process.

3. Methodology

This section provides an overview of the proposed method, which describes the pre-processing stage of data and classification algorithms.

3.1. Overview of the proposed system

An overview of the proposed framework is given in Fig. 1. This system is composed of many phases: datasets, base learners, meta-learners and comparative analysis of the results. In addition, method generalization efficiency, 10-fold cross-validation, is used for all learners and datasets of the classifier.

3.2. Data pre-processing

In this phase, the ranges of the values of the data in datasets may be high. In such a scenario, certain features can significantly or negatively affect algorithms for classification accuracy. Therefore, the data assessments are normalized to the [0,1] range using the min-max normalization technique [23-24]. For mapping a value, of a feature x_i from the range [min(x_i), max(x_i)] to a new range [minx_{new}, maxx_{new}], the normalized feature $\hat{x}_i$ is computed as Eq. 1.

$$\hat{x}_i = \frac{x_i - \min_{x_i}}{\max_{x_i} - \min_{x_i}} (\max x_{new} - \min x_{new}) + \min x_{new} \quad (1)$$

3.3. Classification of algorithms

In this study, three base learners, including J48, MP, SVM and two Meta-Learners Rotation Forest and Bagging are employed as shown in Figure 1.

There are numerous phases of methods related to the datasets and classifiers. In this work, base learners and meta-learner along with several datasets, are experienced for binary classification.

4. Experimental work

In these subsections, we define and present the experimental procedure, measurements of evaluation and results of the experiment.

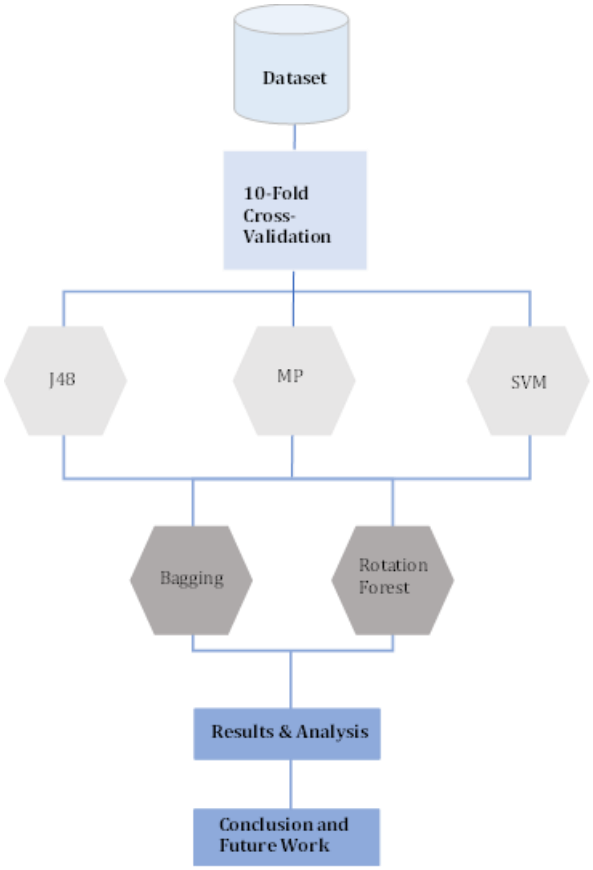


Figure 1. The framework of the method

4.1. Experimental process

In the experimental process, five datasets have been used from the UCI ML Repository [25].

All experiments are performed on base and meta-learners by using WEKA (Waikato Environment for Knowledge Analysis) ML toolkit and JAVA programming language [24]. We have utilized default parameter values for all the classifiers in WEKA.

On the other hand, we have carried out 10-fold cross-validation to all datasets to yield reliable results. The cross-validation is imposed on the original dataset randomly partitioned into 10 equally sized sets, one of which is used as test validation, while the remaining sets are used for training operations. The method is recurring 10 times and calculates the averages of the results.

Dataset characteristics are evaluated concerning the attributes and the number of instances. There are various numerical attribute descriptions illustrated in Table I. The number of instances, attributes, and classes for each dataset are presented in Table I. It is determined by investigating the appropriate data or datasets which are being utilized for binary classification problems.

Table 1. This is the example of table formatting

Datasets	Instances	Attributes	Classes
Abalone	4177	8	29
Balance Scale	625	4	3
Diabetes	768	8	2
German Credit	1000	21	2
Sonar	208	60	2

In this work, various approaches have been carried out along with several datasets, which are considered suitable for the classification. However, the performance metrics are calculated according to the binary classification problems based on the confusion matrix.

4.2. Assessment of measures

This section explains the proposed method's five performance assessment metrics, consisting of accuracy, AUC, precision, recall, and F-measure.

Accuracy reflects how close an agreed number is to a measurement. It is specified further in Eq. (2).

$$\text{Acc} = \left(\frac{TP+TN}{TP+FP+FN+TN}\right) \quad (2)$$

In equation 2, TN, FN, FP and TP show the number of True Negatives, False Negatives, False Positives and True Positives [13,16].

AUC represents the area under the ROC Curve. AUC calculates the whole two-dimensional area beneath the whole ROC curve from (0,0) to (1,1).

Precision is a positive analytical value [22,24]. Precision defines how reliable measurements are, although they are farther from the accepted value.

The equation of precision is shown in Eq. (3).

$$\text{Precision} = \left(\frac{TP}{TP+FP}\right) \quad (3)$$

The recall is the hit rate [13,16,22,24]. The recall is the reverse of precision; it calculates false negatives against true positives. The equation is illustrated in Eq (4).

$$\text{Recall} \quad \left(\frac{TP}{TP+FN}\right) \quad (4)$$

F-measure can be defined as the weighted average [13,16,22,24], of precision and recall. This rating considers both false positives and false negatives. The equation is illustrated in Eq (5).

$$F = 2\,x\frac{Precision * Recall}{Precision + Recall} \quad (5)$$

These criteria are adjusted proportionally in the data by the reference class prevalence in the weighting operation.

4.3. Experimental results

Table 2 presents classification accuracies for all datasets, base and ensemble learners. As it can be observed from Table 2, Rotation Forest with J48 gives highly accurate results than other approaches. In addition to the fact that meta learner bagging produces more accurate results than J48, MP and SVM base learners.

In Table 3, weighted precision values obtained by all base and ensemble classifiers for all datasets are presented.

In Table 4 and 5, weighted recall and weighted F-measure values are illustrated for all datasets, base and ensemble classifiers, respectively.

In Table 6, weighted AUC values are introduced for all datasets, base and ensemble classifiers. According to Table VI, Rotation Forest gives the best results very close or equal to 1.0. So, Rotation Forest can be determined to be a very powerful and effective classifier.

The balance scale, sonar and diabetes datasets have significant outputs concerning the accuracy, precision, recall, F-measure and AUC parameters; however German Credit has somehow satisfactory output and Abalone shows lower outcomes in Table 2-6.

Furthermore, it is analyzed that the Meta learner's rotation forest provides a more accurate outcome. Likewise, Meta learners bagging indicates adequate consequences. In addition, base learners provide positive findings.

Similarly, Figure 2-6, indicates the accuracy, AUC, precision, recall and F-measures values accordingly.

Table 2. Classification accuracies (%) For Uci datasets

	Base Learner			**Meta Learner Bagging**			**Meta Learner Rotation Forest**		
Datasets	**J48**	**MP**	**SVM**	**J48**	**MP**	**SVM**	**J48**	**MP**	**SVM**
Abalone	21.12	26.24	24.11	23.10	**27.15**	23.63	**24.61**	27.00	**27.48**
Balance Scale	76.64	90.72	89.76	84	92.48	90.08	**90.72**	**94.24**	**90.40**
Diabetes	73.82	75.39	65.10	74.61	**76.82**	65.10	**76.30**	76.30	**76.30**
German Credit	70.50	71.50	68.70	73.30	**76.10**	68.60	**74.80**	75.40	**76.70**
Sonar	71.15	82.21	65.87	77.88	**83.65**	62.98	**79.81**	80.77	**85.10**

Table 3. Weighted precision values for Uci datasets

	Base Learner			**Meta Learner Bagging**			**Meta Learner Rotation Forest**		
Datasets	**J48**	**MP**	**SVM**	**J48**	**MP**	**SVM**	**J48**	**MP**	**SVM**
Abalone	**0.36**	**0.43**	0.23	0.17	0.37	0.17	0.30	0.40	**0.43**
Balance Scale	0.73	0.92	0.83	0.81	0.92	0.86	**0.89**	**0.94**	**0.83**
Diabetes	0.73	0.75	0.65	0.74	**0.76**	0.65	**0.76**	**0.76**	**0.76**
German Credit	0.69	0.71	0.49	0.72	**0.75**	0.52	**0.73**	0.75	**0.76**
Sonar	0.71	0.82	0.72	0.78	**0.84**	0.66	**0.80**	0.81	**0.85**

Table 4. Weighted Recall Values For Uci Dataset

	Base Learner			Meta Learner Bagging			Meta Learner Rotation Forest		
Datasets	**J48**	**MP**	**SVM**	**J48**	**MP**	**SVM**	**J48**	**MP**	**SVM**
Abalone	0.21	0.26	0.24	0.23	**0.27**	0.24	**0.25**	**0.27**	**0.27**
Balance Scale	0.77	0.91	0.89	0.84	0.92	0.90	**0.91**	**0.94**	**0.90**
Diabetes	0.74	0.75	0.65	0.75	**0.77**	0.65	**0.76**	0.76	**0.76**
German Credit	0.70	0.71	0.69	0.73	**0.76**	0.69	**0.75**	0.75	**0.77**
Sonar	0.71	0.82	0.66	0.78	**0.84**	0.63	**0.79**	0.81	**0.85**

Table 5. Weighted F-Measure Values For Uci Datasets

	Base Learner			Meta Learner Bagging			Meta Learner Rotation Forest		
Datasets	**J48**	**MP**	**SVM**	**J48**	**MP**	**SVM**	**J48**	**MP**	**SVM**
Abalone	**0.40**	**0.47**	0.10	0.15	0.38	0.03	0.24	0.41	**0.39**
Balance Scale	0.75	0.91	0.86	0.82	0.92	**0.87**	**0.89**	**0.93**	**0.87**
Diabetes	0.74	0.75	**0.79**	0.74	**0.76**	**0.79**	**0.75**	**0.76**	0.75
German Credit	0.69	0.71	0.57	0.72	**0.75**	0.57	**0.74**	0.75	**0.74**
Sonar	0.71	0.82	0.62	0.78	**0.84**	0.59	**0.79**	0.81	**0.85**

Table 6. Weighted Auc values for Uci datasets

	Base Learner			Meta Learner Bagging			Meta Learner Rotation Forest		
Datasets	**J48**	**MP**	**SVM**	**J48**	**MP**	**SVM**	**J48**	**MP**	**SVM**
Abalone	0.59	0.77	0.56	0.70	0.77	**0.59**	**0.72**	**0.78**	0.58
Balance Scale	0.81	0.98	0.91	0.93	**0.99**	0.96	**0.99**	**0.99**	0.94
Diabetes	0.75	0.79	0.50	0.79	**0.82**	0.50	**0.82**	**0.82**	**0.73**
German Credit	0.64	0.73	0.49	0.75	**0.78**	0.49	**0.78**	0.39	**0.69**
Sonar	0.74	0.88	0.64	0.89	**0.91**	0.70	**0.90**	0.89	**0.88**

- High Acc, AUC, Precision, Recall and F- measure is shown in Bold.

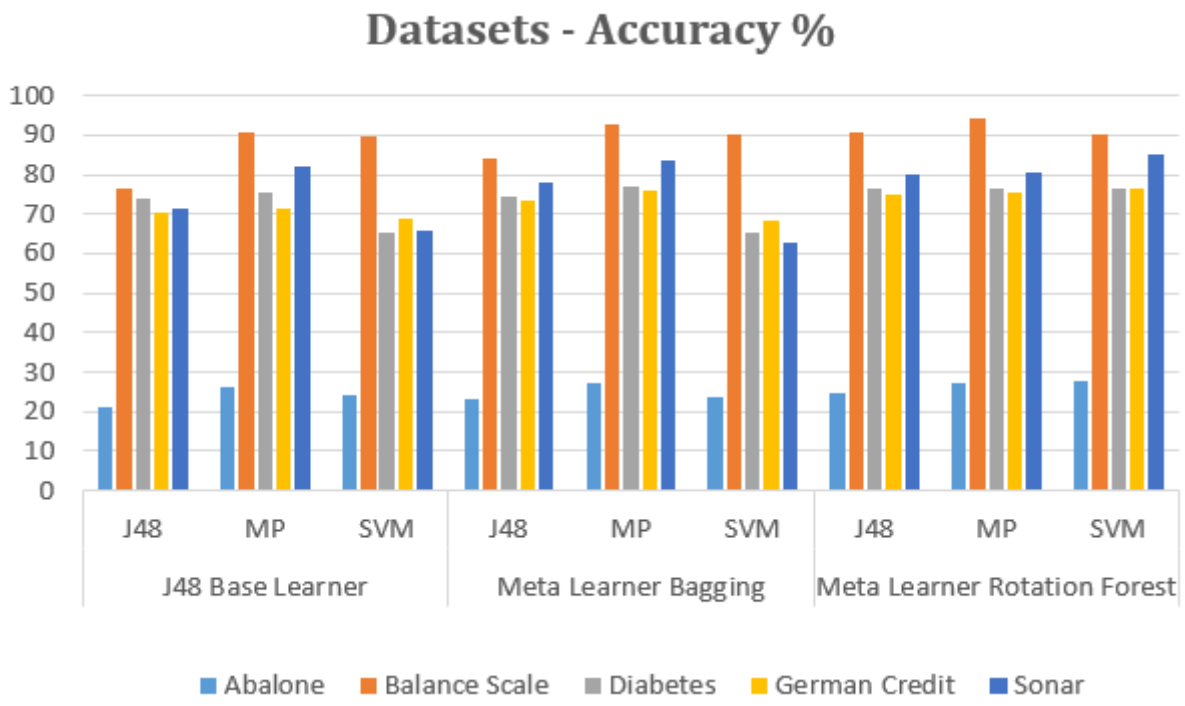


Figure 2. The chart showing the effects between datasets and accuracies

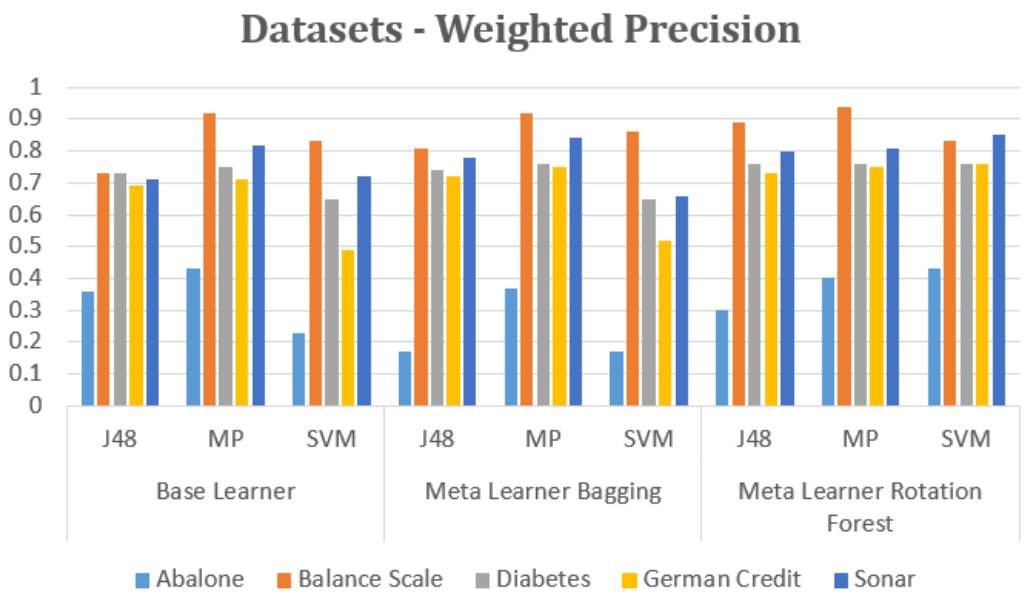


Figure 3. The chart showing the effects between datasets and weighted precision values

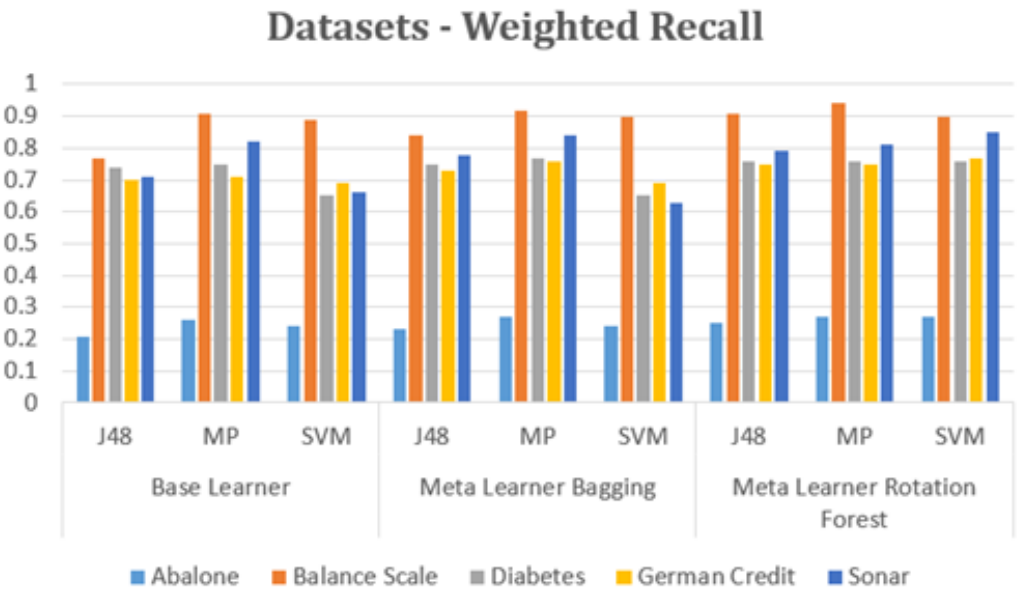


Figure 4. The chart showing the effects between datasets and weighted recall values

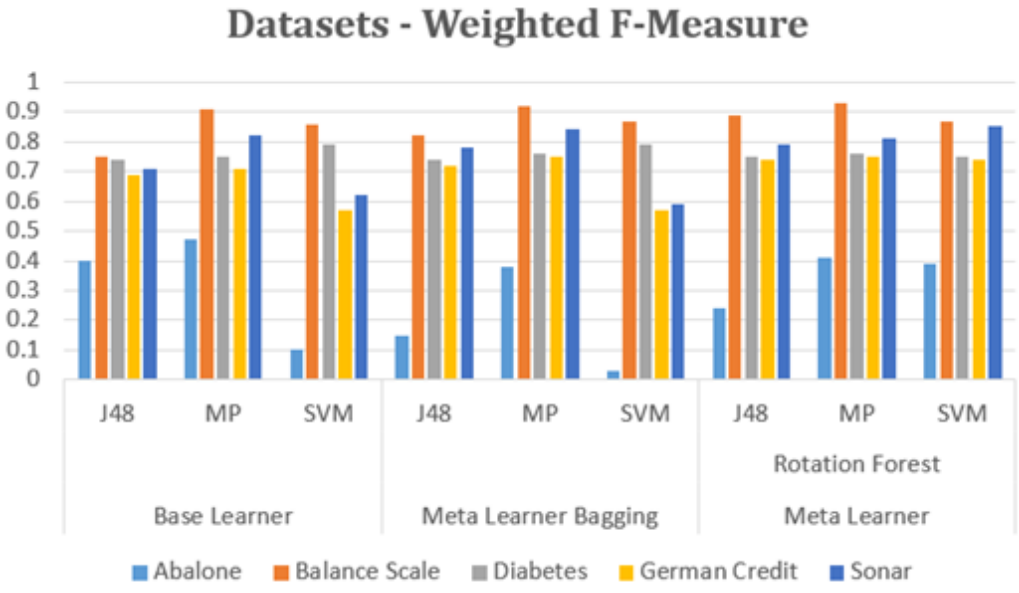


Figure 5. The chart showing the effects between datasets and weighted F-measure

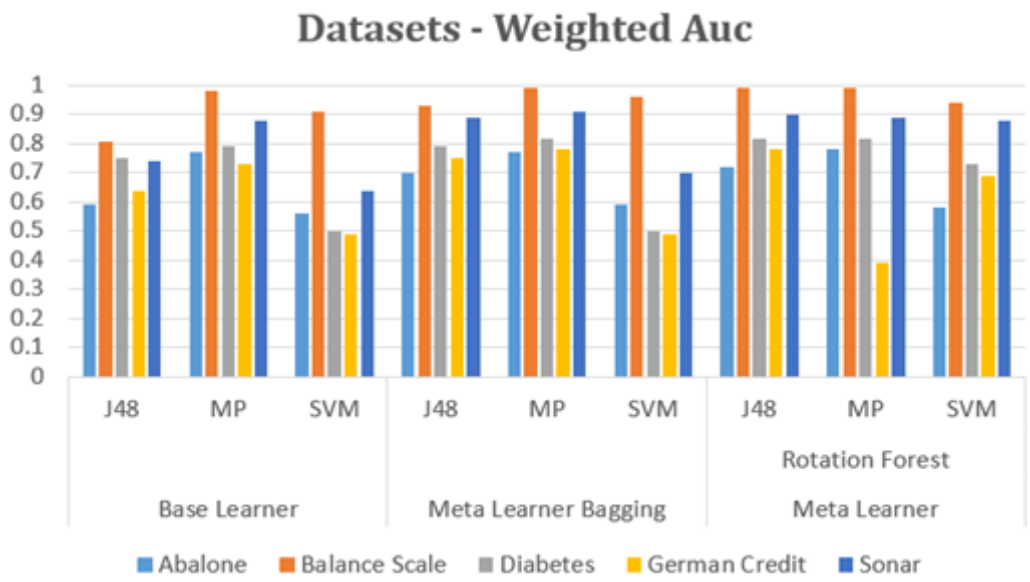


Figure 6. The chart showing the effects between datasets and weighted AUC values

5. Conclusions and future work

This section discusses Base and Meta Learners outcomes and future challenges in the existing Hybrid system. We investigated the various kinds of solutions to relevant problems and analyzed different types of approaches, tools, and techniques, but we couldn't find a single one that could do the entire task at once. Thus, the collaborative approach was proposed to analyze the Hybrid system. This collaborative nature of the proposed system is dependent on two different folds, such as the Base and Meta learner's approach. For the process of data collection, multivariate, categorical, integer and efficient records have been utilized for the Hybrid system. For finding the best results one has to try different methods. We have tried different methods and found the best combination. The results suggest that the use of the feature selection method is advantageous because it reduces complexity and increases accuracy. The performance of J48, MP and SVM with Rotation Forest has been studied using 05 datasets. The main objectives, priority of this proposed system and the key findings of this research work can be summarized as follows, based on the experimental and numerical results:

The Rotation Forest meta-ensemble learning method based on J48 is proposed in this paper. Although Rotation Forest can take more space and consume more time for computations, this method yields more efficient results by using hybrid advantages of base learners' algorithms.

The integration of other hybridization ensemble learning algorithms/approaches and deployment of emerging challenges is the primary focus of our future research.

Author contributions

Abdul Ahad Abro: Original draft and preparation; **Waqas Ahmed Siddique, Mir Sajjad Hussain Talpur, Awais Khan Jumani and Erkan Yaşar:** Reviewed, rewrote, performed part of the literature survey, edited, investigated and designed the architecture and explored software tools.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Urso, A., Fiannaca, A., La Rosa, M., Ravì, V., & Rizzo, R. (2018). Data mining: Classification and prediction. Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, 1, 3, 384-402.
2. Galdi, P., & Tagliaferri, R. (2018). Data mining: accuracy and error measures for classification and prediction. Encyclopedia of Bioinformatics and Computational Biology, 431-436.
3. Ozcift, A., & Gulten, A. (2011). Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. Computer Methods and Programs in Biomedicine, 104(3), 443-451.
4. Panigrahi, R., & Borah, S. (2018). Rank allocation to J48 group of decision tree classifiers using binary and multiclass intrusion detection datasets. Procedia Computer Science, 132, 323-332.
5. Bansal, D., Chhikara, R., Khanna, K., & Gupta, P. (2018). Comparative analysis of various machine learning algorithms for detecting dementia. Procedia Computer Science, 132, 1497-1502.
6. Zhang, C. X., & Zhang, J. S. (2008). RotBoost: A technique for combining Rotation Forest and AdaBoost. Pattern Recognition Letters, 29(10), 1524-1536.
7. Chen, S. F., Gu, H., Tu, M. Y., Zhou, Y. P., & Cui, Y. F. (2018). Robust variable selection based on bagging classification tree for support vector machine in metabonomic data analysis. Journal of Chemometrics, 32(11), e2921.
8. Lamba, R., Gulati, T., Alharbi, H. F., & Jain, A. (2021). A hybrid system for Parkinson's disease diagnosis using machine learning techniques. International Journal of Speech Technology, 1-11.

9. Nandhini, M. (2021). Ensemble human movement sequence prediction model with Apriori based Probability Tree Classifier (APTC) and Bagged J48 on Machine learning. Journal of King Saud University-Computer and Information Sciences, 33(4), 408-416.
10. Pham, B. T., Bui, D. T., Prakash, I., & Dholakia, M. B. (2017). Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. Catena, 149, 52-63.
11. Sun, X., Xu, J., Jiang, C., Feng, J., Chen, S. S., & He, F. (2016). Extreme learning machine for multi-label classification. Entropy, 18(6), 225.
12. Utkin, L. V. (2019). An imprecise extension of SVM-based machine learning models. Neurocomputing, 331, 18-32.
13. Abro, A. A., Taşci, E., & Aybars, U. (2020). A Stacking-based Ensemble Learning Method for Outlier Detection. Balkan Journal of Electrical and Computer Engineering, 8(2), 181-185.
14. Chen, T. (2017). An improved rotation forest algorithm based on heterogeneous classifiers ensemble for classifying gene expression profile. Advances in Modelling and Analysis B, 60(1), 1-24.
15. Khan, A. A., Shaikh, Z. A., Belinskaja, L., Baitenova, L., Vlasova, Y., Gerzelieva Z, Laghari A. A., Abro, A.A., & Barykin, S. (2022). A Blockchain and Metaheuristic-Enabled Distributed Architecture for Smart Agricultural Analysis and Ledger Preservation Solution: A Collaborative Approach. Applied Sciences, 12(3), 1487.
16. Abro, A. A., Khan, A. A., Talpur, M. S. H., & Kayijuka, I. (2021). Machine Learning Classifiers: A Brief Primer. University of Sindh Journal of Information and Communication Technology, 5(2), 63-68.
17. Lu, H., Yang, L., Yan, K., Xue, Y., & Gao, Z. (2017). A cost-sensitive rotation forest algorithm for gene expression data classification. Neurocomputing, 228, 270 -276.
18. Olivares, R., Munoz, R., Soto, R., Crawford, B., Cárdenas, D., Ponce, A., & Taramasco, C. (2020). An optimized brain-based algorithm for classifying Parkinson's disease. Applied Sciences, 10(5), 1827.
19. Khan, A. A., Laghari, A. A., & Awan, S. A. (2021). Machine learning in computer vision: A review. EAI Transactions on Scalable Information Systems, e4.
20. Shuaib, M., Abdulhamid, S. I. M., Adebayo, O. S., Osho, O., Idris, I., Alhassan, J. K., & Rana, N. (2019). Whale optimization algorithm-based email spam feature selection method using rotation forest algorithm for classification. SN Applied Sciences, 1(5), 1-17.
21. Hong, H., Liu, J., Bui, D. T., Pradhan, B., Acharya, T. D., Pham, B. T., ... & Ahmad, B. B. (2018). Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). Catena, 163, 399-413.
22. Abro, A. A., Yimer, M. A., & Bhatti, Z. (2020). Identifying the Machine Learning Techniques for Classification of Target Datasets. Sukkur IBA Journal of Computing and Mathematical Sciences, 4(1), 45-52.
23. Singh, B. K., Verma, K., & Thoke, A. S. (2015). Investigations on impact of feature normalization techniques on classifier's performance in breast tumor classification. International Journal of Computer Applications, 116(19).
24. Abro, A. A. (2021). Vote-Based: Ensemble Approach. Sakarya University Journal of Science, 25(3), 858-866.
25. UCI Machine Learning Repository, 2018, https://archive.ics.uci.edu/ml/index.php.

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Design and implementation of a real-time demonstration setup for dynamic highway tunnel lighting control research studies

Recep Çakmak [*1], **Ayhan Dündar** [2]

[1]*Gümüşhane University, Department of Electrical and Electronics Engineering, Türkiye*
[2]*General Directorate of PTT, Gümüşhane Office, Türkiye*

Keywords
DC/DC Buck converter
LED dimming
Tunnel lighting
Visual perceptions of drivers

Research Article
DOI: 10.31127/tuje.1013374

Received: 22.10.2021
Accepted: 07.12.2021
Published: 01.04.2022

Abstract

This study presents the design and implementation of a computer-based real-time demonstration setup for dynamic highway tunnel lighting control research studies. The energy-efficient and controllable lighting systems stand out worldwide because of energy efficiency objectives and carbon footprint reduction goals of the countries. An LED (Light Emitting Diode) armature-based highway lighting systems are frequently utilized in recent years due to lighting efficiency, energy efficiency, long life, environmentally and cost-friendly structure of the LEDs as well as zero ultraviolet emissions of them. In this study, the LED-based dynamic tunnel lighting control approach has been implemented through the proposed demonstration setup which is a typical real-time software-controlled hardware. The parts of the proposed demonstration setup have been introduced and detailed to help with some upgrade studies in the future. The demonstration setup has been tested in some case studies, then the obtained results have been discussed and introduced. In the future, the proposed system can be upgraded and utilized for advanced control studies thanks to its computer-based and simulation environment connection capability.

1. Introduction

Turkey is a predominantly mountainous country. When looking at the geographic structure of Turkey, it is seen that Turkey is surrounded on three sides by the seas, which are Black Sea, Mediterranean Sea and Aegean Sea. The Mediterranean is in the south, the Aegean is in the west, and the Black Sea is in the north of Turkey. It is also known that mountains form parallel to the sea along the Black Sea and Mediterranean coastlines, and perpendicular to the sea on the Aegean Sea coasts. In addition, there are volcanic mountain formations and mountain formations rising by breaking in the regions in the east of Turkey. These geological features of Turkey create a great obstacle on the roads built for transportation.

Tunnel constructions for highway routes are being more and more important in the world as well as Turkey. Transportation can be accelerated, and optimum protection can be procured for the environment and the landscape through the highway tunnels. Lengths and capacities of the tunnels are increased tremendously day by day, thanks to the emerging tunnel technologies. In long tunnels lighting is an essential issue of the tunnel to provide safely and comfortably driving to the drivers. Emerging carbon economy and energy efficiency targets are becoming a worldwide trend. Therefore, LED (Light Emitting Diode) armature-based tunnel lighting systems have increased progressively in the tunnels due to lighting efficiency, energy efficiency, long life, environmentally and cost friendly structure of the LEDs as well as zero ultraviolet emissions of LEDs [1-4]. However, appropriate lighting is important for drivers' visual perceptions to avoid the black hole effect which reduces visual perception of the drivers. It has been determined that traffic accidents in tunnels frequently experience at the entrances and exits of the tunnels [5-6]. Therefore, the tunnel lighting must always procure maintained visual perceptions for drivers, both day and night, during sudden lighting level changing by their entering and exiting the tunnel. Also, optimum and dynamic lighting control has an important role in energy consumption of the tunnel lighting.

* **Corresponding Author**

*(rcakmak@gumushane.edu.tr) ORCID ID 0000-0002-6467-6240
(ayhan.dundar@ptt.gov.tr) ORCID ID 0000-0002-6707-0420

Cite this article

Çakmak, R. & Dündar, A. (2023). Design and implementation of a real-time demonstration setup for dynamic highway tunnel lighting control research studies. Turkish Journal of Engineering, 7(1), 33-41

Traditional tunnel lighting systems, which are independent of the lighting situation outside the tunnel, are carried out by sodium vapor lamps. These lamps have high power consumption, and their voltage level and light intensity cannot be changed. LED lamps, on the other hand, have both low power consumption and light intensities that can be adjusted by changing the operating voltage [7-8]. Therefore, the LED lamps allow to adjust the illumination levels at the entrance and exit of the tunnels.

The tunnel zones, tunnel lighting criteria and standards have been defined and introduced by [9-10]. Following figure presents the zones in the tunnel and lighting designs of them to avoid encountering the black-hole effect and or bright hole effect and help drivers adapt to the lighting environment in the tunnel when drivers enter the tunnel.

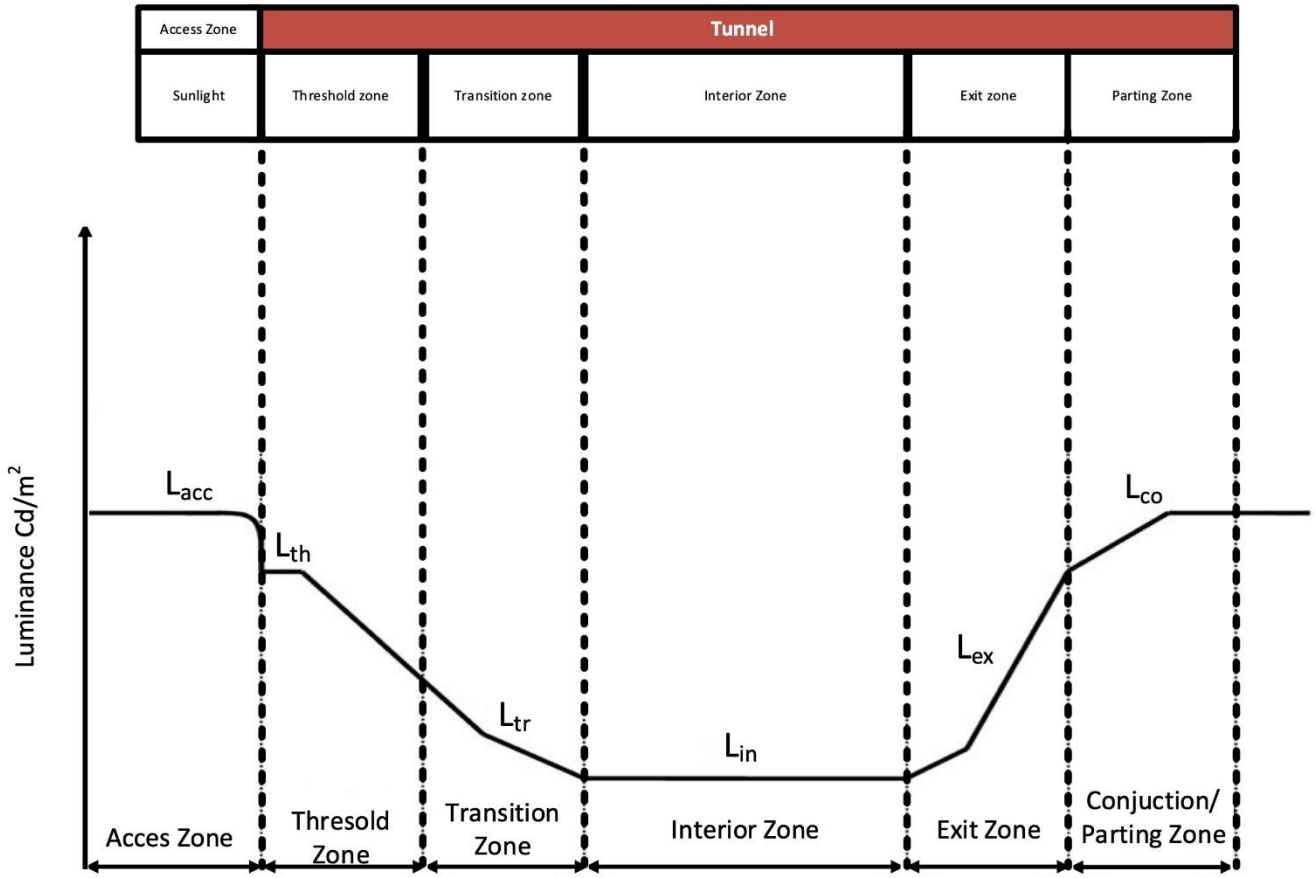


Figure 1. Zones in the tunnel and lighting designs of them. Adapted from [9-10]

There are some of the optimal tunnel lighting and lighting control systems which are proposed by both the literature and tunnel lighting companies. For example, Schréder company proposes intelligent tunnel lighting solutions which help the driver's eyes to adjust easily and quickly [11]. In similarly, the Thorn Lighting company offers tunnel lighting solutions which have fully integrated control systems to provide adaptive lighting system as to level of light outside the tunnel, time of day, speed and density of traffic [12]. An automatic controlled highway tunnel lighting system has been proposed by [13] to meet lighting requirements and to provide energy efficiency. Their proposed automated stepless dimming control strategy provides energy saving 80% more than high pressure sodium lamps, 35 % more than classical rating dimming LED lamps and 20% more than traditional four steps-controlled LED lamps. Zeng et al., [8] have been designed a fuzzy based control algorithm for LED lighted tunnels by utilizing traffic flow and the outside illumination level as the intelligent lighting control parameters. Their proposed fuzzy logic-based LED lighting control system has been utilized in a real highway tunnel and it has been calculated that 87 % decrease in energy consumption for a ten-month period. Xu et al., [14] have been proposed an LED lighting control structure for highway tunnels in order to provide soft and reliable dimming of the LEDs and avoid adaptability problems of the drivers due to brightness changes thanks to adapting the digital addressable lighting interface. They have been utilized step-down power electronics devices to drive the LEDs, and control of the LED driver has been performed by using luminance level sensors and vehicle detectors in the control system. Wang et al., [15] have been asserted a dynamic dimming control system for LED lighting of the tunnels through a demonstration project. They have been reported energy saving the advantages of dynamic dimming-controlled LED based tunnel lighting over the high voltage sodium lamp and constant luminance LED lighting. Qin et al., [16] have been designed an intelligent luminance control method for tunnel lighting by considering the traffic flow in the tunnel. Their proposed system has two operational

strategies which called as day and night operations. In the day operation, the luminance of the tunnel's zones is controlled as to the luminance of the tunnel's entrance and exit. In the night operation, the luminance is increased when the vehicle is entering the tunnel, it decreased when the vehicle is exiting the tunnel through the proposed strategy.

In this study, an LED (Light Emitting Diode) based dynamic tunnel lighting approach has been designed and implemented through devising real time software-controlled hardware by considering the above-mentioned literature. Although utilizing of LED lamps provides incontrovertible energy efficiency due to their low energy consumption [7,8,13], this study focuses on to increase safety driving process of the drivers in a highway tunnel by the proposed LED lighting control system. Thus, it is aimed to provide a dynamic LED dimming mechanism which increases the visual adaptation of the drivers by controlling the light level of the transition areas in the tunnel when the light level in the outdoor environment changes. In order to provide a demonstration and to make a design and development platform for control of LED dimming-based tunnel lighting systems, a real-time simulation and PC controlled hardware implementation system have been devised on an emulated tunnel prototype. The remainder of the paper is organized as follows: Section 2 focuses on material and method of the study; Section 3 presents the results and Section 4 gives the conclusion.

2. Material and Method

The emulated prototype tunnel dimensions have been determined by scaling the entrance, exit, interior and transition regions of the tunnels, considering the above-mentioned literature on tunnel lighting systems. The sensor circuits, which sense the brightness of input and output of the tunnel, and the LED driver circuits and have been designed and data connection with them to the development card have been performed. Arduino Mega 2560 has been utilized as the development card. Since the brightness of the LEDs is directly proportional to their voltage levels, the brightness control has been performed by DC/DC step-down converter and it has been controlled via designed Proportional-Integral (PI) controller. In order to drive the LEDs, the Pulse Width Modulation (PWM) signal which is adjusted by the controller output has been used. The real-time operation of the system has been carried out in MATLAB/Simulink simulation environment. Therefore, after the simulation model has been tuned up and the errors have been fixed the simulation model has been transferred to the Arduino development card through PC communication.

The operation logic of the system can be described simply by following equation:

$$O(t) = \left[K_p e(t) + K_i \int_0^t e(t)dt\right] x L_c \qquad (1)$$

where,

O(t): Operation signal which is sent to the LED driver.
K_p: Proportional coefficient of the PI controller.
K_i: Integral coefficient of the PI controller.
e(t): The difference between reference lighting level and measured lighting level.
L_c: Lighting coefficient which provide different lighting in the zones of tunnel.

In this study, the controller parameters Kp and Ki are tuned through MATLAB Simulink Control System Toolbox. The Kp and Ki are used as 0.5 and 0.005 respectively. The saturation blocks which are used in the output of the PI controller limits the data between 0 and 255 due to PWM blocks of the Arduino operate in 8-bit data.

Following subchapters introduce the parts of the proposed system and devised platform, and their designs.

2.1. Real-time simulation of the system in the MATLAB/Simulink environment

Fig. 2 shows the real-time simulation blocks of the system in the MATLAB/Simulink environment. External (out of the tunnel) illumination data have been read and transferred to the control unit through the light sensors located in the entrance and exterior zones of the tunnel structure. These data have been utilized in compliance with the above-mentioned tunnel lighting essentials. The DC/DC step-down converters are located at the outputs of the control card both acts as a driver for adjusting the light intensity and transfers the necessary information to the feedback circuit in order to control the desired light level. In this system, day and night transition and its application has been performed via the algorithm which changes the operational logic of the system in day and night conditions. Lighting levels outside the tunnel are read by the sensor and switch to night mode when the illumination level decreased to the night level. So, the system continues operating in a way that can be described as the opposite of the day operation.

2.2. Light sensor circuit

The light sensor circuit is shown in Fig. 3. Its structure is like a voltage divider with LDR. The illuminance levels at the entrance and exits of the tunnel are measured through this light sensor circuit.

2.3. DC/DC step-down converter circuit

A DC/DC buck converter, in other words DC/DC step-down converter, is a power electronic device which reduces voltage from its input (supply) to its output (load).

Fig. 4 shows the circuit of the traditional DC/DC step-down converter. The DC/DC step-down converter provides the desired voltage level by reducing the mean value of the input voltage via chopping. The input voltage of the converter is chopped through the pulse width modulation (PWM) signals, which are provided by the PI controller output in this study.

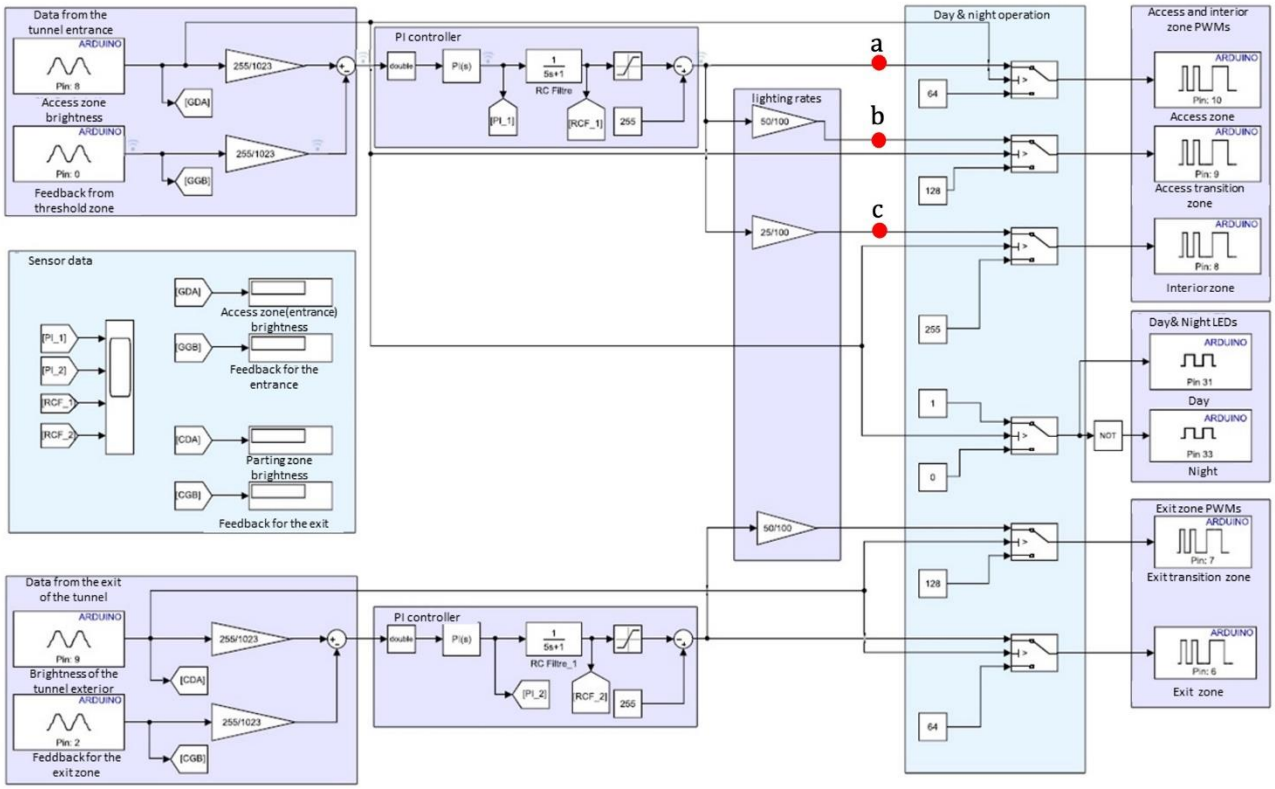


Figure 2. Zones in the tunnel and lighting designs of them

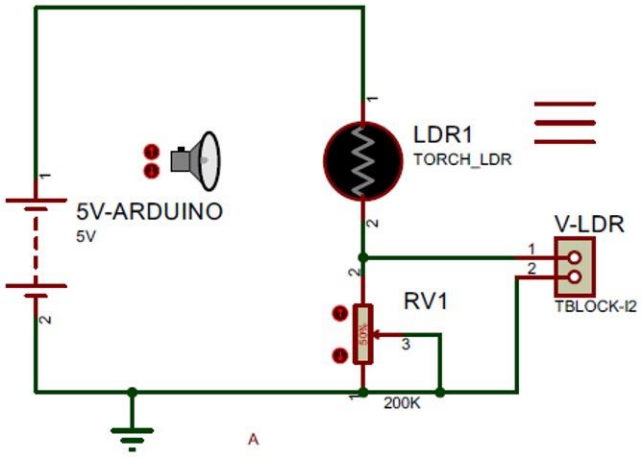


Figure 3. Light sensor circuit

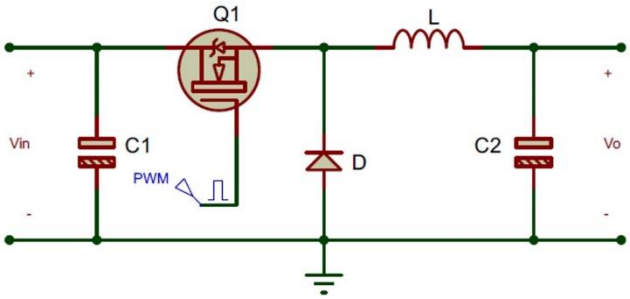


Figure 4. Traditional DC/DC step-down converter circuit diagram

An average voltage is obtained from the output of the DC/DC step-down converter as to the on or off state of the Q1 switch which is given in Fig. 4. As it can be seen in Eq. (2), the average voltage depends on the duty cycle (D) which determine the on-off state of the Q1 switch.

$$D = \frac{Vo}{Vin} \text{ and } 0 \leq D \leq 1 \quad (2)$$

In Eq. (2), Vo is the output voltage of the converter and Vin is the input voltage of the converter. The sum of the on and off times of the Q1 switch is called the period which is denoted by *Ts.* It is presented in Fig. 5.

The graph in Fig. 5 shows the voltage between the terminals of the Q1 switch. Accordingly, the switch Q1 is open during the t_{on} time, and the switch is closed during the time indicated by t_{off}. So, the mean value of the output voltage will be changed by duty cycle (D) as it presented in Eq. (3).

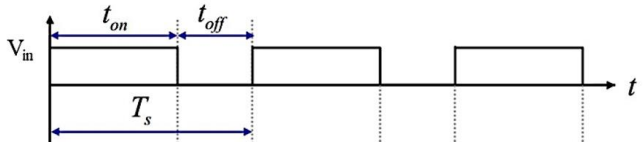


Figure 5. Chopping the input voltage as to duty cycle which determine on and off time of the Q1 switch.

$$Vo = D.Vin \quad (3)$$

The role of the duty cycle and PWM signals in this proposed system can be explained as follows. When the duty cycle is changed by the controller, the PWM signals change, then the voltage of the LEDs changes.

The following figure shows that the response of the controller and PWM signals when the error (the brightness difference between inside and outside of the tunnel) is changed.

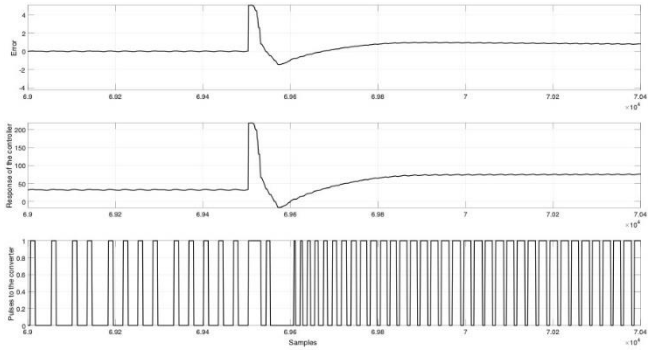


Figure 6. The response of the controller and PWM signals when the error is changed.

Fig. 7 shows the DC/DC step-down converter circuits used in this study. Due to its structure, this circuit allows the voltage in the range of 5 V to 35 V given to its input to be adjusted between 0 V and input voltage according to the duty cycle sent by the control unit. Module LEDs operating with 12 V are used in the circuit in the tunnel lighting system that was applied. For this reason, LED lighting has been controlled to give an output between 0 V and 12 V by applying 12 V voltage to the DC/DC step-down converter input.

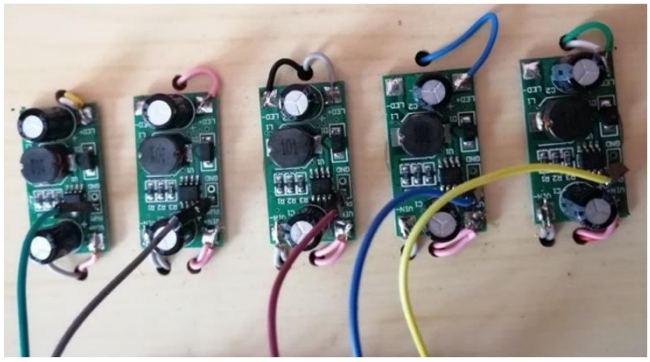


Figure 7. DC/DC step-down converter circuits used in this study.

2.4. The feedback circuit

The feedback circuit is given in Fig. 8. This compares illumination of the entrance and exit transition zones with the exteriors of the tunnel. It is basically a voltage divider made with a resistor and a potentiometer. In addition, a voltage follower circuit with LM358 OPAMP is designed and in order to provide the system more stable and to create a high impedance at the output of the voltage divider. This circuit is placed to end of the voltage divider. The voltage follower circuit is presented in Fig. 8.

In Fig. 9, the printed circuit board (PCB) layout diagram and 3D model of the feedback circuit which includes the voltage divider and voltage follower.

2.5. PWM driver and signal inversion circuit

Fig. 10 shows the PWM driver, and the signal inversion circuit used in this study. This circuit basically consists of five transistor switch circuits. The signal inversion circuit creates a 180° phase difference in the signal which comes from Arduino MEGA 2560 control card. The boughten ready to use DC/DC step-down converter, which is shown in Fig. 7, operates inversely proportional to the duty cycle because of its production nature. So, this signal inversion has been utilized to ensure operation of the DC/DC converter as it is introduced in Eq. (2). Fig. 11 shows printed circuit board (PCB) layout diagram and 3D model of the designed and implemented signal inversion circuit.

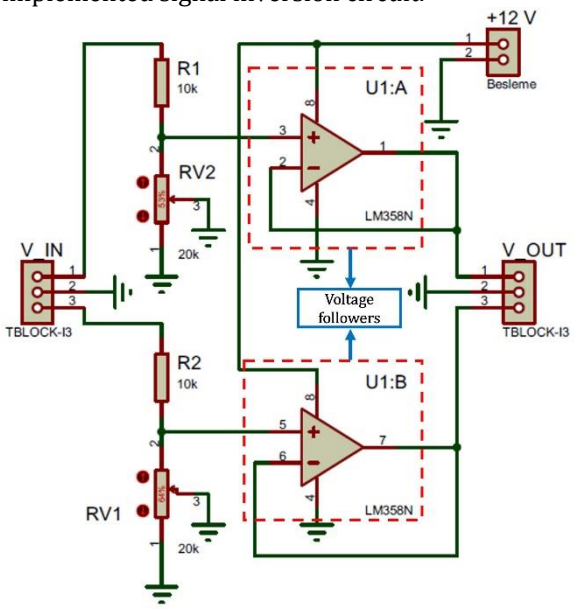


Figure 8. The feedback circuit with voltage follower circuits.

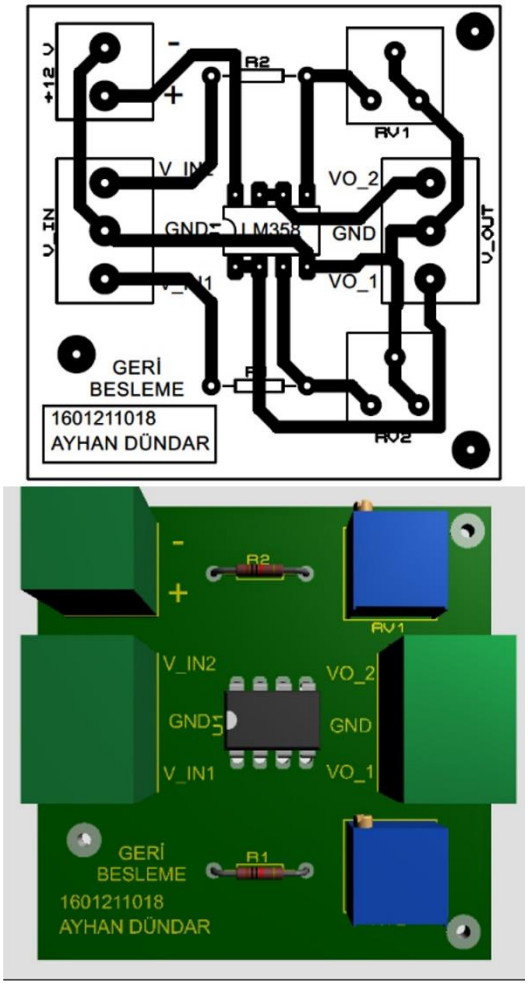


Figure 9. PCB layout diagram and 3D model of the feedback circuit

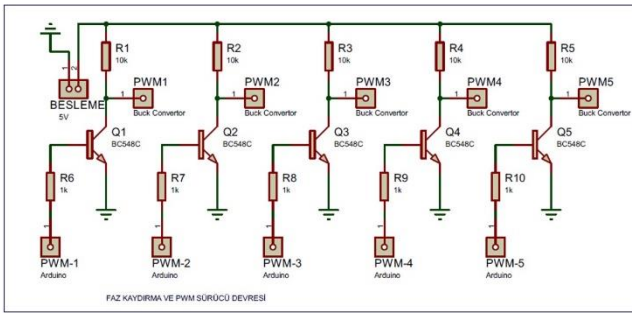


Figure 10. PWM driver and signal inversion circuit.

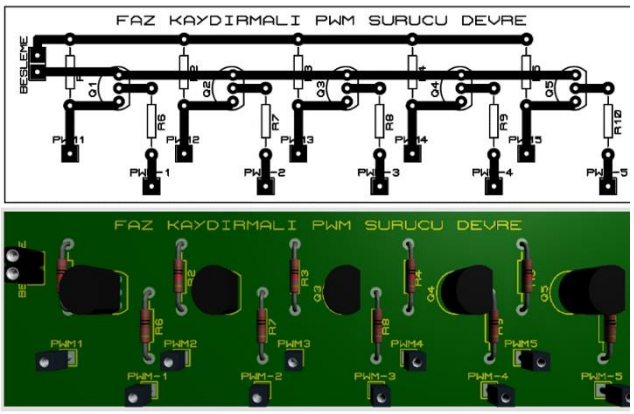


Figure 11. PCB layout diagram and 3D model of the signal inversion circuit

2.6. Arduino MEGA 2560 development card

In this study, Arduino MEGA 2560 development card has been used to acquire and process of the sensor data. These data are utilized in the devised algorithm, which is embedded to Arduino MEGA 2560 via MATLAB/Simulink, to control the illumination of the LEDs thanks to the DC/DC step-down converter and its driver. Arduino MEGA 2560 is an open-source microcontroller board based on ATmega2560 processor. The operating speed of the card is 16 MHz. It has 54 digital input/output ports, 15 of which can be used as PWM output ports. The Mega2560 has 16 analog inputs with 10-bit resolution each. They normally operate in the 0-5 V range. The used Arduino MEGA 2560 is shown in Fig. 12.

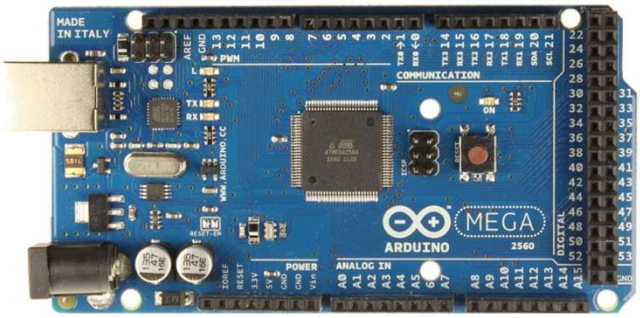


Figure 12. Arduino MEGA 2560 development card

2.7. Devised real time PC controlled demonstration platform

The hardware of the devised real time PC-controlled demonstration platform to develop and to test dynamic LED dimming-based highway tunnel lighting control system is shown in the Fig. 13. The parts of the platform which are labeled by numbers are introduced in follows.

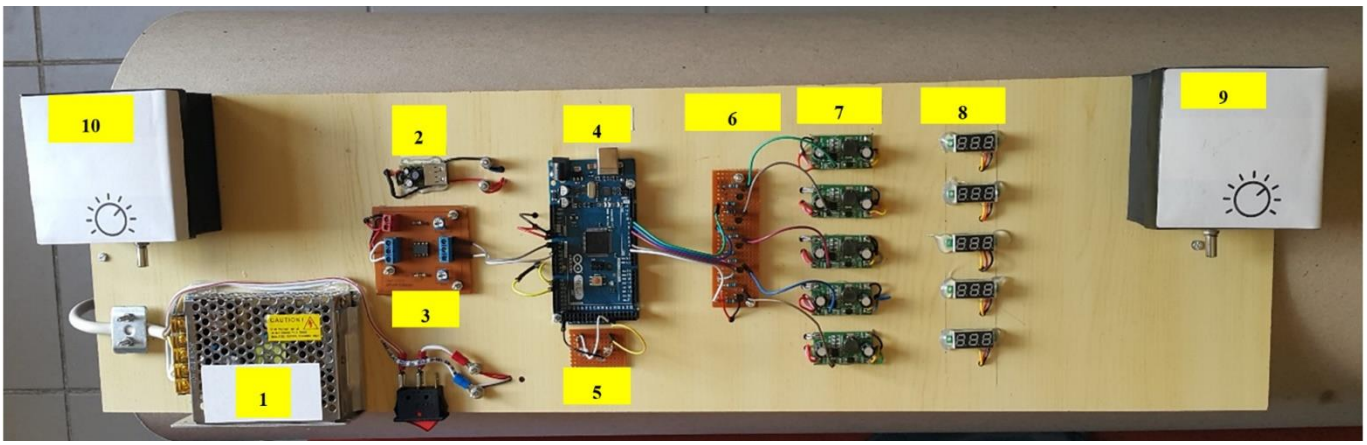


Figure 13. Devised Real Time PC Controlled Demonstration Platform

1: Power Supply (12V, 60W)
2: 5V 3A DC/DC step-down voltage regulator which is used to power supply the Arduino card
3: The feedback circuit
4: Arduino MEGA 2560
5: Sensor to sense day and night
6: PWM driver and signal inversion circuit
7: DC/DC step-down converters which drive the LEDs in the tunnel
8: DC voltmeters shows the voltage of the LEDs located in the different zones of the tunnel
9: Illumination sensor to sense brightness in exit exterior zone of the tunnel
10: Illumination sensor to sense brightness in entrance exterior zone of the tunnel

2.8. Flowchart of the simulation

The flowchart of the simulation which is uploaded to the Arduino MEGA2560 card is shown in Fig. 14. The operation mode of the system is controlled by "Is less than 750?" condition statement which is given in the flowchart of the simulation. The 750 value is a threshold value, and it determines the operation condition as night or day. The value which is determined as 750 has been specified experimentally by trial-error approach such that the value represents the night conditions in the experimental setup.

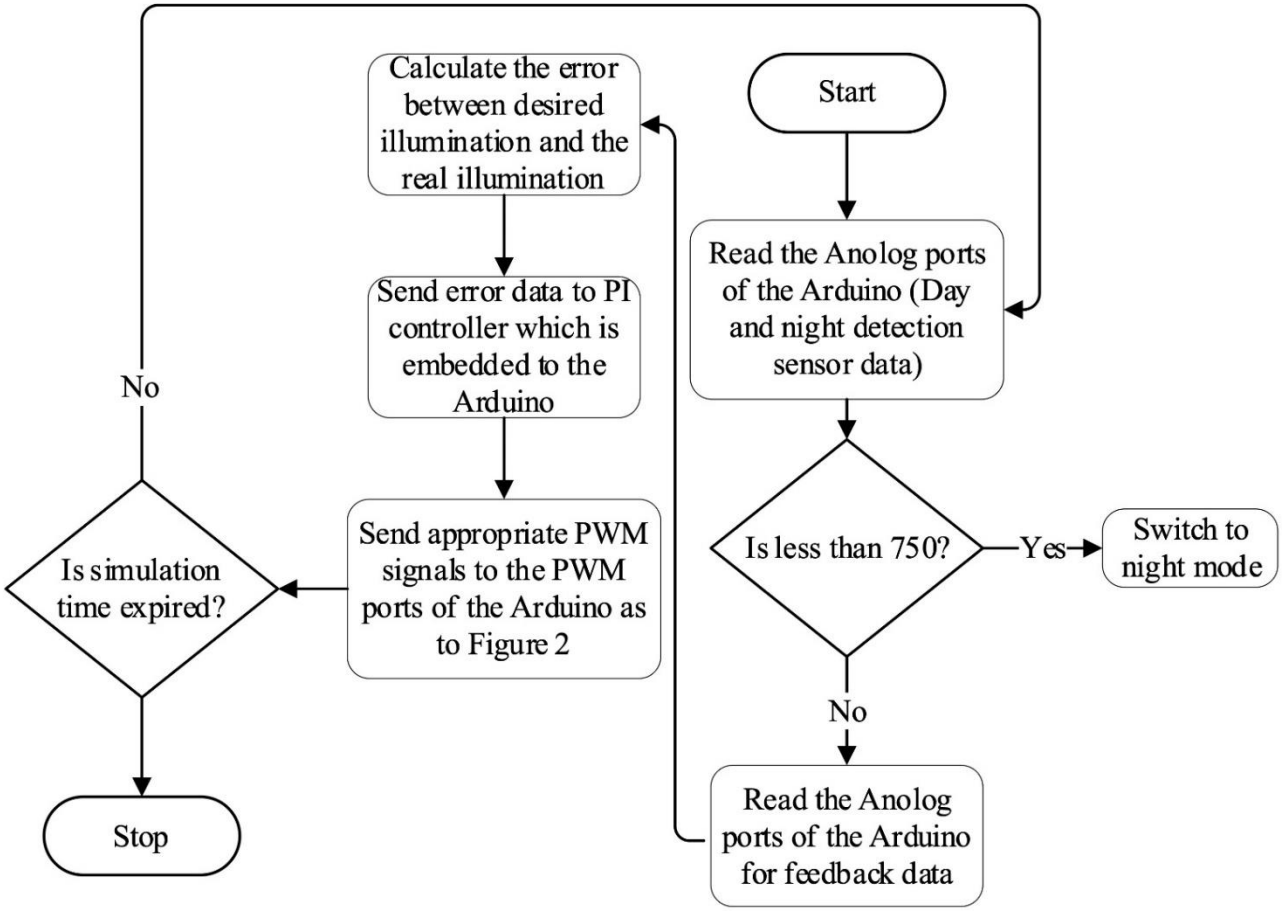


Figure 14. Flowchart of the simulation

3. Results

Some of obtaining the data set through in the case study are presented in Table 1. The PI output is converted to 8-bit data, which is the operating range of Arduino PWM ports as shown in Fig.2. In the system, the lighting levels of the access zone, the transition zone and the interior zone are controlled by PWM-10, PWM-9 and PWM-8, respectively. Since the PWM-10 value should be the same as the outdoor lighting value, the PI output is directly connected to this port and the required value of feedback is taken from this port.

The illumination level of the transition zone has been determined as 50% of the access zone. So, 50% of the PI output value is sent to PWM-9 port. The illumination level of the interior zone has been determined as 25% of the access zone. So, 25% of the PI output value is sent to PWM-8 port. Although it is not shown in Table 1, illumination control of the tunnel's exit zone has been controlled in the same manner as above rates.

As it can be seen in Table 1, A8 is the analog input port of the Arduino, and the actual light value is read from this port. When the table is carefully examined, it is shown that the value 1023 is the 10-bit input value of the Arduino and represents a maximum voltage of 5 V. When the value of the A8 port decreases, A0 feedback port is constantly brought to the A8 value via the PI controller. In addition, it is seen that the values decrease in the input of PWM blocks due to the decreasing amount of light. This shows that the amount of light emitted by the LEDs in the tunnel is reduced by the DC/DC step-down converter, which is connected to the PWM outputs. The results of the above-mentioned operation are also can be seen in the sixth column of the Table 1 which presents the voltage (Volt) of the led whose driver is connected to the PWM 10 port.

Thanks to the day and night transition algorithm, when the A8 input value falls below 750, it is seen that LED lightings in the tunnel increase from the tunnel entrance to the interior area. Then they decrease the light level from the middle area to the exit.

Fig. 15 presents the PI controller output changes in the case study. Since the light level, which are read from A8 Port is at the maximum value at the beginning the simulation, PI output is equal to zero. As it can be seen clearly from the Fig. 15, the decreasing of light while over time during the simulation cause changes in the output of the PI controller.

Table 1. Some of obtained data set through a case study

Illumination sensor data in entrance exterior zone		**Feedback**	**Error**	**PI Output**	**Input of PWM blocks which drive the LEDs. (The measured points are given in the Fig. 2 as a, b and c)**			**The voltage (Volt) of the led whose driver is connected to the PWM 10 port**	**Interpreted lighting level for Port (A8)**
Port (A8)		**Port (A0)**			**PWM-10**	**PWM-9**	**PWM-8**		
Data as 10 Bit	Volt	Data as 10 Bit							
1023	5	1023	0	0,000021	255	127.5	63.75	12	Max
1000	4.89	996	4	164.4	91.81	4587	22.92	11.73	
990	4.84	988	2	170.7	84.77	42.11	20.77	11.61	High
980	4.79	978	2	172.7	79.44	40.08	20	11.50	
970	4.74	969	1	178.9	75.53	38.31	18.91	11.38	
940	4.59	939	1	182.9	72.13	35.81	17.99	11.03	
910	4.45	908	2	187.3	68.12	33.98	16.67	10.67	Medium
890	4.35	887	3	190.6	64.97	32.44	15.92	10.44	
870	4.25	869	1	193.8	61.99	30.91	15.06	10.21	
840	4.11	839	1	200.2	54.67	27.71	13.98	9.85	
810	3.96	808	2	210.8	44.87	22.75	11.32	9.50	Low
790	3.86	787	3	222.5	34.63	16.12	8.31	9.27	
760	3.71	758	2	239.55	16.34	8.68	4.14	8.91	
750	3.67	Transition zone for night and day (Night < 750 < Day)							
<750		Constant light level for night conditions			64	128	255	10.28	Medium

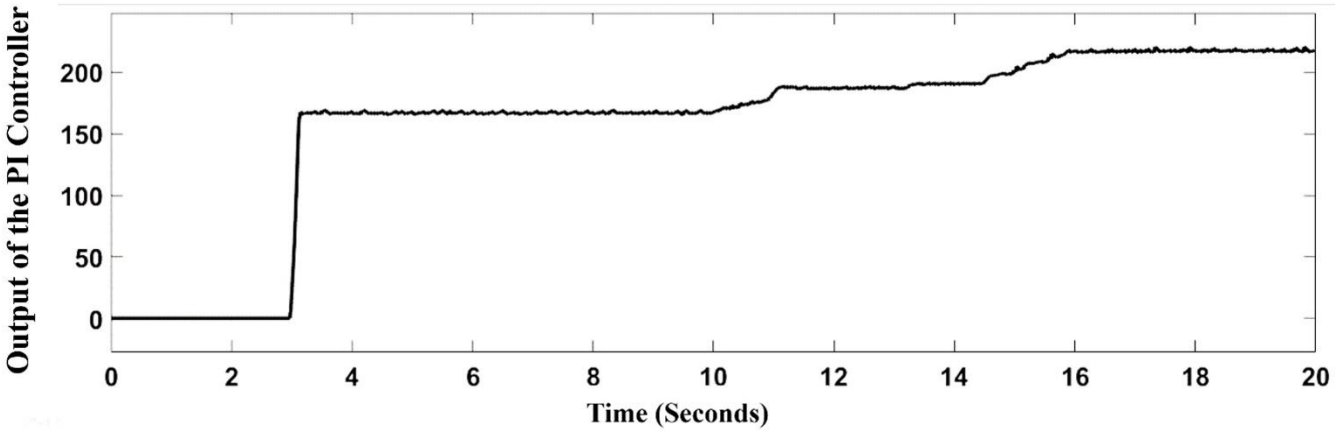


Figure 15. Changes of PI output in the case study

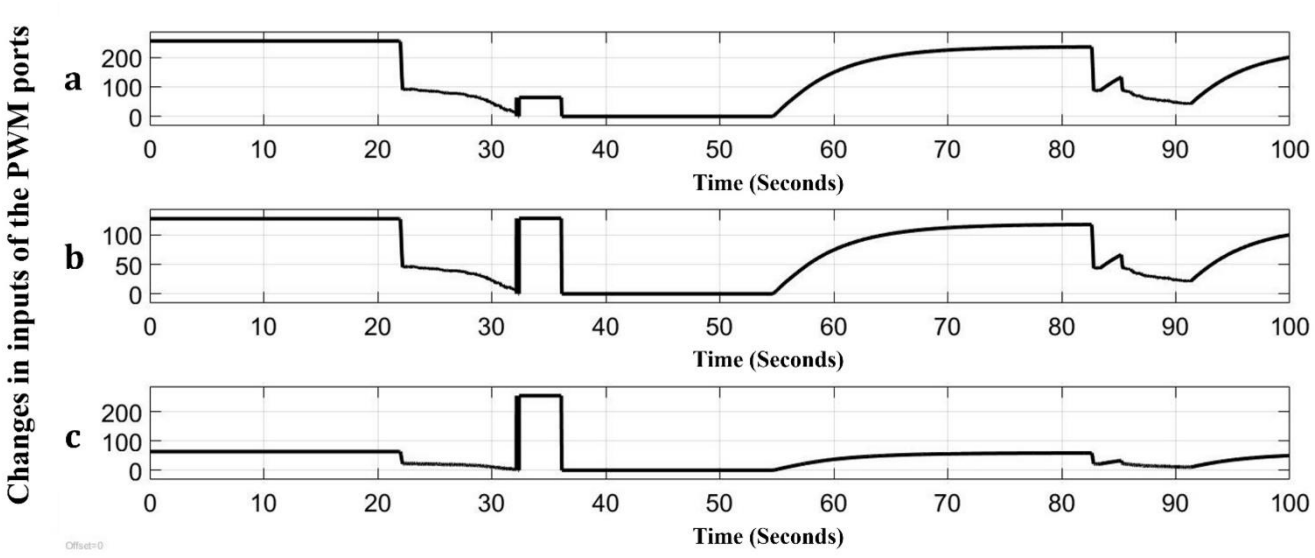


Figure 16. Changes in inputs of the PWM ports in the case study

In Fig. 16, changes in inputs of the PWM ports in the case study are presented. These data have been logged for following scenario. The platform has been tested at the maximum amount of external illumination about 22 seconds. Then the external illumination is decreased, and it has been observed that there is a decrease in inputs of the PWM ports. At the 32nd seconds, the system has been switched to the night mode, and it has been remained in the night mode for about 5 seconds. At the 37th seconds, the system has been switched to the day mode, and the external illumination has been increased to maximum outdoor lighting level in approximately 70 seconds. Between the 80th and 100th seconds, the system has been simulated as if a cloudy weather had occurred such that the external illumination level is undulated.

It can be noticed that from the Fig. 16, when the input of the PWM-10 value takes the maximum value of 255, the input of PWM-9 value takes the maximum value of 128 and the input of PWM-8 value takes the maximum value of 64. The case results prove that the illumination level decreases dynamically from the entrance of the tunnel to the interior zone according to the external lighting level changes, as it proposed in this study.

4. Conclusion

In this study, an LED (Light Emitting Diode) based dynamic tunnel lighting approach has been designed and implemented through devising real time software-controlled hardware. The parts of the proposed platform have been introduced and detailed in order to help with some upgrade studies in the future. The devised system has been tested in some case studies to see the effectiveness of it. The case study results have been discussed and introduced. The results show that the system is performed properly. The devised system can be upgraded and utilized for advanced control studies thanks to its PC and simulation environment connection capability. In the future, it is planned that to develop an advanced controller and to test them through this platform. Also, this platform can be used for tunnel lighting and control practices in some courses in the Electrical and Electronics Engineering Departments.

Acknowledgement

This paper is a revised and extended version of the presented abstract paper at the IOCENS21 (International Online Conference on Engineering and Natural Sciences).

Author contributions

Recep Çakmak: Conceptualization, Methodology, Visualization, Simulation, Investigation, Writing-Original draft, Writing-Reviewing and Editing. **Ayhan Dündar:** Simulation, Hardware Implementation and Investigation.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Perdahçı, C. A, Durak, M, Kılıç, Y. & Altun, B. (2013). Tünel Aydınlatma Sistemlerinde Led Teknolojisi. 3e Electrotech Dergisi, Ekim 2013.
2. Rüstemli, S. & Avcil, S. (2018). Tünel Aydınlatmasında LED Armatür Kullanımı. Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 23(2), 168-181.
3. Mao, B., Niu, P. & Huang, C. (2008). The design of the drive control chip for the solar LED lighting system. Mod Applied Science, 2, 75-80.
4. Moretti, L., Cantisani, G., & Di Mascio, P. (2016). Management of road tunnels: Construction, maintenance and lighting costs. Tunnelling and Underground Space Technology, 51, 84-89.
5. Yeung, J. S. & Wong, Y. D. (2013). Road traffic accidents in Singapore expressway tunnels. Tunnelling and Underground Space Technology, 38, 534-541.
6. Bassan, S. (2016). Overview of traffic safety aspects and design in road tunnels. IATSS research, 40(1), 35-46.
7. Wu, M. S., Huang, H. H., Huang, B. J., Tang, C. W. & Cheng, C. W. (2009). Economic feasibility of solar-powered led roadway lighting. Renewable energy, 34(8), 1934-1938.
8. Zeng, H., Qiu, J., Shen, X., Dai,, G., Liu, P. & Le, S. (2011). Fuzzy control of LED tunnel lighting and energy conservation. Tsinghua Science & Technology, 16(6), 576-582.
9. Liu, H. Y. (2005). Design criteria for tunnel lighting. World 2005 Long Tunnels, 363,372.
10. Commission internationale de l'éclairage (CIE) (1990). Guide for the lighting of road tunnels and underpasses. CIE.
11. Schréder, Schréder Tunnel Lighting Solutions. (2021), Retrieved from http://www.tunnelsonline.info/Uploads/storefront/adverts/ITD151207wpb.pdf
12. Thorn, Thorn Tunnel Lighting. (2021). Retrieved from http://www.thornlighting.com/download/TunnelINT.pdf
13. Fan, S., Yang, C., & Wang, Z. (2010, October). Automatic control system for highway tunnel lighting. In International Conference on Computer and Computing Technologies in Agriculture (pp. 116-123). Springer, Berlin, Heidelberg.
14. Xu, L. H., Zhou, Y. & Huang, Y. G. (2014). Design of Highway Tunnel LED Lighting Control System. In CICTP 2014: Safe, Smart, and Sustainable Multimodal Transportation Systems (pp. 406-412).
15. Wang, D., Jiang, H., Ma, J. & Zheng, X. (2012, June). Dynamic Dimming Control Method Research on Tunnel LED Lighting Based on LED Controllability. In 2012 2nd International Conference on Remote Sensing, Environment and Transportation Engineering (pp. 1-4). IEEE.
16. Qin, L., Dong, L. L, Xu, W. H., Zhang, L. D. & Leon, A. S. (2017). An intelligent luminance control method for tunnel lighting based on traffic volume. Sustainability, 9(12), 2208.

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Tribological properties of MoS_2 particles as lubricant additive on the performance of statically loaded radial journal bearings

Hasan Baş[*1]

[1]*Karadeniz Technical University, Department of Mechanical Engineering, Türkiye*

Keywords
Additive
Friction
Lubrication
Journal bearing
Molybdenum disulfide
MoS_2

Research Article
DOI: 10.31127/tuje.1016153

Received: 29.10.2021
Accepted: 30.01.2022
Published: 01.04.2022

Abstract

It is very important to reduce power losses for plain bearings used in industry and especially in automobiles. In recent years, inorganic compounds have been added to the engine oils to reduce friction in plain bearings and increase their performance. In this study, the effects of additive oil obtained by adding 1% by weight Molybdenum Disulfide (MoS_2) to the base hydraulic (or lubricating) oil (Shell Tellus 10) on the statically loaded radial journal bearing performance were experimentally investigated. As a result of experiments, it was determined that MoS_2 added engine oil showed less friction behavior by forming a better oil film compared to the base oil at increasing bearing load and temperatures, and therefore exhibited less wear and power loss.

1. Introduction

Hydrodynamic bearings are used in high-speed rotary device [1]. The radial bearings, which carry load in the radial direction and work on the principle of the hydrodynamic lubrication, are one of the common types of the bearings used in industry. They are especially used in environments where cushioning of vibrations and resistance and silence to vibratory forces are essential. Journal bearings have a great effect on reducing power losses in mechanical systems. The friction factor and thus the power loss is a function of viscosity, load, and speed. There was a great deal of research at that time trying to find the best combinations of materials and lubricants that would give the lowest coefficient of friction.

The use of suitable lubricants to reduce the friction and wear of mechanical contacts is very important for power loss [2]. To reduce maintenance period and power loss of rotary machine, some of additive oils have been added to base mineral oil in various industries. To improve lubricant oil properties, additives have been used. These contain antiwear additives, extreme pressure additives, viscosity control additives, film-forming additives, and sediment control additives [3]. Recently, attention has been directed towards nanoparticles since they have unique properties when compared to their majority equivalents.

Studies in the use of nanoparticles as additives have shown to reduce friction and wear [4,5]. The choice of appropriate lubricant has a special effect on the performance of the bearings. The major losses that immerge in a car engine can be friction between moving parts. The main role of the lubricant is to retain the surface of the two metals dampish, that separates them from each other by creating a suitable layer on the surfaces with friction, and the heat and created abrasive particles will be annihilated [6]. A good lubricant should have a high flash point, low pour point and high viscosity index (VI). Nano-additive oil-based lubricants expose wonderful lubrication performance for tribological applications [7,8]. Due to their proper lubrication properties, the use of MoS_2 nanoparticles as solid lubricants has been highly regarded, recently. MoS_2 nanoparticles has a hexagonal crystalline structure.

* **Corresponding Author**
*(bas@ktu.edu.tr) ORCID ID 0000-0002-5653-3813

Cite this article
Baş, H. (2023). Tribological properties of MoS_2 particles as lubricant additive on the performance of statically loaded radial journal bearings. Turkish Journal of Engineering, 7(1), 42-48

The self-lubricity properties are related to the large area between the S-Mo-S sandwich layers and the weak Van der Waals forces and the pure positive charge at the surface, resulting in propagation of electrostatic repulsion. The layers exist with weak molecular forces and can easily slide over each other, [6, 9-12]. The effect of MoS_2 nano-additives on the frictional properties of the lubricating oil was studied by Rajendhran et al [13]. The results divulged that 0.5 wt% nano-additives added to the pure oil could distinctly improve the tribological properties. Numerous studies have been carried out to investigate the tribological properties of MoS_2 nano-doped oil-based lubricants. The results demonstrated that the lubricant characteristics of the prepared nano-fluids is attractively improved with the combine of MoS_2 nanoparticles [12, 14-19].

There are three lubrication regimes for the tribology of the journal bearing, which are generally defined as boundary, mixed and full film lubrication, firstly obtained by the McKee brothers [20]. In this graph, there is usually a friction coefficient on the vertical axis, and only velocity or an ηn/P dimensionless expression can be on the horizontal axis.

The variation of the "friction coefficient [μ]" in the journal bearing versus the "bearing parameter ηn/P " is shown in Fig. 1. Where n is the speed of the shaft (rev./s), p is the nominal pressure of the bearing (N/m^2), and η is the dynamic viscosity of the lubricant ($N.s/m^2$).

The friction coefficient, the viscosity of the lubricant used in the bearing, the strain applied to the system by the external load carried by the bearing, and the relative speed difference between the sliding speeds of the surfaces can instantly be seen in the graph. With the help of this graphic, practical information can be obtained about whether the bearing works in the critical operating region, that is, in the region close to metal-metal contact.

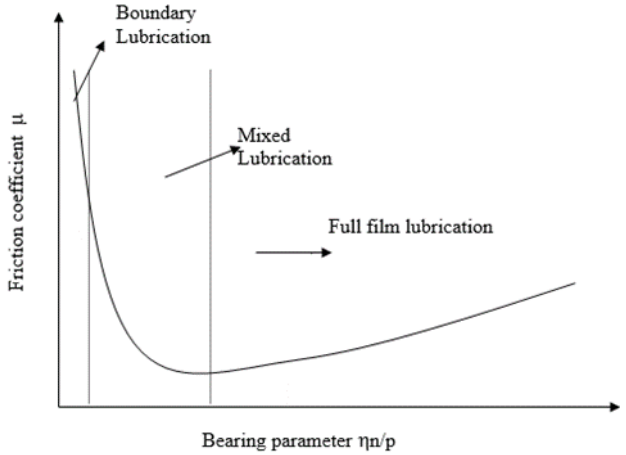


Figure 1. A Schematic Stribeck curves showing three lubrication regimes

Since the minimum film thickness is generally dependent on surface roughness, bed load and relative velocity, an acceptable minimum value is not shown. However, in order to avoid metal-to-metal contact and to keep friction losses at a minimum level, %20 increase in the sum of the surface roughness can be determined as the minimum value. In this study, a measurement system was used that shows the metal-to-metal contact between the bearing and the shaft during the tests. This study experimentally compares the effects of MoS_2 on the base engine oil (Shell Tellus 10) in the statically loaded radial journal bearing. In addition, the effect of the MoS_2 concentration particle's on the tribological behavior of lubricants is explored.

To determine the friction behavior, the bearing system was modified and constructed. The structure of the particles was characterized by scanning electron microscopy (SEM).

2. Method

To investigate the effects of additive engine oil on bearing performance, a journal bearing test rig designed specifically for this research. All experiments were carried out using the base engine oil (Shell Tellus 10) and by adding 1% by weight Molybdenum Disulfide (MoS_2) to the base engine oil at the flow rate of 55 cm^3/min. The optimum mixture of nanoparticle was determined according to the results obtained from studies in the literature [21-23]. Accordingly, molybdenum disulfide was added to the oil at 1% wt.

The bearing oil temperatures were measured using a thermometer at the outlet of the bearing and dynamic viscosity values were taken from the calibration chart of viscosity-temperature for this type of oil, given Table 1.

Table 1. Typical physical characteristics of the base oil, Shell Tellus 10

Shell Tellus C Oil		
Viscosity Garde (ISO 3448)		10
ISO Oil Type		HL
Viscosity cSt	40 °C	10
	@ 0 °C	2.5
Viscosity Index (IP34)		70
Density @ 15 °C (IP 160) kg/l		0.877
Falsh Point (IP34) (PMCC) °C		166

Tests carried out two different oil temperature, each test oil temperature remained in the range 25-30 °C and 50-60 °C with matching dynamic viscosity 0.02 and 0.01 $N.s/m^2$, respectively. After running-in (initial) process, and final (the end of working at 253N and 553N; 1.22 and 2.66 MPa pressures) surface roughness of the journal bearings and the shaft was measured by using surface roughness tester (Mahr-Germany Perthen Perthometer) in pursuant of DIN 4768. Surface roughness parameters such as center line average (Ra), average peak to valley height (Rz) and maximum peak to valley height (Rmax) were measured by using a driving unit at a tracing speed of 0.5 mm/s with traversing and cut-off lengths, 4.8 mm. In order to read measured roughness values, the device was connected to a measuring indicator. The average values of five roughness measurements were taken. The surface roughness results before and after the tests are given in Table 2. As seen in Table 2, the roughness of the bearing and shaft surfaces changed very little during the tests.

Table 2. Surfaces roughness values before and after experimental studies

Before (Bearing)		After (Bearing)		Before (Shaft)		After (Shaft)	
Ra (μm)	Rz (μm)	Ra (μm)	Rz (μm)	Ra (μm)	Rz (μm)	Ra (μm)	Rz (μm)
0,25	1,45	0,18	0,92	0,36	1,17	0,34	2,02

2.1. Bearing Material

In order to minimize wear in plain bearings, the selection of the bearing material working with the shaft is very important [24]. For this reason, containing tin (6-40%), aluminum-based ZA-27 plain bearing was used as the bearing material in the experiments. To determine the chemical compositions of the alloy, the atomic absorption method was used. The results of chemical analysis of the plain bearing and measured MoS_2 particles size range were given in Table 3.

Table 3. Chemical compositions of the alloys and measured MoS_2 particles size range

Alloy	Chemical composition (Weight %)				Size range of MoS_2 particles (±1μm)
ZA-27	Zn	Al	Cu	Mg	
	70,8	27.2	2.01	0,02	3,5-11

2.2. Physical Methods

Experimental test rig of the journal bearing is given in Fig. 2 as; a) schematic representation, b) photo and c) technical drawing with dimensions. The rig consists of a rigid tubular steel of square profile frame equipped with a pneumatically operated pressurized oil supply, a filtered oil tank, a modified electronic circuit with shaft and bearing housing, a direct current variable speed control unit, assembly of split journal bearing and electronic instrumentation for indicating shaft speed and motor current. The apparatus operates a 220-240 V mono phase electrical supply and an air supply at a minimum pressure of 0.6 MPa. The pressurized oil system composed of a double acting pneumatic cylinder driving a hydraulic oil cylinder in a reciprocating manner. Integrated in the system are two over pressure relief valves, which limit the oil supply pressure to the bearing and hydrostatic pad to approximately 0.15 and 1.6 MPa respectively.

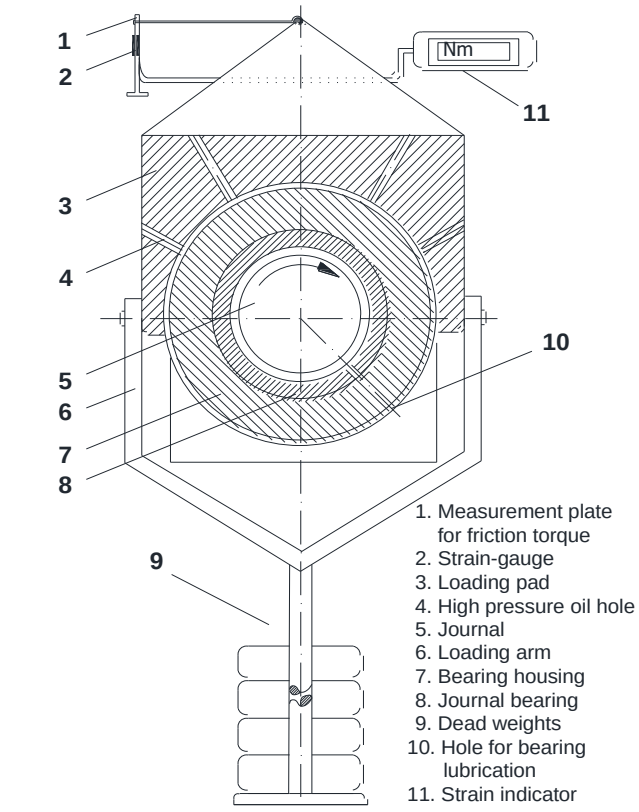
In this experimental test rig, the hydrostatic pad is constructed using mild-steel, and accurately lapped to fit the journal bearing housing with suitable tolerance. Without affecting the sensitivity of friction torque measurements, 11 oil jets used for separate the hydrostatic pad and outside of the bearing cap, enables the bearing load to be transmitted to the bearing, which is increase up to maximum 553 N. Oil leakage from the bearing and hydrostatic pad is collected in the drip tray and returned to the reservoir. The test oil is collected in another tray and returned to the gap between shaft and bearing via pump. The journal bearing housing is made of iron and has a correctly ground bore to support thin-walled bearings used in automobile engines. The sleeve was made of steel hardened to 722 HB.

The rotational speed could be adjusted in the range of 0-1100 rpm (0-2.88 m/s) using a direct current speed control unit.

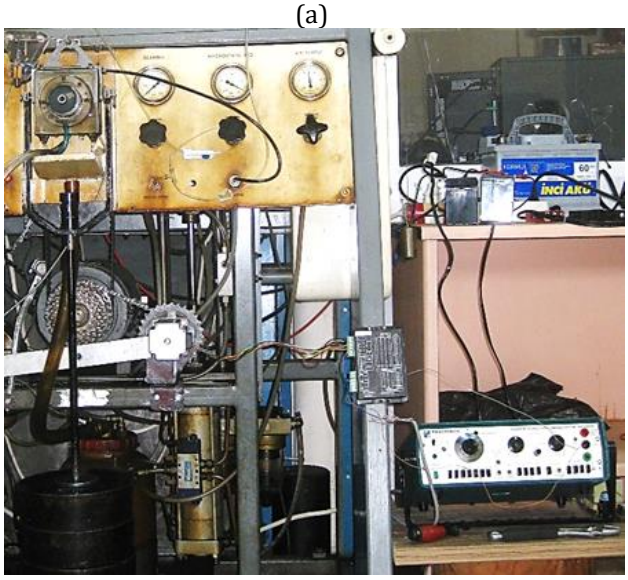
In order to be able to measure the measurements in the boundary and mixed friction regions in detail at the test duration, the stepper motor and equipment's were mounted to the system using one-way clutch gear, Fig 2-b. The effective bearing length was taken as 18 mm (Fig. 2-c) to operate at a pressure of 2.66 MPa because the maximum allowable load of the test rig was 553 N. The diameter of shaft and thin-walled journal bearing are 50.760 and 50.850 mm, respectively. The bearings had a bearing clearance of 45 μm, Fig. 2-c.

By using a Wheatstone bridge circuit with strain gauges on the torque plate, the friction torque on the journal bearing was measured. Two strain gauges on the bridges were used as active gauges, others used as passive gauges. The torque signal obtained from the bridge circuit was calibrated and transmitted to the recorder device and this enabled monitoring and recording the signals. Using strain gauges signals, the recorded signal values were then converted into friction torque considering the dimension of bearing. Before executing tests, the journal bearing surfaces were exposed to running-in process under a pressure of 1.2 MPa at a constant rotational shaft speed (500 rpm, 1.3 m/s) for 10 minutes. After this process, experimental tests were carried out at rotational speeds ranging from 0 rpm to 300 rpm (0.78 m/s) for 5 minutes at each shaft speed. The pressures of hydrostatic pad and lubricating oil were selected as 1.2 and 0.1 MPa respectively. To investigate the effects of additive engine oil on bearing performance, base engine oil (Shell Tellus 10) and MoS_2 nano particles were used by adding 1% by weight to the base oil. The structure of the particles was characterized by scanning electron microscopy (SEM), shown in Fig. 3. The MoS_2 nanoparticles have a platelet-like shape with an average diameter of 7.2 μm, (Table 3). Purity of the nanoparticles is 99%. The average particle size and density of the MoS_2 additive at 25°C are 3.20 μm and 5.06 g/cm^3, respectively.

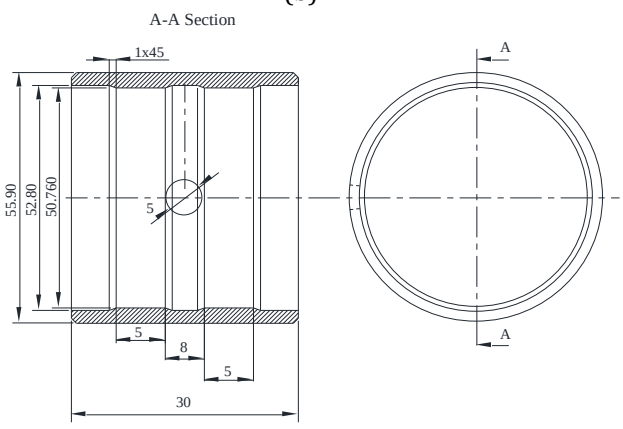
After each experiment, the bearing and shaft surfaces were cleaned with acetone-isopropanol mixture and the tests were prepared for initial conditions. The results were evaluated taking into consideration the formation of a boundary lubrication, mixed lubrication and liquid film lubrication. Experiments were carried out up to 300 rpm; 553 and 353 N bearing loads. Additive rates were selected as 1wt%. The running time of the tests were implemented in the range of 600-700s and 20-55°C under 70% relative humidity. Information on test conditions is also given in Table 4.



(a)



(b)



(c)

Figure 2. The journal bearing test rig; (a) Schematic, (b) Photo (c) Dimensions of the plain bearing

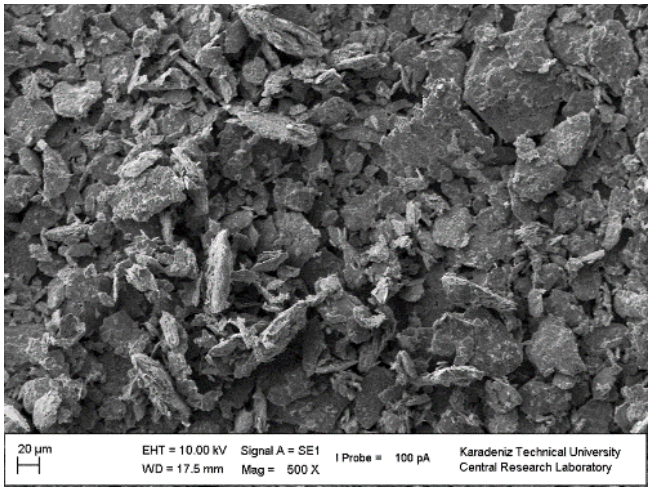


Figure 3. SEM morphologies of the MoS_2 nanoparticles

3. Results and Discussion

In the experiments, the base oil (Shell Tellus 10) and 1 wt.% MoS_2 additive were used to determine the effects of MoS_2 particles on the friction behavior in the bearing. As shown in Fig. 4, using MoS_2 additives, lower friction torque values were obtained in the tests compared to the base oil at 27°C test condition. This difference is even more pronounced in the boundary and mixed friction regions. For example, an average increase of 0.3 and 0.6 Nm is observed in the friction torque at 253N and 553N bearing loads in the boundary friction region, respectively. It is seen that the friction torque values in the bearing increase with the increasing bearing load. As a result, it has been determined that the effect of MoS_2 additive is more dominant in metal-to-metal contact operating conditions than that of the liquid film region. In the bearing tests with different oils performed at 253 N bearing load, the transition speed values are very close to each other and are around 14 rpm. As shown in Fig. 5, using of MoS_2 additives at 253N, the minimum friction torque value was obtained as 0.037 Nm. While using the base oil at the same test conditions, this value was 0.041 Nm. It was detected that MoS_2 particles has a more effects to decrease the friction torques than the base oil in the bearing.

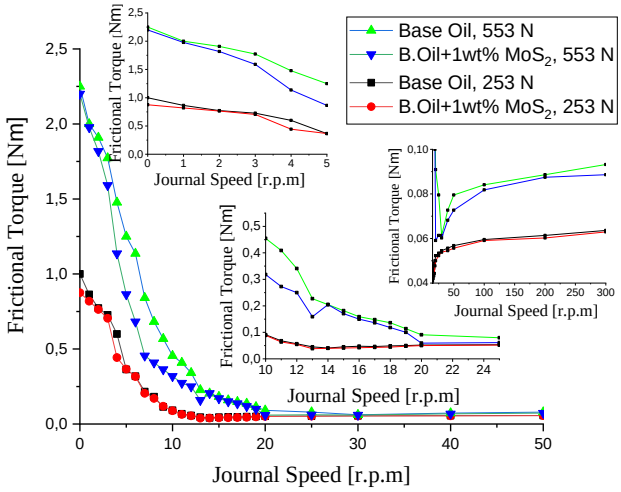


Figure 4. Variation of friction torque with rotational speed in two different oil and load tests

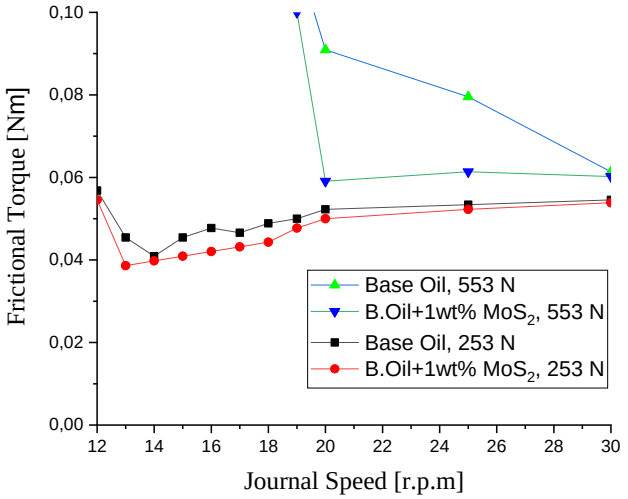


Figure 5. Variation of friction torque with rotational transition speed values in two different oil and load tests

In addition, the rotational transition speed values decrease with the use of MoS_2 additives, hence, frictional torque reaches at the minimum value at 27°C test condition. This means that, MoS_2 additives facilitates the formation of oil film between the bearing and the shaft. Using the base oil at 553 N bearing load, friction torque and rotational transition speed values increases and these values are 0.06 Nm and 30 rpm, respectively. Using MoS_2 additives these values are 0.057 Nm and 20 rpm, as shown in Fiq 5. In the fluid friction region, this difference is approximately 0.02 Nm at low load and 0.01 Nm at high load.

In order to determine the friction coefficient in the plain bearing at 27°C test conditions, firstly, the peripheral friction force values in the bearing were measured, and then, the friction coefficient values were obtained by taking the ratio of this friction force to the bearing load. Variation of frictional coefficient with rotational speed for two different oil (base oil and 1 wt% MoS_2 additive) and load was given in Fig. 6. As shown in Fig. 6, higher friction coefficient values were obtained in the base oil tests compared to the MoS_2 additive oil. These differences were even more in the boundary and mixed friction regions. It is seen that the friction coefficient values formed in the bearing decreases with increasing bearing load. The reason for the decrease may be that, compared to the friction force, the increase in the bearing load is higher, or the internal friction of the oil decreases as a result of the thinning of the oil film between the surfaces. Hence, the effect of surface roughness was effective in terms of friction coefficient under metal-to-metal contact operating conditions.

In addition, the rotational transition speed values decrease with the use of MoS_2 additives, hence, frictional coefficient reaches at the minimum value. Using the base oil at 253 N, the values of transition speed and frictional coefficient are 14 rpm and 0.007, respectively. With the use of MoS_2 additives, the values decreased to 13 rpm and 0.006, respectively. This means that, MoS_2 additives facilitates the formation of oil film between the bearing and the shaft. Using the base oil at 553 N bearing load, friction coefficient and rotational transition speed values increases and these values are 0.015 and 30 rpm, respectively. Using MoS_2 additives these values are 0.009 and 20 rpm, respectively, as shown in Fig. 7. In the fluid friction region, this difference is approximately 0.001 Nm at low load and 0.015 Nm at high load.

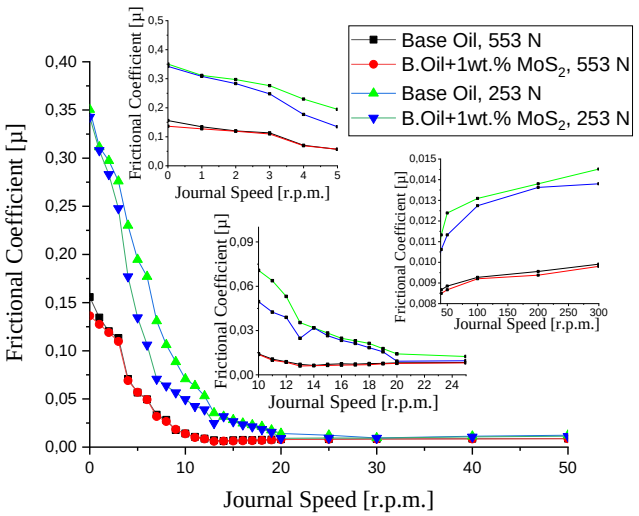


Figure 6. Variation of frictional coefficient with rotational speed in two different oil and load tests

Using the MoS_2 particles as additives, the effect of oil temperature on the friction torque is given in Fig. 8. As shown in Fig. 8, higher friction torque values were obtained at higher oil temperatures in the tests.

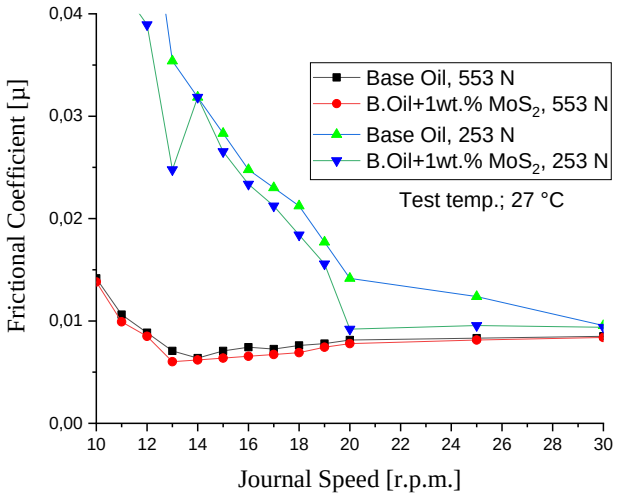


Figure 7. Variation of friction coefficient with rotational transition speed values in two different oil and load tests

Compared to the base oil, tests using oil with MoS_2 additives reveal lower temperature differences. This difference decreasing with the load As a result, it has been determined that the effect of surface As In the fluid friction region, the friction torque in the bearing decreases with the increasing oil temperature. While the friction torque decreases due to increasing temperature, the transition speed is increases. With the increase of temperature in the fluid friction region, the friction torque also decreases, while the transition speed increases. As a result, it has been determined that the effect of oil temperature is effective in terms of friction torque under metal-to-metal contact operating conditions.

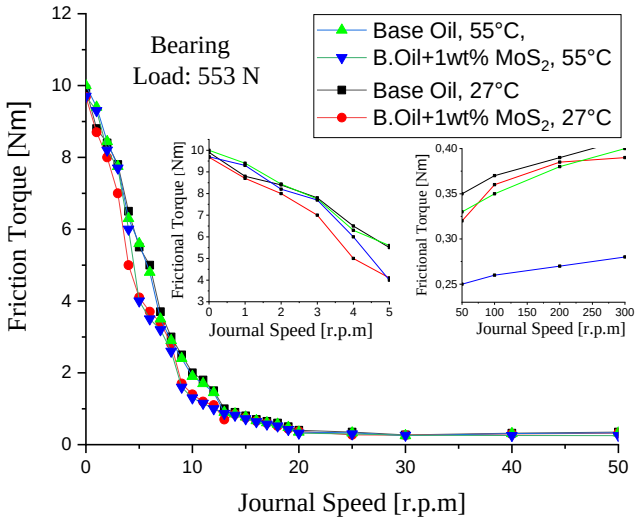


Figure 8. Variation of friction torque with speed using two different oil and temperature tests

The effects of oil temperature with different oil (base oil and 1 wt.% MoS_2 additive) on the friction coefficient were given in Fig. 9. As shown in Fig. 9, higher friction coefficient values were obtained in the base oil tests compared to the MoS_2 additive oil at 55 °C test conditions. These differences were more in the boundary and mixed friction regions. It was seen that the friction coefficient values formed in the bearing decreases with the increasing bearing load at high working oil temperatures, 55°C. The reason for the decrease may be that compared to the friction force the increase in the bearing load is higher, or the internal friction of the oil decreases as a result of the thinning of the oil film between the shaft and bearing surfaces. Hence, the surface roughness was effective in terms of friction coefficient under metal-to-metal contact operating conditions.

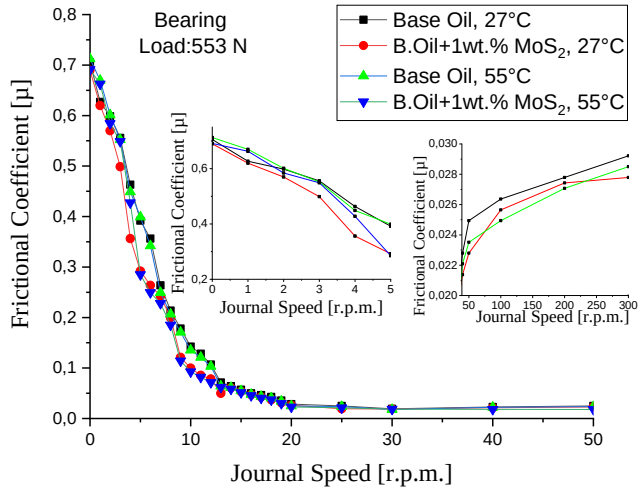


Figure 9. Variation of frictional coefficient with speed in two different oil and temperature at 553 N

Additionally, variation of frictional coefficient with transient speed in two different oil and temperature at 553 N test were detailed in Fig.10.

The rotational transition speed values decrease with decreasing oil temperature for both of MoS_2 additives and base oil. But using MoS_2 additives these decreases were more than that of the base oil. It can be said that, oil viscosity decreased with increasing temperature, and oil film formation becomes easier. Friction coefficient reaches to the minimum value when transient rotational speed has occurred. Using the base oil at 553 N, the values of transition speed and frictional coefficient were 30 rpm and 0.018, respectively, at the 55°C test condition. With the use of MoS_2 additives at the same oil temperature, friction coefficient value decreased to 0.016. Using the base oil at 27 °C test condition, friction coefficient and rotational transition speed values decreased and these values were 0.021 and 30 rpm, respectively. Using MoS_2 additives these values were 0.017 and 20 rpm, respectively. Hence, increasing with oil temperature, oil film formation between shaft and bearing becomes more difficult.

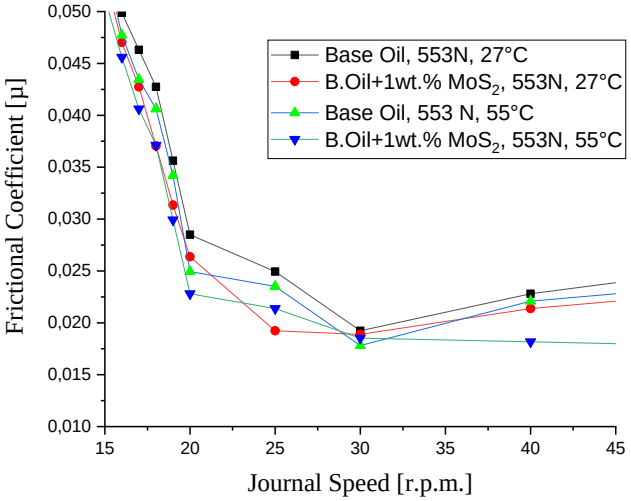


Figure 10. Variation of frictional coefficient with transient speed in two different oil and temperature at 553 N

4. Conclusion

The aim is to further reduce friction by adding inorganic compounds as additives to the engine oils used between the shaft and the bearing. The results of this study show that all the influential factors such as base oil, additive oil, test pressure and test speed have a significant effect on controlling the frictional behavior of thin-walled plain bearings. In this study, by adding 1% by weight Molybdenum Disulfide (MoS_2) to the base (Shell Tellus 10) engine oil, the effects of the MoS_2 additives on the bearing performance were examined and the following results obtained.

1. Compared to the base oil, the MoS_2 additives exhibits less friction behavior at increasing bearing load and temperatures.
2. MoS_2 additives is more effective in the mixed friction region than that of liquid film region. As a results, it is possible to evaluate whether tribofilm formation may occur in boundary or mixed lubrication on the surfaces of the shaft and bearing.
3. The friction reduction effect of MoS_2 additives are increasing with higher bearing load.
4. As a result of these tests, owing to MoS_2 additives facilitates the formation of oil film between the bearing

and shaft, it is concluded that MoS_2 additives reduces wear and increases journal bearing life.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Esmaeili, M., Oskouei, A. R., Mirhadizadeh, S. A., TSE, W. T., & Hoshyar, N. (2016). Prediction of Hydrodynamic bearing performance based on effective parameters by Neural Network. *International Journal of Engineering and Management Sciences*, *7*(2), 92-99.
2. Cortes, V., Sanchez, K., Gonzalez, R., Alcoutlabi, M., & Ortega, J. A. (2020). The performance of SiO_2 and TiO_2 nanoparticles as lubricant additives in sunflower oil. *Lubricants*, *8*(1), 10. https://doi.org/10.3390/lubricants8010010
3. Rudnick, L. R. (Ed.) (2017). Lubricant Additives: Chemistry and Applications, 3rd ed.; Chemical Industries; CRC Press Taylor & Francis Group: Boca Raton, FL, USA.
4. Lee, K., Hwang, Y., Cheong, S., Choi, Y., Kwon, L., Lee, J., & Kim, S. H. (2009). Understanding the role of nanoparticles in nano-oil lubrication. *Tribology Letters*, *35*(2), 127-131.
5. Cortes, V., Sanchez, K., Gonzalez, R., Alcoutlabi, M., & Ortega, J. A. (2020). The performance of SiO2 and TiO2 nanoparticles as lubricant additives in sunflower oil. *Lubricants*, *8*(1), 10. https://doi.org/10.3390/lubricants8010010
6. Mousavi, S. B., Heris, S. Z., & Estellé, P. (2020). Experimental comparison between ZnO and MoS2 nanoparticles as additives on performance of diesel oil-based nano lubricant. *Scientific Reports*, *10*(1), 1-17. https://doi.org/10.1038/s41598-020-62830-1
7. Beheshti, A., Huang, Y., Ohno, K., Blakey, I., & Stokes, J. R. (2020). Improving tribological properties of oil-based lubricants using hybrid colloidal additives. *Tribology International*, *144*, 106130.
8. Wu, H., Zhao, J., Xia, W., Cheng, X., He, A., Yun, J. H., ... & Jiang, Z. (2017). A study of the tribological behaviour of TiO2 nano-additive water-based lubricants. *Tribology International*, *109*, 398-408.
9. Elianov, O., Garusi, S., Rosentsveig, R., Cohen, S. R., Feldman, Y., Pinkas, I., ... & Shay, B. (2018). Deposition of metal coatings containing fullerene-like MoS_2 nanoparticles with reduced friction and wear. *Surface and Coatings Technology*, *353*, 116-125.
10. Hu, E. Z., Xu, Y., Hu, K. H., & Hu, X. G. (2017). Tribological properties of 3 types of MoS_2 additives in different base greases. *Lubrication Science*, *29*(8), 541-555.
11. Osim, W., Stojanovic, A., Akbarzadeh, J., Peterlik, H., & Binder, W. H. (2013). Surface modification of MoS 2 nanoparticles with ionic liquid–ligands: towards highly dispersed nanoparticles. *Chemical Communications*, *49*(81), 9311-9313.
12. Tannous, J., Dassenoy, F., Lahouij, I., Le Mogne, T., Vacher, B., Bruhács, A., & Tremel, W. (2011). Understanding the tribochemical mechanisms of IF-MoS_2 nanoparticles under boundary lubrication. *Tribology Letters*, *41*(1), 55-64.
13. Rajendhran, N., Palanisamy, S., Periyasamy, P., & Venkatachalam, R. (2018). Enhancing of the tribological characteristics of the lubricant oils using Ni-promoted MoS2 nanosheets as nano-additives. *Tribology international*, *118*, 314-328.
14. Ripoll, M. R., Tomala, A., Gabler, C., Dražić, G., Pirker, L., & Remškar, M. (2018). In situ tribochemical sulfurization of molybdenum oxide nanotubes. *Nanoscale*, *10*(7), 3281-3290.
15. Lahouij, I., Vacher, B., Martin, J. M., & Dassenoy, F. (2012). IF-MoS2 based lubricants: influence of size, shape and crystal structure. *Wear*, *296*(1-2), 558-567.
16. Song, H., Wang, B., Zhou, Q., Xiao, J., & Jia, X. (2017). Preparation and tribological properties of MoS2/graphene oxide composites. *Applied Surface Science*, *419*, 24-34.
17. Liu, L., Huang, Z., & Huang, P. (2016). Fabrication of coral-like MoS2 and its application in improving the tribological performance of liquid paraffin. *Tribology International*, *104*, 303-308.
18. Charoo, M. S., Wani, M. F., Hanief, M., & Rather, M. A. (2017). Tribological properties of MoS2 particles as lubricant additive on EN31 alloy steel and AISI 52100 steel ball. *Materials Today: Proceedings*, *4*(9), 9967-9971.
19. Baş, H. (2021). Investigation of effects of boron additives on performance of cam mechanisms. *International Journal of Automotive Engineering and Technologies*, *10*(1), 60-66.
20. McKee, S. A., & McKee, T. R. (1932). Journal-bearing friction in the region of thin-film lubrication. *SAE Transactions*, 371-377.
21. Mushtaq, Z., & Hanief, M. (2021). Evaluation of Tribological Performance of Jatropha Oil Modified With Molybdenum Disulphide Micro-Particles for Steel–Steel Contacts. *Journal of Tribology*, *143*(2). https://doi.org/10.1115/1.4047752
22. Xu, Z. Y., Hu, K. H., Han, C. L., Hu, X. G., & Xu, Y. F. (2013). Morphological influence of molybdenum disulfide on the tribological properties of rapeseed oil. *Tribology Letters*, *49*(3), 513-524. https://doi.org/10.1007/s11249-012-0092-8
23. Srinivas, V., Thakur, R. N., & Jain, A. K. (2017). Antiwear, antifriction, and extreme pressure properties of motor bike engine oil dispersed with molybdenum disulfide nanoparticles. *Tribology Transactions*, *60*(1), 12-19. https://doi.org/10.1080/10402004.2016.1142034
24. Çuvalcı, H., & Baş, H. (2004). Investigation of the tribological properties of silicon containing zinc–aluminum based journal bearings. *Tribology international*, *37*(6), 433-440.

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Investigation of the discharge flow rate patterns at real-time traffic signal control intersections

Nihat Can Karabulut [1], Murat Özen*[2], Oruç Altıntaşı[3]

[1]*Yıldız Technical University, Department of Civil Engineering, Türkiye*
[2]*Mersin University, Department of Civil Engineering, Türkiye*
[3]*İzmir Katip Çelebi University, Department of Civil Engineering, Türkiye*

Keywords
Real-time managed intersection
Queue length
Green time
Time headway
Discharge flow rate

Research Article
DOI: 10.31127/tuje.1020765

Received: 09.11.2021
Accepted: 16.02.2022
Published: 01.04.2022

Abstract

This study investigates the effects of variable queue lengths and green times on discharge flow rates at real-time managed intersections in Mersin, Turkey. For this purpose, traffic flow data were collected at two different signalized intersections during morning peak hours for two days. The traffic data including the time headways, queue lengths for each cycle were derived from video records via MATLAB coding while the signal timing data were obtained from Mersin Metropolitan Municipality. The impact of variable queue lengths and green times on discharge flow rate were evaluated separately via analysis of variance (ANOVA) tests. The results indicated that time headways of the first vehicles in the queue were statistically larger than the time headways of the remaining vehicles in the queue (p-value<0.05). On the other hand, the time headways of the remaining vehicles in the queue were not found statistically different at 95 confidence level (p-value>0.05). Furthermore, the effect of the variable green time on discharge flow rate revealed that the significant difference was only observed for the first twelve seconds of the green time.

1. Introduction

Traffic congestion has become one of the major problems of urban transportation as a result of the rapid increase in vehicle ownership. Geometrically well-designed signalized intersections and their appropriate management can reduce the consequences of this problem. Fixed-time and real-time signalization systems are used to manage the traffic flow in urban areas. However, fixed-time signalized systems have some drawbacks, when the traffic is oversaturated during the peak hours, which causes longer queue lengths. To eliminate these problems, traffic engineers have developed signalized systems that manage intersections with real-time traffic flow data. In these systems, the green times of the approach legs are determined by considering the instant traffic demand.

Discharge flow rate and saturation flow can be used to evaluate performance of the signalized intersections. The relationship of between these two parameters, green time, queue length and intersection geometry have been studied in many studies in the literature. Khosla and Williams [1] found that discharge flow rate tends to decrease after the 60 seconds of green time, but it is not statistically significant. Denney et al. [2] concluded that as the green time increases, the discharge flow rate decreases. On the other hand, Lin and Thomas [3] stated that the discharge flow rate increases with increase in effective green time. Stanić et al. [4] concluded that discharge flow rate of the first four vehicles in the queue is significantly lower than the discharge flow rate of the other vehicles in the queue. More specifically, it was observed that discharge flow rate increases 10% to 15% after the first four vehicles. Chaudhry and Ranjitkar [5] observed the increasing trend in discharge flow rate toward the back of the queue. This result was also supported by other studies [6-7]. On the other hand, some studies concluded that discharge flow rate decreases toward the back of the queue [8-10].

* **Corresponding Author**
(karabulutnihatcan@gmail.com) ORCID ID 0000-0002-4294-0215
*(ozen.murat@mersin.edu.tr) ORCID ID 0000-0002-1745-7483
(oruc.altintasi@ikcu.edu.tr) ORCID ID 0000-0002-4217-1890

Cite this article
Karabulut, N. C., Özen, M., & Altıntaşı, O. (2023). Investigation of the discharge flow rate patterns at real-time traffic signal control intersections. Turkish Journal of Engineering, 7(1), 49-55

A majority of these studies have been carried out on intersections managed with fixed-time signalization systems, however, it is very limited for real-time managed intersections which is the main motivation of this research. This study aims to make a discussion on the effect of variable queue lengths and green times on traffic flow at real-time managed intersections. For this purpose, traffic flow data were collected at two different real-time managed signalized intersections during morning peak hours in Mersin, Turkey. Relations between green time and queue length and discharge flow rate were statistically investigated.

2. Study Area

To examine the effect of the variable queue lengths and green times on the discharge flow rate, two isolated intersections were selected on major arterials located in Mersin, Turkey (see Figure 1). These intersections are fully actuated and the green times were allocated based on the arrival flow rates. One of them was 4-leg (namely CNR) and the other one was roundabout (namely Kipa) as shown in Figure 1. The following points were considered while selecting these intersections:

- Intersections with three lanes in each approach were selected to minimize the effect of right and left turning vehicles.
- Intersections with low heavy vehicle volume were selected to reduce negative effects of heavy vehicles.
- Intersections with lower pedestrian movements were selected to eliminate negative effects of pedestrian movements.
- Intersection with smaller approach slopes (i.e., less than 2%) were selected to reduce negative effects of high slopes.

Video camera data were collected during the morning peak hours of 7:30-9:30 on October 8-9, 2019. The approach legs with higher arrival flow rates were considered, since the effect of variable green times can be observable only if the approach leg was under saturation flow condition. Thus, east, west and north approach legs were taken into consideration for the Kipa intersection, while only west approach leg was considered for the CNR intersection. In addition, the right and left turning vehicles were not considered for this study; thus, only the middle lane of each approach leg was analyzed.

Figure 1. General view of Kipa and CNR intersections

3. Methodology

Figure 2 presents the proposed methodology of the study. For each approach leg of the Kipa and CNR intersections, traffic flow video data and signal timing data (green and red times) were obtained from Transportation Department of the Mersin Metropolitan Municipality. Time headway of the vehicles passing through the green time were calculated by deciphering video camera records in MATLAB® programming software, will be later converted to the discharge flow rate. Note that only cycles with queue length greater than 8 vehicles were considered based on the Highway Capacity Manual recommendations [11]. In overall, 337 cycles were analyzed from two intersections (254 cycles for Kipa intersection and 83 cycles for CNR intersection) (see Table 1).

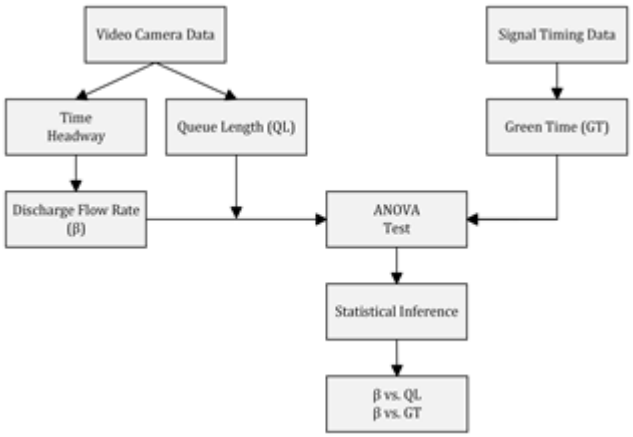


Figure 2. Methodological framework of the study.

Table 1. Traffic flow characteristics at Kipa and CNR intersections

Intersection		Kipa			CNR
Approach Leg		West	East	North	West
Number of Cycles		125	53	76	83
Green Time (sec.)	Min.	31	19	20	18
	Avg.	56	32	24	22
	Max.	72	35	30	33
Red Time (sec.)	Min.	30	48	54	63
	Avg.	46	67	76	69
	Max.	78	77	91	90
Cycle Time (sec.)	Min.	76	76	77	84
	Avg.	105	101	102	94
	Max.	130	114	120	111
Queue Length (veh./cycle/lane)	Min.	8	8	8	8
	Max.	25	15	15	13
Arrival Cycle Volume (veh./cycle/lane)	Min.	9	8	8	8
	Max.	33	15	15	16
Traffic Composition (%)	Car	94	91	95	93
	Minibus	5	3	3	5
	Others	1	6	2	2

To study effect of queue lengths, vehicles in the queue were grouped with four vehicles. Number of vehicles in each group (N), their average time headways ($\overline{h}$) and corresponding discharge flow rates (β) are presented in Table 2 and 3.

As it is seen in Table 2, 6 groups were obtained for Kipa intersection. The number of vehicles (N) in each group varies between 36 and 1.016 and the average time headways ($\overline{h}$) vary between 1.69 and 2.73 seconds (see Table 2.). In addition, 3 groups were obtained for CNR intersection. The number of vehicles (N) in each group varies between 126 and 332, and the average time headways ($\overline{h}$) vary between 1.75 and 2.35 seconds (see Table 2.).

Table 2. Discharge flow rate with respect to queue positions at Kipa intersection

Kipa	Queue Position					
	1–4	5–8	9–12	13–16	17–20	21–25
N	1,016	1,016	618	249	79	36
$\overline{h}$	2.73	2.03	1.91	1.85	2.03	1.69
β	1,319	1,773	1,885	1,946	1,773	2,130

Analysis of variance (ANOVA) test was used to study effects queue lengths on the discharge flow rates and the following hypothesis was tested at 0.05 significance level.

H_0: The time headways of the vehicles in all groups are the same.
H_1: Time headways of the vehicles in at least one group is different.

Table 3. Discharge flow rate with respect to queue positions at CNR intersection

CNR	Queue Position		
	1 – 4	5 – 8	9 – 13
N	332	332	126
$\overline{h}$	2.35	1.83	1.75
β	1,532	1,967	2,057

Since the green time ranges at Kipa and CNR intersections are significantly different, different green time intervals were used to analyze effects of green times on discharge flow rates at Kipa and CNR intersections. Specifically, Kipa intersection was analyzed with green time intervals of 12 seconds and CNR intersection was

analyzed with green time intervals of 6 seconds (see Figure 4). As an example, a total 801 were entered Kipa intersection from during the cycles with a green time between 25 and 35 seconds. 268 vehicles entered the intersection during the first 12 seconds of green time, 398 vehicles entered the intersection during the second 12 seconds of green time and 135 vehicles entered during the remaining green time period.

Analysis of variance (ANOVA) test was used to study effects variable green times on discharge flow rates and the following hypothesis was tested at 0.05 significance level.

H_0: Time headways of the vehicles during each 6 or 12 second intervals of the green time are the same.
H_1: Time headways of the vehicles during at least one the 6 or 12 second interval of the green time is different.

4. Results

The traffic composition was very similar for all approaches; almost 90% of the vehicles were private cars followed by the minibuses with 2.8%-4.7%. Cycle volume indicating the number of the passing vehicles during each cycle showed that arrival flow of the west direction of the Kipa intersection was almost twice of the other approaches. The higher green times were observed for the west direction of the KIPA intersection ranging from 31 sec to 72 sec with an average of 56 sec (see Table 4). The other approaches did not seem to be congested in which the green times were ranging from 20 sec to 35 sec. For the CNR intersection, green times ranged from 18 sec to 33 sec. Except for the west approach of Kipa intersection, the maximum queue lengths were almost same for the remaining three approaches with 13-15 vehicles (see Table 4). In this study, queue lengths were calculated based on the number of stopped vehicles at the end of the red time. The vehicles entering to the queue after the start of green time were not considered.

Table 4. Average time headways of vehicles at 12 and 6 second green time intervals at Kipa and CNR intersections

Green Time (sec.)	Green Interval (sec.) – Kipa Intersection											
	[0 – 12)		[12 – 24)		[24 – 36)		[36 – 48)		[48 – 60)		[60 – 72)	
	$\bar{h}$	N	$\bar{h}$	N	$\bar{h}$	N	$\bar{h}$	N	$\bar{h}$	N	$\bar{h}$	N
[15 – 25)	2.68	179	1.90	215	2.04	4	-	-	-	-	-	-
[25 – 35)	2.66	268	1.96	398	2.33	135	-	-	-	-	-	-
[35 – 45)	2.53	212	2.09	267	2.49	185	2.61	26	-	-	-	-
[45 – 55)	2.71	90	2.24	135	2.14	128	2.44	105	3.30	20	-	-
[55 – 65)	2.72	126	2.07	185	2.06	189	2.13	173	2.65	107	3.30	11
[65 – 75)	2.67	163	2.02	239	2.06	228	2.17	219	2.16	220	2.28	149
Green Time (sec.)	Green Interval (sec.) – CNR Intersection											
	[0 – 6)		[6 – 12)		[12 – 18)		[18 – 24)		[24 – 30)		[30 – 36)	
	$\bar{h}$	N	$\bar{h}$	N	$\bar{h}$	N	$\bar{h}$	N	$\bar{h}$	N	$\bar{h}$	N
[15 – 20)	2.59	40	1.92	66	1.76	66	2.12	6	-	-	-	-
[20 – 25)	2.49	80	2.08	131	1.78	139	1.89	88	-	-	-	-
[25 – 30)	2.87	18	2.20	35	1.91	40	1.63	40	2.04	10	-	-
[30 – 35)	2.18	14	2.24	20	1.89	23	1.68	24	1.77	22	1.69	5

4.1.The effects of queue length on discharge flow rate

The results obtained for Kipa and CNR intersections were consistent. It was found that the time headways of the first four vehicles in the queue were statistically larger than the time headways of the remaining vehicles in the queue (p-value<0.05). On the other hand, the time headways of the remaining vehicles in the queue were not statistically different (p-value>0.05).

Kipa intersection reached the maximum flow rate between 13th and 16th vehicles in the queue, which was about 1,950 veh/hour/lane (see Figure 3). Note that the number of vehicles in the last queue group in Kipa intersection was quite low. Therefore, it was not taken into consideration in determining the maximum discharge flow rate. The CNR intersection reached the maximum flow rate between 9th and 13th vehicles in the queue, which was about 2,050 veh/hour/lane (see Figure 3).

4.2. The effects of variable green time on discharge flow rate

For Kipa intersection, the time headways of the vehicles in the first 12 seconds of the green time were statistically larger than the time headways of the remaining vehicles during the green time (p-value<0.05).

On the other hand, it is not possible to say anything statistically about the relationship between other groups. This result indicated that the discharge flow during the first 12 seconds of the green time were lower than the discharge flow rates of remain green time period (see Figure 4). Therefore, it can be concluded that first few vehicles passing through in the first 12 seconds of the green time were subjected to significant initial losses.

For CNR intersection, the time headways of the vehicles in the first 6 seconds of the green time may said that statistically larger than the time headways of the remaining vehicles during the green time (p-value<0,05). On the other hand, it is not possible to say anything

statistically about the relationship between other groups. This result indicated that, the discharge flow during the first 6 seconds of the green time were lower than the discharge flow rates of remain green time period (see Figure 5). Therefore, it can be concluded that first few vehicles passing through in the first 6 seconds of the green time were subjected to significant initial losses. For the cycles with less than 25 seconds of green time, the maximum discharge flow rate was obtained between 12th and 18th second green. Otherwise, the maximum discharge flow rate was obtained between 18th and 24th second green (see Figure 5).

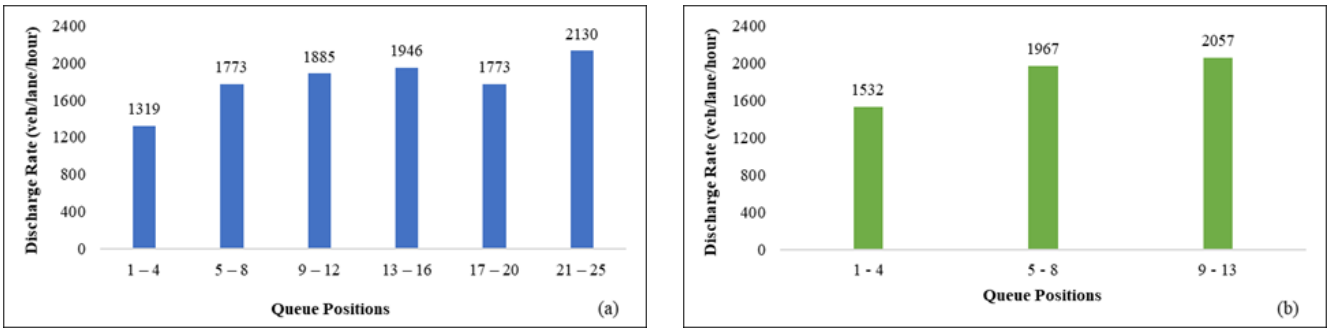


Figure 3. Discharge flow rates of queue groups at (a) Kipa and (b) CNR intersections

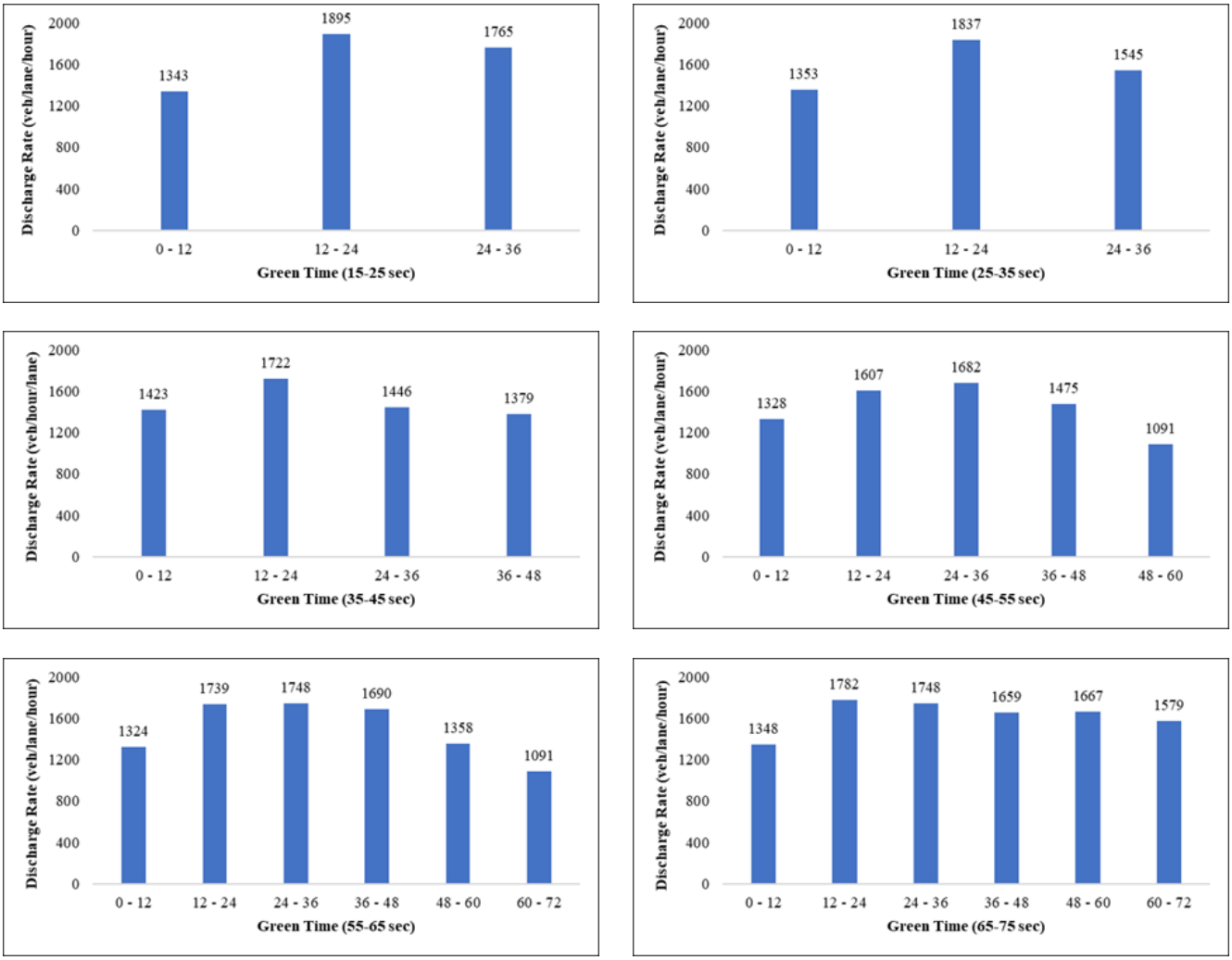


Figure 4. Discharge flow rate of vehicles entering the Kipa intersection at 12 second intervals

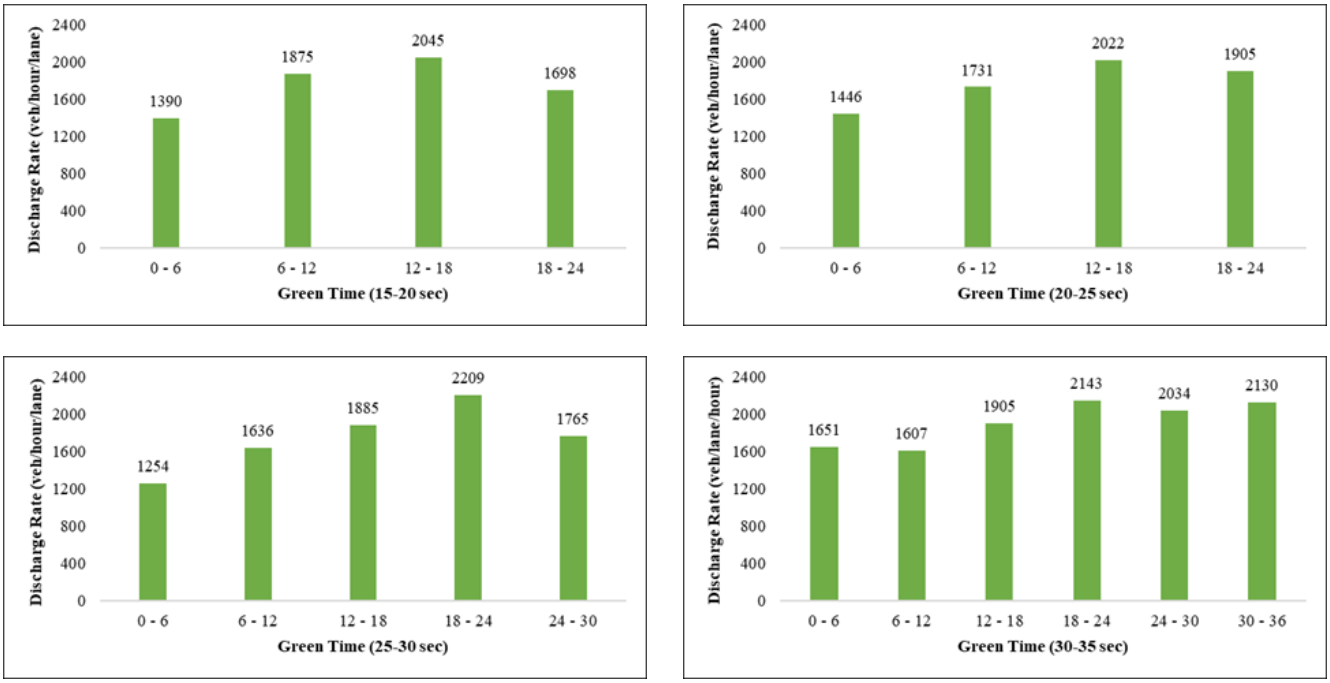


Figure 5. Discharge flow rate of vehicles entering the CNR Intersection at 6 second intervals

5. Conclusion

Two real-time signalized intersection approaches in the Mersin city center were studied to analyze the impacts variable green times and queue lengths on the discharge flow rates. The results can be summarized as follows:

- Time headways of the first four vehicles in the queue were statistically larger than the time headways of the remaining vehicles in the queue. Therefore, the discharge flow rate of the first four vehicles was lower than the discharge flow rate of the remaining vehicles. More specifically, the discharge flow rate increased almost 30% after the first four vehicles. This means that the first four vehicles in the queue were subjected to significant initial losses. In addition, time headways of the vehicles in the queue statistically remained constant after the first four vehicles in the queue. Therefore, it can be concluded that saturation flow was obtained with the fifth vehicles in the queue. This result is consistent with some studies in the literature [1, 4, 11, 12].
- Time headways of the vehicles in the first 12 seconds of the green at the Kipa intersection and in the first 6 seconds of the green time at the CNR intersection are statistically larger than the time headways of the other vehicles. Therefore, the results of the queue length green time analyses confirm each other. Therefore, it was statistically revealed that the initial lost times encountered by the first vehicles in the queue have negative effect on the traffic flow.

Acknowledgement

The authors would like to thank Transportation Department of the Mersin Metropolitan Municipality for sharing the video recordings and signalization data.

Author contributions

Nihat Can Karabulut: Literature review, Software, Analysis **Murat Özen:** Conceptualization, Writing-Original draft preparation, Methodology **Oruç Altıntaşı:** Conceptualization, Writing-Original draft preparation, Methodology

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Khosla, K., & Williams, J. C. (2006). Saturation flow at signalized intersections during longer green time. Transportation Research Record, 1978 (1), 61-67.
2. Denney, Jr. R. W., Curtis, E. & Head, L. (2009). Long green times and cycles at congested traffic signals. Transportation Research Record, 2128 (1), 1-10.
3. Lin, F. B. & Thomas, D. R. (2005). Headway compression during queue discharge at signalized intersections. Transportation Research Record, 1920 (1), 81-85.
4. Stanić, B., Tubić, V. & Čelar. N. (2011). Straight lane saturation flow and its rate in Serbian cities. Transport, 26 (3), 329-334.
5. Chaudhry, M. S. & Ranjitkar, P. (2013). Capacity and signal timing analysis of signalized intersections with increasing saturation flow rate. Transportation Research Board 92nd Annual Meeting (No. 13-3396).
6. Dey, P. P., Nandal, S. & Kalyan, R. (2013). Queue discharge characteristics at signalised intersections under mixed traffic conditions. European Transport, 55(7), 1-12.

7. Tanyel, S., Koyuncu, M. & Çaliskanelli, S. P. (2018). Sürücü davranışlarının sinyalize kavşak başarımı üzerindeki etkisi. Teknik Dergi, 29 (5), 8563-8588.
8. Li, H. & Prevedouros, P. D. (2002). Detailed observations of saturation headways and start-up lost times. Transportation Research Record, 1802 (1), 44-53.
9. Day, C. M., Sturdevant, J. R., Li, H., Stevens, A., Hainen, A. M., Remias, S. M. & Bullock, D. M. (2013). Revisiting the cycle length: Lost time question with critical lane analysis. Transportation Research Record, 2355(1), 1-9.
10. Gao, L. & Alam, B. (2014). Optimal discharge speed and queue discharge headway at signalized intersections. In 93rd Annual Meeting of Transportation Research Board, CD-ROM. Washington, DC, United States.
11. Transportation Research Board (TRB) (2010). Highway capacity manual (5th Edition). Washington, D.C.: Transportation Research Board, National Research Council.
12. Cohen, S. L. (2002). Application of car-following systems to queue discharge problem at signalized intersections. Transportation Research Record, 1802(1), 205-213.

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Investigation of acceleration on non-structural building elements under earthquake effect

Mustafa Halûk Saraçoğlu*[1], Ahmet Özkaya[2]

[1]*Kütahya Dumlupınar University, Department of Civil Engineering, Türkiye*
[2]*Emet Municipality, Türkiye*

Keywords
Turkish Building Earthquake Code 2018
Non-structural building elements
Equivalent earthquake load
Spectral acceleration coefficient

Research Article
DOI: 10.31127/tuje.1021866

Received: 10.11.2021
Accepted: 25.01.2022
Published: 01.04.2022

Abstract

Non-structural elements of a building are often not considered in structural analysis. Therefore, due attention is not paid to them during the design process of the building. However, non-structural elements are exposed to seismic forces and therefore they are exposed to damage and failures that have psychological and economic consequences in some critical and vital structures during seismic events. Depending on the type of building the total cost of non-structural elements and the ratio of the project cost are many times higher than the cost of structural elements. In this study, the equivalent earthquake load was investigated within the scope of the design principles of non-structural building elements under the effect of the earthquake defined in Turkish Building Earthquake Code 2018.

1. Introduction

Earthquakes, which are caused by the shaking of the earth by seismic waves that occur as a result of unexpected energy in the Earth's crust, cause damage to structures according to their magnitude. Plastic deformations or collapses occur in columns, beams, floors and foundations of the building, which cannot bear the energy generated during an earthquake. If the structural elements damaged beyond the target level, the damage of the non-structural elements could cause loss of life, property and function.

The design of non-structural building elements was compulsory by the Turkish Building Earthquake Code 2018 (TBEC 2018) [1]. There is a lack of project controls and implementation supervision in our country about this subject area. Studies conducted on this important issue were continue in our country which is at the earthquake zone [2-3].

Atu [2] has studied about the architectural details damaged by earthquake and precautions to be taken in her master thesis in the year 2000. According to this study; it has concluded that precautions should be taken and that the subject is not sufficiently covered in the codes. Özkaya [4] evaluate the performance targets, measures and the approach of the TBEC 2018 to this issue in order to avoid the losses and negativities that may occur. Girgin [5] has explained the importance of seismic design of secondary systems due to the damage during several earthquakes and studies for dynamic analysis of secondary systems with the effect of dynamic and physical characteristics in his master thesis. Tüzün [6] has presented new approaches in seismic design of buildings and seismic protection methodologies for mechanical installations and emphasized that the earthquake protection of the installations is a multidisciplinary study and requires the interdisciplinary collaborations. İpek et al. [7] investigate the evaluation of non-structural systems in terms of earthquake effects in their study. İpek [8] have made the definition, classification and examination of non-structural elements in terms of dynamic behavior of a structure under the effect of earthquake loads in his study. Mısır et al. [9] studied about the reducing the damages of non-structural elements and performed dynamic tests on non-structural elements using a shaking table. Taş [10] investigate the earthquake

* **Corresponding Author**

*(mhaluk.saracoglu@dpu.edu.tr) ORCID ID 0000-0003-3842-5699
(ahmetozkaya0003@gmail.com) ORCID ID 0000-0002-8347-5436

Cite this article

Saraçoğlu, M. H. & Özkaya, A. (2023). Investigation of acceleration on non-structural building elements under earthquake effect. Turkish Journal of Engineering, 7(1), 56-63

behavior of non-structural wood interior equipment. Beşir and Dereci [11] investigate the risks and measures that can be taken by non-structural elements in housing indoor during earthquakes.

In this study, the equivalent earthquake load (which is abbreviated as F_{ie}) calculation for the non-structural element or equipment was studied within the scope of the design principles of non-structural building elements under the effect of the earthquake defined in TBEC 2018. By making calculations for some non-structural elements related to the subject, the variation of the equivalent earthquake load acting according to the location of the elements in the structure has been examined and the importance of the subject has been revealed. Even if the non-structural elements are the same element, the manufacturing details may be different, as they may receive different earthquake loads. The objective of this study is to reveal this situation by calculating the earthquake loads acting on non-structural building elements.

2. Material and Method

In TBEC 2018, Design Principles of Non-Structural Building Elements under the Effect of Earthquake are explained under the Section 6 (TBEC 2018).

With the performance-based approach, minimum damage will occur at the targeted performance level in structural and non-structural elements. Damage loss can also be minimized after the earthquake.

According to TBEC 2018; it is compulsory to make an earthquake analysis according to the rules given in Section 6, for;

- all kinds of protrusions (such as balcony, parapet, chimney, console) that connected to the main structural system but self-sustaining and which could harm people or the main structural system or prevent the use of the building in case of damage in an earthquake,
- facade and partition panels,
- architectural elements,
- mechanical and electrical equipment,
- and their connections to the structure.

However, it is not compulsory to make an earthquake analysis for the furniture in the building, the equipment temporarily located in the building and not connected to the building, and the non-structural elements in the buildings with Earthquake Design Class 4 (DTS-4).

Regarding element-based calculations also states that non-structural elements and equipment must be fixedly attached to the structure and the fasteners must have the capacity to meet the equivalent earthquake loads and displacements given in Section 6 of TBEC 2018. Additional capacity due to friction will not be taken into account in the calculation of the connection elements (weld, bolt, dowel, rivet, etc.) that connect the equipment to the structure under earthquake effect. Connection elements must have the strength that ensure the load transfer from equipment to the structure continuously.

The non-structural element or equipment should be considered as part of the building carrier system according to TBEC 2018 when the total weight is greater than 10% of the floor that locates. In this case, the mass of the non-structural element or equipment and the rigidity properties of its connection to the building will be considered in the earthquake analysis of the building carrier system.

The calculation of equivalent earthquake load acting horizontally to the center of gravity of the non-structural element or equipment and acting on the non-structural element or equipment with the use of Equation 6.1 of TBEC 2018 is as follows:

$$F_{ie}^{Eq6.1} = \left(m_e\ A_{ie}\ B_e\right)/R_e \tag{1}$$

In this equation m_e is the working mass of the non-structural element or equipment, A_{ie} is the maximum total acceleration affecting the area where the non-structural element or equipment on the floor i is connected to the floor under DD 2 earthquake ground motion level, B_e is the magnification factor applied to the non-structural element or equipment, Re is the behavior coefficient defined for the non-structural element or equipment. According to TBEC 2018, A_{ie} acceleration should be calculated separately using equivalent earthquake load, modal analysis and time history analysis methods and the maximum one should be used in the calculations.

The inequality about earthquake load acting on non-structural elements according to the Equation 6.5 of TBEC 2018 is as follows:

$$F_{ie} \geq F_{ie}^{Eq6.5} = 0.3\, m_e\ I\ S_{DS}\ g \tag{2}$$

According to TBEC 2018, equivalent earthquake load acting on non-structural elements calculated with the Equation 6.1 cannot be less than given with the Equation 6.5 of the TBEC-2018.

$$F_{ie}^{Eq6.1} \geq F_{ie}^{Eq6.5} \tag{3}$$

The flow chart of equivalent earthquake load calculation for the non-structural element or equipment is shown in Fig. 1.

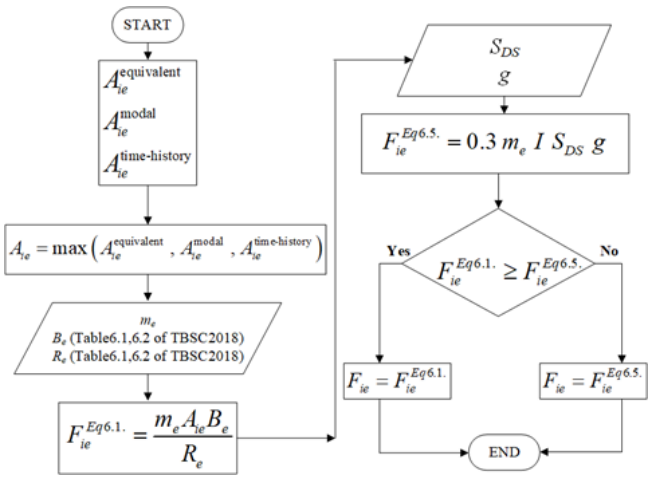


Figure 1. Flow chart of equivalent earthquake load calculation

3. Numerical study

In this numerical example; equivalent earthquake loads on various non-structural elements of nine-story building is investigated. There is a 1m^3 volume of water tank, chimney and a 1m height parapet in the roof floor. Main stairs continue from the first floor to the ninth floor. In the intervening floors, there are railings at the stairs, combi boilers in the kitchen, panel radiators in the rooms, suspended ceilings on the ceilings and facade coatings up to the first floor.

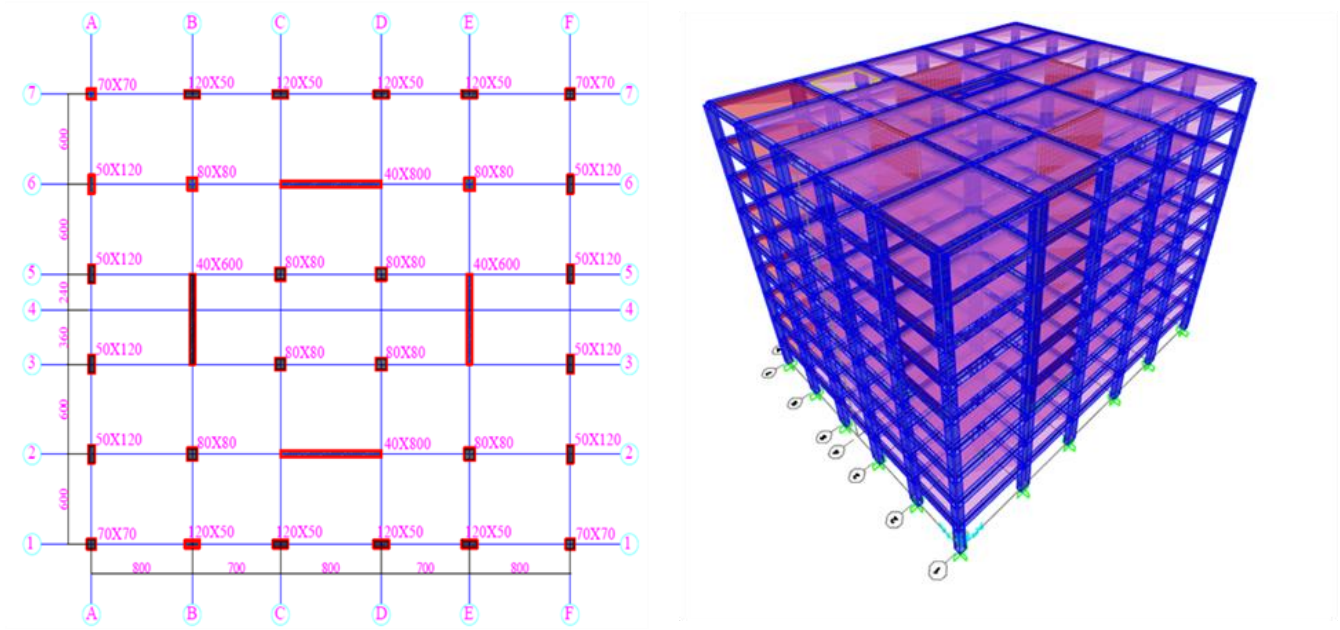


Figure 2. Column and shear wall plan and 3d view of the building

The column and shear wall plan and three-dimensional view of the sample building analyzed by using SAP 2000 v21 are shown in Fig. 2.

Parameters used in the numerical example is shown in Table 1.

Table 1. Parameters used in the analysis

Parameter	Value
Behavior coefficient of carrier system (R)	7
Coefficient of building importance (I)	1
Height of the building as meters	31.5
Building height class (BYS)	4
Building usage class (BKS)	3
Earthquake design class (DTS)	1

In all elements of the example the same material properties is defined. Concrete class is C30/37, E_c=32 GPa, F_{ck}=30 MPa and the reinforcement class is B420C, E_y=200 Gpa, F_{yk}=420 MPa.

The local soil class was taken as ZD and the earthquake ground motion level as DD-2 from the AFAD (Disaster and Emergency Management Presidency) earthquake map belonging to the coordinates of the ground values (Latitude: 39.344763° Longitude 29.26084°) on which the example model will be constructed. Ground data are shown in Table 2.

Table 2. Ground data

Parameter	**Value**
S_S	0.850
S_1	0.201
S_{DS}	0.986
S_{D1}	0.442
PGA (g)	0.352
PGV (cm/sn)	19.618

Table 3. Earthquake records

Number	Acceleration Records	Stations	Latitude	Longitude	Depth	Type	magnitude
1	20070920061911_4302_ap_Acc_E	4302	39.2147	29.3885	16.90	MW	4.4
2	20130609141856_4305_ap_Acc_E	4305	39.1392	29.0220	15.61	ML	4.1
3	20200125174947_4310_ap_Acc_E	4310	39.0293	27.8430	7.84	MW	4.1
4	20201030115124_4304_mp_Acc_E	4304	37.8790	26.7030	14.90	MW	6.6
5	20201211143749_4306_mp_Acc_E	4306	39.9901	28.1961	7.00	MW	4.1
6	20201211143749_4314_mp_Acc_E	4314	39.9901	28.1961	7.00	MW	4.1
7	20210209155154_4307_mp_Acc_E	4307	38.5965	31.6318	7.01	MW	4.7
8	20210209155154_4312_mp_Acc_E	4312	38.5965	31.6318	7.01	MW	4.7
9	20210209155154_4313_mp_Acc_E	4313	38.5965	31.6318	7.01	MW	4.7
10	20210214210815_4311_mp_Acc_E	4311	38.1878	30.0456	6.05	MW	3.9
11	20210630030010_4301_mp_Acc_E	4301	39.0295	29.6541	7.01	MW	3.8

A total of 11 earthquake acceleration records were taken from stations 4302-4305-4310-4304-4306-4314-4307-4312-4313-4311 and 4301 for analysis according to the Mode Addition Method in the Time History Domain. The earthquake acceleration records given in Table 3 are arranged by scaling.

According to TBEC 2018, the maximum total acceleration acting on the element or equipment will be defined as the maximum one among the accelerations calculated according to the Equivalent Earthquake Load Method, Modal Analysis and Time History Analysis.

The natural periods for the first ten modes of the presented example are shown in Table 4.

The accelerations were calculated separately by using different methods of TBEC-2018. The accelerations to be used in the earthquake calculation of non-structural elements in X direction are tabulated in Table 5.

Table 4. Natural periods (sec)

Mode Number	Natural Periods
1	1.55188
2	1.30825
3	1.30219
4	0.37947
5	0.32502
6	0.29713
7	0.16634
8	0.14508
9	0.13159
10	0.11882

The calculated accelerations for Y direction, to be used in the earthquake simulations of non-structural elements are summarized in Table 6.

Table 5. The A_{ie} accelerations in X direction to be used in the analysis of non-structural elements (m/sn²)

Story	Equivalent Earthquake Load	Modal Analysis	Time History Analysis	Selected Acceleration
1	0.29	0.26	-0.18	0.29
2	0.96	0.50	-0.10	0.96
3	1.86	0.57	-0.22	1.86
4	2.93	0.51	-0.30	2.93
5	4.10	0.37	0.05	4.10
6	5.32	0.30	-0.01	5.32
7	6.56	0.41	0.05	6.56
8	7.78	0.81	0.13	7.78
9	8.95	1.80	0.69	8.95

Table 6. The A_{ie} accelerations in Y direction to be used in the analysis of non-structural elements (m/sn²)

Story	Equivalent Earthquake Load	Modal Analysis	Time History Analysis	Selected Acceleration
1	0.25	0.25	-0.21	0.25
2	0.82	0.49	-0.25	0.82
3	1.59	0.53	0.14	1.59
4	2.48	0.53	0.09	2.48
5	3.44	0.27	-0.09	3.44
6	4.42	0.19	-0.06	4.42
7	5.37	0.30	0.01	5.37
8	6.28	0.71	0.12	6.28
9	7.14	1.71	0.59	7.14

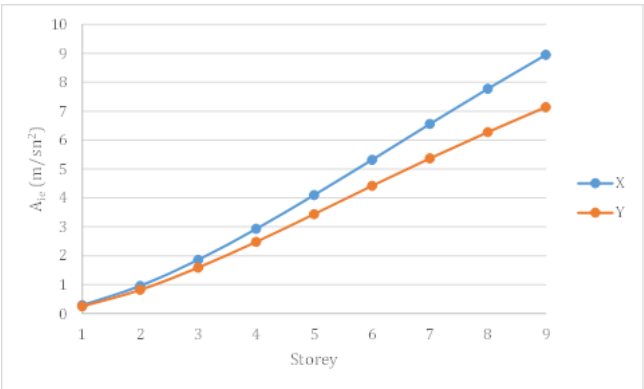


Figure 3. Selected A_{ie} accelerations

As seen from the Fig. 3, selected acceleration is greater in the X direction. Calculations for non-structural building elements were made separately for both the x-direction and the y-direction.

With the Equation 6.5 of TBEC 2018 as expressed in Eq. 2, it has been explained that restrictions on non-structural building elements is imposed.

Calculation of the equivalent earthquake loads in every floor for non-structural building elements as suspended ceiling, chimney, facade coating, combi boiler, railing, panel radiator, parapet and water tank for x direction and y direction is shown in Tables 7 and 8 respectively.

Table 7. Equivalent earthquake loads on various non-structural building elements in the X direction

Story	Element	m_e	B_e	R_e	A_{ie} (m/sn²)	F_{ie} Eq 6.1(N)	F_{ie} Eq 6.5(N)	F_{ie} FINAL(N)
9th FLOOR	RAILING	7	2,5	2,5	8,95	62,65	20,31	62,65
	WATER TANK	1000	2,5	6	8,95	3729,17	2901,80	3729,17
	CHIMNEY	300	2,5	2,5	8,95	2685,00	870,54	2685,00
	PARAPET	550	2,5	2,5	8,95	4922,50	1595,99	4922,50
8th FLOOR	RAILING	7	2,5	2,5	7,78	54,46	20,31	54,46
	COMBI BOILER	33	1	2,5	7,78	102,70	95,76	102,70
	PANEL RADIATOR	25	2,5	6	7,78	81,04	72,54	81,04
	SUSPENDED	27	1	2,5	7,78	84,02	78,35	84,02
	FACADE COATING	11	1	1,5	7,78	57,05	31,92	57,05
7th FLOOR	RAILING	7	2,5	2,5	6,56	45,92	20,31	45,92
	COMBI BOILER	33	1	2,5	6,56	86,59	95,76	95,76
	PANEL RADIATOR	25	2,5	6	6,56	68,33	72,54	72,54
	SUSPENDED	27	1	2,5	6,56	70,85	78,35	78,35
	FACADE COATING	11	1	1,5	6,56	48,11	31,92	48,11
6th FLOOR	RAILING	7	2,5	2,5	5,32	37,24	20,31	37,24
	COMBI BOILER	33	1	2,5	5,32	70,22	95,76	95,76
	PANEL RADIATOR	25	2,5	6	5,32	55,42	72,54	72,54
	SUSPENDED	27	1	2,5	5,32	57,46	78,35	78,35
	FACADE COATING	11	1	1,5	5,32	39,01	31,92	39,01
5th FLOOR	RAILING	7	2,5	2,5	4,1	28,70	20,31	28,70
	COMBI BOILER	33	1	2,5	4,1	54,12	95,76	95,76
	PANEL RADIATOR	25	2,5	6	4,1	42,71	72,54	72,54
	SUSPENDED	27	1	2,5	4,1	44,28	78,35	78,35
	FACADE COATING	11	1	1,5	4,1	30,07	31,92	31,92
4th FLOOR	RAILING	7	2,5	2,5	2,93	20,51	20,31	20,51
	COMBI BOILER	33	1	2,5	2,93	38,68	95,76	95,76
	PANEL RADIATOR	25	2,5	6	2,93	30,52	72,54	72,54
	SUSPENDED	27	1	2,5	2,93	31,64	78,35	78,35
	FACADE COATING	11	1	1,5	2,93	21,49	31,92	31,92
3rd FLOOR	RAILING	7	2,5	2,5	1,86	13,02	20,31	20,31
	COMBI BOILER	33	1	2,5	1,86	24,55	95,76	95,76
	PANEL RADIATOR	25	2,5	6	1,86	19,38	72,54	72,54
	SUSPENDED	27	1	2,5	1,86	20,09	78,35	78,35
	FACADE COATING	11	1	1,5	1,86	13,64	31,92	31,92
2nd FLOOR	RAILING	7	2,5	2,5	0,96	6,72	20,31	20,31
	COMBI BOILER	33	1	2,5	0,96	12,67	95,76	95,76
	PANEL RADIATOR	25	2,5	6	0,96	10,00	72,54	72,54
	SUSPENDED	27	1	2,5	0,96	10,37	78,35	78,35
	FACADE COATING	11	1	1,5	0,96	7,04	31,92	31,92
1st FLOOR	RAILING	7	2,5	2,5	0,29	2,03	20,31	20,31
	COMBI BOILER	33	1	2,5	0,29	3,83	95,76	95,76
	PANEL RADIATOR	25	2,5	6	0,29	3,02	72,54	72,54
	SUSPENDED	27	1	2,5	0,29	3,13	78,35	78,35
	FACADE COATING	11	1	1,5	0,29	2,13	31,92	31,92

Equivalent earthquake loads for all non-structural elements are calculated for each floor in the relevant rows in Tables 7 and 8. Again, in the column F_{ie} calculated in the aforementioned tables, if the condition of Equation 6.5 of TBEC-2018 is not satisfied, the minimum value that satisfies the condition of the equation is specified in the next column. The appropriate values for the equivalent earthquake loads calculated from Equation 6.5 of TBEC-2018 are shown in red in the last column. While calculations were made for standard five non-structural elements on the eight floors, on the ninth floor, which is the roof floor, calculations were made for three elements other than these elements, such as the parapet, chimney and water tank.

Table 8. Equivalent earthquake loads on various non-structural building elements in the Y direction

Story	Element	m_e (kg)	B_e	R_e	A_{ie} (m/sn²)	F_{ie} Eq 6.1 (N)	F_{ie} Eq 6.5 (N)	F_{ie} FINAL (N)
9th FLOOR	RAILING	7	2,5	2,5	7,14	49,98	20,31	49,98
	WATER TANK	1000	2,5	6	7,14	2975,00	2901,80	2975,00
	CHIMNEY	300	2,5	2,5	7,14	2142,00	870,54	2142,00
	PARAPET	550	2,5	2,5	7,14	3927,00	1595,99	3927,00
8th FLOOR	RAILING	7	2,5	2,5	6,28	43,96	20,31	43,96
	COMBI BOILER	33	1	2,5	6,28	82,90	95,76	95,76
	PANEL RADIATOR	25	2,5	6	6,28	65,42	72,54	72,54
	SUSPENDED CEILING	27	1	2,5	6,28	67,82	78,35	78,35
	FACADE COATING	11	1	1,5	6,28	46,05	31,92	46,05
7th FLOOR	RAILING	7	2,5	2,5	5,37	37,59	20,31	37,59
	COMBI BOILER	33	1	2,5	5,37	70,88	95,76	95,76
	PANEL RADIATOR	25	2,5	6	5,37	55,94	72,54	72,54
	SUSPENDED CEILING	27	1	2,5	5,37	58,00	78,35	78,35
	FACADE COATING	11	1	1,5	5,37	39,38	31,92	39,38
6th FLOOR	RAILING	7	2,5	2,5	4,42	30,94	20,31	30,94
	COMBI BOILER	33	1	2,5	4,42	58,34	95,76	95,76
	PANEL RADIATOR	25	2,5	6	4,42	46,04	72,54	72,54
	SUSPENDED CEILING	27	1	2,5	4,42	47,74	78,35	78,35
	FACADE COATING	11	1	1,5	4,42	32,41	31,92	32,41
5th FLOOR	RAILING	7	2,5	2,5	3,44	24,08	20,31	24,08
	COMBI BOILER	33	1	2,5	3,44	45,41	95,76	95,76
	PANEL RADIATOR	25	2,5	6	3,44	35,83	72,54	72,54
	SUSPENDED CEILING	27	1	2,5	3,44	37,15	78,35	78,35
	FACADE COATING	11	1	1,5	3,44	25,23	31,92	31,92
4rd FLOOR	RAILING	7	2,5	2,5	2,48	17,36	20,31	20,31
	COMBI BOILER	33	1	2,5	2,48	32,74	95,76	95,76
	PANEL RADIATOR	25	2,5	6	2,48	25,83	72,54	72,54
	SUSPENDED CEILING	27	1	2,5	2,48	26,78	78,35	78,35
	FACADE COATING	11	1	1,5	2,48	18,19	31,92	31,92
3rd FLOOR	RAILING	7	2,5	2,5	1,59	11,13	20,31	20,31
	COMBI BOILER	33	1	2,5	1,59	20,99	95,76	95,76
	PANEL RADIATOR	25	2,5	6	1,59	16,56	72,54	72,54
	SUSPENDED CEILING	27	1	2,5	1,59	17,17	78,35	78,35
	FACADE COATING	11	1	1,5	1,59	11,66	31,92	31,92
2nd FLOOR	RAILING	7	2,5	2,5	0,82	5,74	20,31	20,31
	COMBI BOILER	33	1	2,5	0,82	10,82	95,76	95,76
	PANEL RADIATOR	25	2,5	6	0,82	8,54	72,54	72,54
	SUSPENDED CEILING	27	1	2,5	0,82	8,86	78,35	78,35
	FACADE COATING	11	1	1,5	0,82	6,01	31,92	31,92
1st FLOOR	RAILING	7	2,5	2,5	0,25	1,75	20,31	20,31
	COMBI BOILER	33	1	2,5	0,25	3,30	95,76	95,76
	PANEL RADIATOR	25	2,5	6	0,25	2,60	72,54	72,54
	SUSPENDED CEILING	27	1	2,5	0,25	2,70	78,35	78,35
	FACADE COATING	11	1	1,5	0,25	1,83	31,92	31,92

As an example, for the calculation of equivalent earthquake load for a railing in x direction in first floor with the use of Equation 6.1 of TBEC 2018 is as follows:

$$F_{ie}^{Eq6.1} = \left(7 \times 0.29 \times 2.5\right) / 2.5 = 2.03 \text{ N} \qquad (4)$$

The equivalent earthquake load acting on non-structural elements according to the Equation 6.5 of TBEC 2018 is as follows:

$$F_{ie}^{Eq6.5} \geq 0.3 \times 7 \times 1 \times 0.986 \times 9.81 = 20.31 \text{ N} \qquad (5)$$

According to TBEC 2018, equivalent earthquake load acting on non-structural elements calculated with the Equation 6.1 of TBEC-2018 cannot be less than the Equation 6.5 of TBEC-2018.

$$F_{ie}^{Eq6.5} = 20.31 \text{ N} \geq F_{ie}^{Eq6.1} = 2.03 \text{ N} \qquad (6)$$

From such value, 20.31 N is selected as an equivalent earthquake load for a railing (a non-structural building element) for the first floor.

As another example on calculation of equivalent earthquake load in y direction for a facade coating on the seventh floor with the use of Equation 6.1 of TBEC 2018 is as follows:

$$F_{ie}^{Eq6.1} = \left(11 \times 5.37 \times 1\right) / 1.5 = 39.38 \text{ N} \qquad (7)$$

The equivalent earthquake load acting on non-structural elements according to the Equation 6.5 of TBEC 2018 is as follows:

$$F_{ie}^{Eq6.5} \geq 0.3 \times 1 \times 1 \times 0.986 \times 9.81 = 31.92 \text{ N} \quad (8)$$

According to TBEC 2018, equivalent earthquake load acting on non-structural elements calculated with the Equation 6.1 cannot be less than the value obtained from the Equation 6.5 of TBEC-2018.

$$F_{ie}^{Eq6.5} = 31.92 \text{ N} < F_{ie}^{Eq6.1} = 39.38 \text{ N} \quad (9)$$

As a result, 39.38 N is selected as an equivalent earthquake load of a facade coating (a non-structural building element) in the seventh floor.

In Tables 7 and 8, if $F_{ie}^{Eq\ 6.5} \geq F_{ie}^{Eq\ 6.1}$, loads in red indicate the selected equivalent earthquake load.

Calculated equivalent earthquake loads for every floor of facade coatings and railings as non-structural element are shown in Fig. 4.

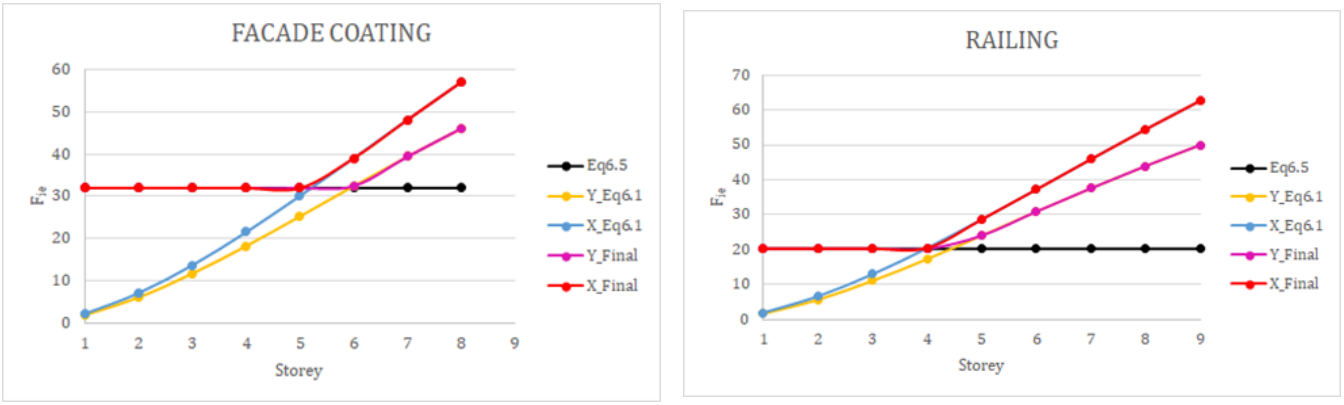


Figure 4. F_{ie} loads for facade coatings and railings for every floor

4. Discussion

The coefficient A_{ie} is directly proportional to the acceleration and displacement that acting on non-structural elements. The equivalent earthquake load acting on non-structural elements is also directly proportional to the coefficient A_{ie}, namely acceleration. In this case, since the displacements acting on the floors are different, even if the same non-structural element is the same, different accelerations and earthquake loads will be affected due to different displacements.

It can be seen from the Table 7 that equivalent earthquake load of the railing (a non-structural building element) does not satisfy the limits of the Equation 6.5 of TBEC 2018 for the first, second and third floors of the building. However, for the other floors of the building it can be seen that the limits are satisfied.

As seen from this example, especially in buildings with a high number of stories the earthquake load that affects to non-structural elements will increase.

For nonstructural element facade coating, from the Fig. 4 it can be seen while the equivalent earthquake forces in y direction were those calculated from Equation 6.5 of TBEC-2018 up to sixth floor, for the stories above the sixth floor Equation 6.1 of TBEC-2018 is preferred. Equivalent earthquake forces in the x direction, like y direction, they were calculated from Equation 6.5 of TBEC-2018 for the stories up to the fifth floor and for the above Equation 6.1 of TBEC-2018 is adopted.

5. Conclusion

Equivalent earthquake loads must be determined for the design of non-structural elements. Even if the non-structural elements are the same, on different locations on building the earthquake loads acting on them may be different. These loads used in the analysis of non-structural elements are calculated using the accelerations.

According to TBEC 2018, these accelerations should be calculated with three different methods and the maximum one should be used in the design. These three methods are Equivalent Earthquake Load Method, Modal Analysis and Time History Analysis respectively.

In this study, a nine-story building with different types of non-structural elements is considered as an example. The maximum total acceleration affecting the area where the non-structural element or equipment on the floor i is connected to the floor under DD 2 earthquake ground motion level is calculated with three methods. Equivalent earthquake load is determined by using the maximum of these accelerations.

Equivalent earthquake load was found to be conservative but governing case in determination of designing the non-structural elements in buildings.

The results of this study also demonstrated that the attention should be paid for the connection, fixing or assembly details. Because of the equivalent earthquake loads acting on them, they could be different during the projecting and manufacturing phase.

The cost of buildings today is quite high. The ratio of the project cost of non-structural elements to the project cost of structural elements is significantly high. The total

cost of non-structural elements depending on the type of building (hospital, commercial building, residence, industry, etc.), mechanical-electrical installations, architectural elements, and other items and equipment are many times higher than the cost of structural elements. The cost of non-structural elements damaged during the earthquake is high.

Damages occurring in non-structural elements of manufacturing buildings such as factories after an earthquake may also cause the interruption of working activities. This situation will adversely affect the development of the production and service sectors as well as the society and the psychosocial life of individuals.

Author contributions

Mustafa Halûk Saraçoğlu: Conceptualization, Methodology, Writing-Original draft preparation, Investigation, Writing-Reviewing and Editing. **Ahmet Özkaya:** Methodology, Software Data curation, Software, Validation. Visualization, Investigation

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Turkish Building Earthquake Code (TBEC-2018), Afet ve Acil Durum Yönetimi Başkanlığı, Resmi Gazete, Sayı: 30364 (Mükerrer), 18 Mart 2018
2. Atu, S. (2000). Architectural details damaged by earthquake and precautions to be taken. Doctoral Thesis, İstanbul Technical University, Institute of Science, İstanbul, 81p.
3. Büyükkaragöz, A. & Cantürk, R. (2018). The behaviour of non-structural elements under earthquake effect in industrial buildings. GU J Sci, Part C, 6(2), 426-434.
4. Özkaya, A. (2021). Design of non-structural building elements. Master's Thesis, Kütahya Dumlupınar University, Institute of Science, Kütahya, 67p
5. Girgin, S. C. (2008). Dynamic analysis of nonstructural (secondary) systems. Master's Thesis, Dokuz Eylül University, Institute of Science, İzmir, 91p
6. Tüzün, C. (2019). New Approaches in Seismic Design of Buildings and Seismic Protection Methodologies for Mechanical Installations. 14th National Congress on Installation Engineering, 987-995, İzmir, Turkey.
7. İpek C, Kuzucuoğlu, A. H. & Kıstır, M. R. (2015). Evaluatıon of nonstructual systems ın terms of earthquake effects. International Burdur Earthquake & Environment Symposium (IBEES2015), 197-206, Burdur, Turkey.
8. İpek, C. (2015). Analysis of non-structural systems under earthquake effect. 5th International Earthquake Symposium, 187-199, Kocaeli, Turkey.
9. Mısır, İ. S., Özçelik, Ö., Girgin, S. C., Kahraman, S. & Baran, T. (2010). Reducing vulnerability of non-structural elements and shaking table tests. İMO Newsletter, 25(152), 24-26.
10. Taş, H. H. (2010). Investigation of earthquake behavior of non- structural wood interior equipment. Doctoral Thesis, Süleyman Demirel University, Institute of Science, Isparta, 153p.
11. Beşir, Ş. E. & Dereci, Ş. (2021). Risks and measures that can be taken by non-structural elements in housing indoor during earthquakes. International Social Mentality and Researcher Thinkers Journal, 7(42), 350-360.

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Application of machine learning algorithms in the investigation of groundwater quality parameters over YSR district, India

Jagadish Kumar Mogaraju[*1]

[1]*Indira Gandhi National Open University, Life Sciences, Kadapa, India*

Keywords
Groundwater
Machine learning
Support Vector Machines
Classification Trees
Random Forest

Research Article
DOI: 10.31127/tuje.1032314

Received: 04.12.2021
Accepted: 02.03.2022
Published: 01.04.2022

Abstract

Human life sustained for decades due to the availability of basic needs, and freshwater is one of them. However, groundwater quality is constantly under pressure. This can be attributed to anthropogenic activities not limited to urban areas but to rural zones. Machine learning methods like linear discriminant analysis (LDA), Classification and Regression Trees (CART), k-Nearest Neighbour (KNN), Support Vector Machines (SVM) and, Random Forest (RF) models were used to analyse groundwater quality variables. The mean accuracy of each classifier was calculated, and the obtained mean accuracies were 77.5% (LDA), 87% (CART), 96% (KNN), 93.5% (SVM) and 96% (RF). RF and KNN models were selected as optimal models with higher accuracy. This study made it apparent that machine learning algorithms can estimate and predict water quality variables with significant accuracy. In this study, the observations and variables were compared with the water quality index and drinking water limits provided by the Bureau of Indian Standards. The water quality index for each observation was calculated. If at least four variables have a higher value than prescribed limits, it was assigned a value of 1; if more than four variables reported higher values, it was assigned a value of 2.

1. Introduction

Machine learning seeks to predict an outcome by extracting patterns from big datasets, usually in the form of an algorithm [1]. Machine learning is an advanced tool to understand groundwater quality variables over a study area [2]. Machine learning tools were used in the planning of several irrigation projects all over the world [3]. Mining is prevalent in the study area and is considered an essential economic source [4]. The irrigation, salinization, ion exchange, carbon dissolution and weathering processes can affect the groundwater quality. Some of them can be due to anthropogenic activities [5]. This region's deterioration of groundwater quality is mainly due to overexploitation and contamination [6]. Agriculture has been an essential economic source in India, and 60% of the populace depend on it for livelihood [7]. The quality of groundwater can be due to the percolation and infiltration of pollutant-laden rainwater; however, domestic and agricultural activities get involved in some places [8]. Crystallines, shale, limestone and quartzite are some litho units that can affect groundwater quality [9-10]. Approximately 91000 hectares of land in the study area are irrigated by water from the local canals. One thousand three hundred sixty-eight minor irrigation tanks irrigate 47000 hectares [9]. The rise in the groundwater level was 2.11m in alluvium, 2.50m in limestones, 3.82m in shales, 5.35m in crystallines and 7.32 in quartzites [9]. The groundwater quality contamination due to nitrates and pesticides was studied using machine learning models like Extreme Gradient Boosting (XGB), Support Vector Machines (SVM) and Artificial Neural Networks (ANN) and were evaluated for their accuracy [11]. Support Vector Regression (SVR), Artificial Neural Network (ANN), Random Forest (RF), and Adaptive Boosting (Adaboost) models were used to forecast and evaluate water quality indexes [3]. The ensemble models of RF and Boosted Regression Trees (BRT) were investigated with Multivariate Discriminative Analysis (MDA) [12]. SVM, MDA and BRT models were used in installing a framework for evaluating nitrate contamination in groundwater [13]. Artificial Intelligence techniques like SVM, Naïve Bayes classifier and Particle Swarm Optimization (PSO) were used to predict the water quality index [14]. ANN models

* **Corresponding Author**
*(jagadishmogaraju@gmail.com) ORCID ID 0000-0002-6461-8614

Cite this article
Mogaraju, J. K. (2023). Application of machine learning algorithms in the investigation of groundwater quality parameters over YSR district, India. Turkish Journal of Engineering, 7(1), 64-72

were compared with GIS tools to delineate the potential zones of groundwater in Ethiopia [15]. SVM, Multivariate Adaptive Regression Splines (MARS), k-nearest neighbour (KNN), ANN, RF, BRT, penalized discriminant analysis (PDA), flexible discriminant analysis (FDA), quadratic discriminant analysis (QDA) and linear discriminant analysis (LDA) models were compared in combination with GIS tools to assess groundwater quality in Iran [16]. RF, C5.0 and MARS models integrated with GIS were used in potential groundwater mapping [17]. The collected feature vectors were subjected to machine learning-based feature selection to determine the best feature sets for predicting soil water content [18]. Determining the soil surface humidity in vegetated areas is problematic; hence, polarimetric decomposition models and machine learning-based regression models were used to solve the problem [19].

Machine Learning has been one of the most quickly evolving areas in Artificial Intelligence research. Decision trees are highly effective and straightforward to understand. On the other hand, individual trees can be susceptible to slight data changes. Even greater prediction can be accomplished by using this variability to develop many trees from the same data [20]. Several Machine Learning algorithms have acquired popularity in part owing to their transparency. The Decision Tree algorithm, also known as the Classification and Regression Trees (CART) algorithm, is one of them. The CART algorithm is a classification algorithm used to construct a decision tree using Gini's impurity index.

The main purpose of this study is to investigate and report the prediction accuracies of the variables of groundwater quality using machine learning algorithms. The primary contribution of this study is to extend the role of machine learning in understanding the subsurface hydrology parameters and their characteristics on spatial domain.

2. Method

2.1. Data collection

The datasets used in this study were collected from the Central Ground Water Board (CGWB), Government of India. These datasets can be accessed at [21-22]. The shapefile used in producing location map of the study area was downloaded from the Website of Geodata, The University of Texas at Austin [23]. The location map of the study area is revealed in Figure 1.

The water samples were collected from 56 places, and a Physico-chemical analysis was performed. The water analysis was focused on Bicarbonates (HCO_3), Calcium (Ca), Chloride (Cl), Electric Conductivity (EC), Fluoride (F), Magnesium (Mg), Sodium (Na), Nitrates (NO_3), pH, Residual Sodium (RSC), SAR, SO_4, Total Hardness (TH) and Total Alkalinity. These variables were compared with the drinking water limits proposed by the Bureau of Indian Standards (BIS) (IS 10500:2012). The values of every variable were compared with the BIS limits and then assigned labels like 'N' (Normal) and 'H' (High). If at a given observation, the number of 'H's are ≤ 4, then it was assigned a value of '1' (Manageable), and if the number of 'H's are >4, then a value of '2' (High) was assigned. In order to pass the data onto the machine learning tools, the factor levels were kept to a minimum of 2. Unfortunately, some of the data is unavailable. Instead of opting for data imputation techniques, Inverse Distance Weighting was used to compensate for the missing values. 'R' version 4.1.1 with packages like 'caret,' 'mlbench,'randomForest' and other packages that were integrated with caret packages were used in this analysis. The data was split into 'training' and 'test' data. 80% of the observations (rows) were used as training data, and the remaining 20% was used as test or validation data. One linear algorithm (LDA), two non-linear algorithms (CART & KNN) and advanced algorithms (SVM & RF) were used. These algorithms were run using 10-fold cross-validation. Accuracy and Kappa values were obtained.

This work was framed for classification. Hence, accuracy was mainly considered to select the appropriate model instead of R^2 and RMSE values. Mean accuracy post comparing five models were considered to select the appropriate model. The skill of the selected model was estimated using test data. A confusion matrix was prepared using both training and test or validation data. Fourteen numerical variables and a 1-factor variable with two levels, i.e., '1' and '2', were used. The descriptive statistics are represented in table 1.

2.2. Study area

Waterlogging is caused by the intensive use of surface water in irrigation project command areas. The increased use of groundwater for agriculture, industry, and home purposes produces ongoing depletion of water levels, well drying, and water quality issues [24]. Thus, water resource management is necessary to protect aquifers and ensure they continue to provide water at a reasonable cost. In the drought-prone Cuddapah district, integrated geological, hydrological (surface and groundwater), and geochemical elements have been researched to develop and manage water resources. Crystallites, quartzites, shales, and limestones are the primary lithological units. Canal water irrigates about 91 000 acres of land in the Cuddapah area. In addition, 1368 minor irrigation tanks irrigate a registered ayacut of roughly 47 000 ha [9].

In the entire district, 503 spring channels originating from rivers/streams have been identified, with the ability to irrigate around 8700 acres. In quartzites, 5.35 m (crystallines), 3.82m (Shales), 2.50m (Limestone), and 2.11m in alluvium, the average seasonal rise in groundwater level is 7.32 m, 5.35 m(crystallines,) 3.82 m (shales), 2.50m in limestones, and 2.11m in alluvium [9]. Large amounts of groundwater are accessible in mining sites, which can be used and managed appropriately by the irrigation department/cultivators. According to groundwater assessment studies, the district has 584 million m^3 of groundwater accessible for future irrigation [8]. According to chemical analysis, the groundwater quality in various rock units is within legal limits for irrigation and residential use; however, specific conductance, chloride, and fluoride levels are high in a few spots. This could be due to untreated effluents, a faulty drainage system, or fertilizer application [25].

Table 1. Descriptive statistics

		Valid	Mean	Std. Deviation	Minimum	Maximum
HCO3	1	32	230.650	91.415	28.418	391.530
HCO3	2	24	441.555	203.365	213.751	972.903
Ca	1	32	39.455	19.418	7.865	112.000
Ca	2	24	79.923	30.369	41.270	162.424
Cl	1	32	142.035	81.228	19.883	350.685
Cl	2	24	526.148	812.553	50.012	4220.785
EC	1	32	960.865	399.591	137.111	1662.483
EC	2	24	2763.739	3219.268	1064.999	17435.179
F	1	32	0.497	0.253	0.056	0.983
F	2	24	0.897	0.502	0.370	2.147
Mg	1	32	29.314	11.706	3.985	51.672
Mg	2	24	61.970	33.765	25.642	172.373
Na	1	32	117.287	72.153	12.436	292.843
Na	2	24	427.485	711.630	96.625	3645.152
NO3	1	32	26.125	20.753	5.435	116.000
NO3	2	24	74.934	109.238	0.776	558.252
pH	1	32	3.432	1.381	0.604	7.410
pH	2	24	5.467	1.310	3.046	7.945
RSC	1	32	1.415	0.989	0.159	3.621
RSC	2	24	4.197	2.690	1.000	12.858
SAR	1	32	2.134	1.240	0.257	4.522
SAR	2	24	6.603	8.958	1.418	45.873
SO4	1	32	64.895	33.530	6.859	131.419
SO4	2	24	206.106	299.020	53.558	1564.692
TH	1	32	219.010	85.925	36.014	371.911
TH	2	24	454.290	209.380	212.140	1073.733
TA	1	32	190.081	74.262	23.293	320.926
TA	2	24	372.755	185.069	196.160	961.798

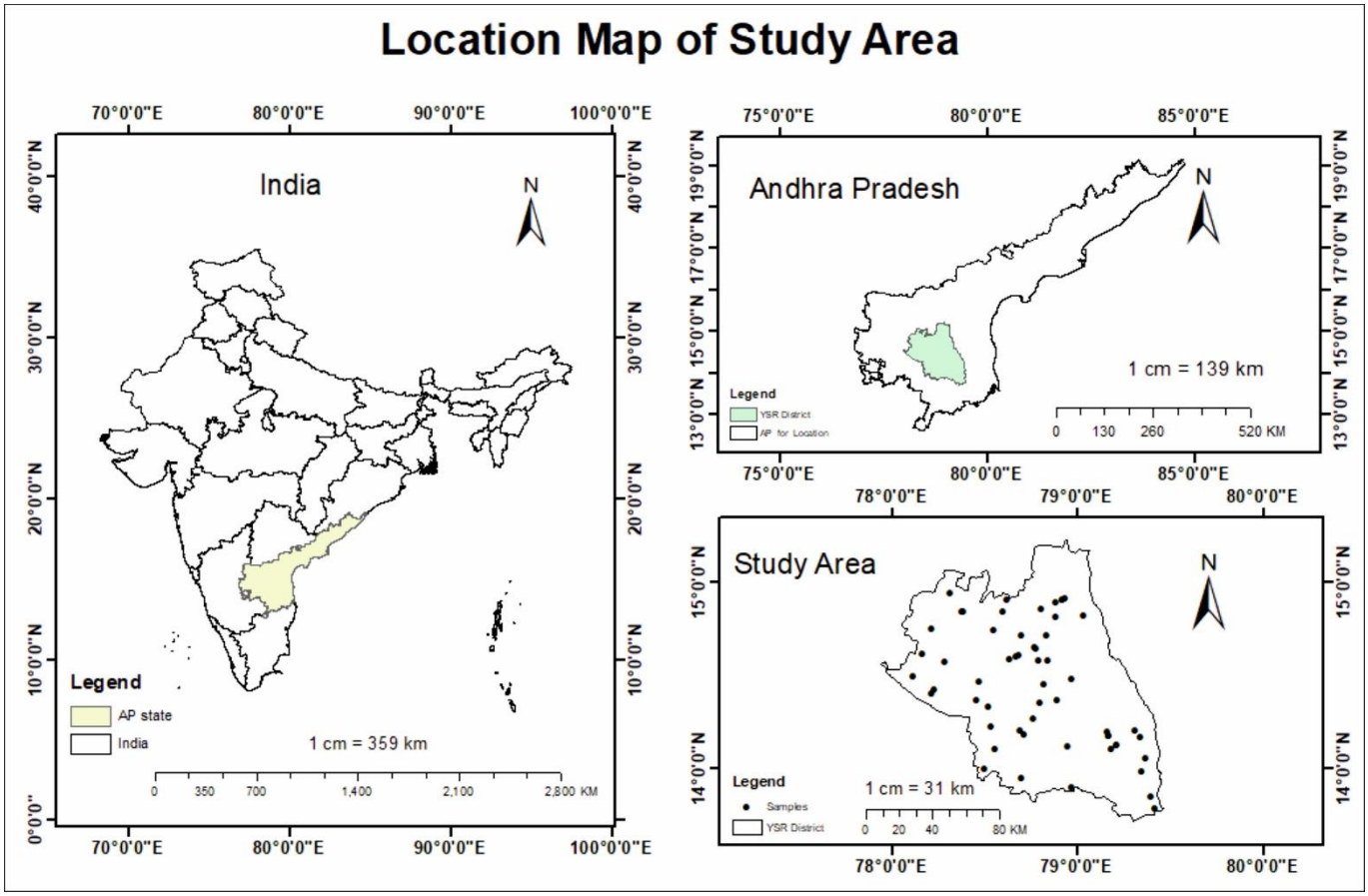


Figure 1. Location map of the study area

2.3. Methodology

The detailed methodology is shown in figure 2. In the pre-processing step, linear discriminant analysis (LDA) is the most often used dimensionality reduction technique. The target is to project a dataset onto a lower-dimensional space with excellent class-separability to minimize overfitting ("dimensionality's curse") and reduce computational costs. Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) are linear transformation approaches for dimensionality reduction. PCA is an "unsupervised" technique. It ignores class labels and aims to find the directions (known as principal components) that maximize a dataset's variance. Unlike PCA, LDA is "supervised," which means it calculates the directions ("linear discriminants") that will represent the axes that maximize the separation between several classes. Although it may appear logical that LDA is superior to PCA for multi-class classification tasks with known class labels, this is not necessarily the case.

Data classification can be done in a variety of ways. Two widely used data categorization and dimensionality reduction techniques are Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). When the within-class frequencies are unequal, and their performances have been tested on randomly produced test data, it is handled using Linear Discriminant Analysis. This approach maximizes the ratio of between-class variation to within-class variance in each given data set, ensuring maximum separability. The classification challenge in speech recognition is tackled with the help of Linear Discriminant Analysis.

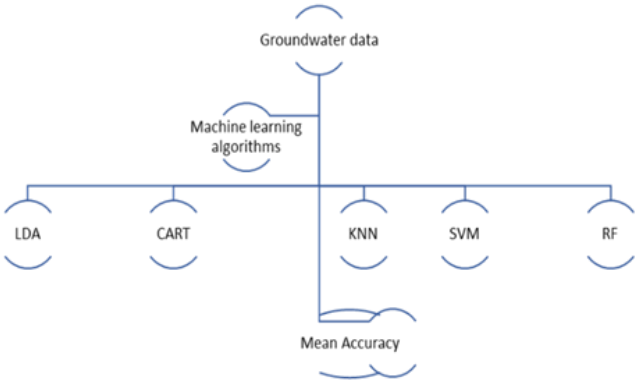


Figure 2. Methodology

2.3.1. Linear Discriminant Analysis (LDA)

In comparison to Principal Components Analysis, we opted to design an algorithm for LDA in the hopes of delivering better classification. The most significant distinction between LDA and PCA is that PCA focuses on feature classification, whereas LDA focuses on data classification. When PCA transforms data sets to a different space, the shape and position of the original data sets change. In contrast, LDA seeks to provide more class separability and create a decision zone between the given classes. This strategy also aids in a better understanding of the feature data distribution [26].

$$S_B = \sum_{i=1}^{c} Ni(mi - m)(mi - m^T)$$

Where 'S_B' is the between-class matrix, 'm' is the overall mean, 'm_i' is the sample mean, and 'N_i' is the size of the respective classes.

Discriminant analysis is a method for distinguishing between two or more groups that meet the study's requirements. Linear combinations of discriminating variables that measure qualities on which the groups are predicted to differ are generated, resulting in a model extrapolating to the rest of the region. Linear discriminant analysis (LDA) has previously been used in New Zealand hydrological studies to distinguish between groups of rivers at base flow conditions [27]. Discriminant analysis has also been used to investigate the distribution of nitrate in groundwater [28]. However, only limited studies have used LDA to discriminate between zones of varying redox status. It is a straightforward machine learning algorithm with a wide range of applications.

2.3.2 Classification and Regression Trees (CART)

The Classification and Regression Trees (CART) algorithm is a decision tree classification technique that uses Gini's impurity index as a splitting condition [29]. CART is a binary tree created by continuously splitting each node into two child nodes. Statistician Leo Breiman coined the term to characterize Decision Tree algorithms to solve classification or regression predictive modeling problems. A Decision Tree is a technique for predictive analysis. The decision tree is the predictive model used here. It is used to go from observations about an item represented by branches to the item's target value, represented by leaves. Decision trees are popular machine learning approaches due to their readability and simplicity. The nodes in the decision tree are divided into sub nodes based on an attribute's threshold value. The CART algorithm uses the Gini Index criterion to find the best homogeneity for the sub nodes. The root node is used as the training set, and the best attribute and the threshold value are used to divide it into two parts. In addition, the subsets are divided using the same rationale. This process is continuously repeated until the tree has the last pure sub-set or the maximum number of leaves conceivable in that growing tree. Tree Pruning is another name for this [30].

$$GI = \sum_{i=0}^{c} Pi(1 - Pi)$$

'GI' is the Gini Index, and 'P' is the estimated output.

The Supervised Learning category includes the k-Nearest Neighbour method used for classification and regression. It is a flexible approach that may also fill in missing values and resample datasets.

2.3.3 K-Nearest Neighbour (KNN)

As the name suggests, the K-Nearest Neighbour algorithm uses k-Nearest Neighbours or Data points to forecast the class or continuous value for a new Datapoint. The nearest neighbours are the data points

with the shortest distance in feature space from our new data point. Moreover, k is the number of data points we consider in our method implementation. As a result, while utilizing the KNN method, the distance metric and the K value are two key factors. The most often used distance measure is Euclidean distance. We can also employ Hamming, Manhattan, and Minkowski distances depending on needs. It considers all of the data points in the training dataset when predicting class/constant value for a new data point. Instead of learning and storing weights, the entire training dataset is saved in memory. As a result, the whole training dataset represents the KNN model [31].

There is overfitting of data/high variance at low K levels. As a result, the test error is significant while the training error is low. Because the nearest neighbour to that point is that point itself, the error is always zero in train data when K=1. As a result, with smaller K values, test error is considerable even while training error is minimal. This is referred to as overfitting. The test error decreases when the value for K is increased. However, after a specific K value, bias/underfitting occurs, and test error increases. So, we may say that the test data error is high at first (due to variation). It drops and stabilizes, and with a higher K value, it rises again (due to bias). When the test error stabilizes and is low, the K value is optimal. We can choose K=8 for our KNN algorithm implementation based on the error curve [32].

It classifies data into a category that is quite similar to the new data [33]. Distance-based approaches are often employed to solve data categorization problems. The k-nearest neighbour classification technique is one of the most extensively used distance-based algorithms (k-NN). This classification compares the distances between the test sample and the training samples to get the final classification result. The conventional k-NN classifier works well with numerical data.

$$R^* \leq R_{knn} \leq R^* (2\text{-}MR^*/M\text{-}1)$$

Where R* is the Bayes error rate, R_{knn} is the k-NN error rate, and M is the number of classes.

2.3.4 Support Vector Machines (SVM)

SVMs (Support Vector Machines) are a novel machine learning technique based on Statistical Learning Theory (Vapnik-Chervonenkis or VC-theory). For the estimate of dependencies and predictive learning from finite data sets, VC-theory has a solid mathematical foundation. SVM is dependent on the Structural Risk Minimisation principle, which aims to reduce both empirical risk and model complexity while maintaining good generalizability.

SVMs (supervised vector machines) are supervised machine-learning algorithms used in classification and regression models. SVMs are more powerful than regression models, but they work best with limited datasets. First, every data point is plotted in an n-dimensional space, with n equalling the number of characteristics. Then a hyperplane is created to divide (classify or sort) the clusters physically. This approach uses the hyperplane to maximize the distance (or margin) between classes while ignoring outliers. When linear separation is not achievable, kernels alter data to make it more separable [34]. SVM (support vector machines) is a supervised learning algorithm that may be used to solve classification and regression problems such as support vector classification (SVC) and support vector regression (SVR) (SVR). However, it is only used for small datasets because processing them takes too long.

$$\text{SVM}= [\tfrac{1}{n} \sum_{i=1}^{n} \max(0,1) - yi(w^T xi - b))] + \lambda \parallel w \parallel^2$$

Where 'w' is the average vector, x_i is a p-dimensional real vector, and 'b' is the boundary.

2.3.5 Random Forests (RF)

RF is a supervised ML algorithm commonly used to solve classification and regression problems. It creates decision trees from various samples, using the majority vote for classification and the average for regression. One of the essential characteristics of the Random Forest Algorithm is that it can handle data sets with both continuous and discrete variables, as in regression and classification. It outperforms the competition when it comes to categorization difficulties [35]. Bagging is a random forest ensemble approach. Bagging takes a random sample of data from the complete set and puts it in a virtual bag. As a result, row sampling is used to substitute the samples (Bootstrap Samples) provided by the Original Data in each model. Row sampling with replacement is known as the "bootstrap" step [36]. Because random forests are built from subsets of data, and the final output is based on average or majority rating, overfitting is avoided. It is, on the whole, slower. Random forest selects data at random, forms a decision tree, and averages the results. It does not rely on any formulas. Bagging, or bootstrap aggregation, is used by the Random Forest classifier to create an ensemble of classification and regression tree (CART)-like classifiers [37].

$$Ni_j = w_j C_j - W_{left(j)} C_{left(j)} - W_{right(j)} C_{right(j)}$$

Where 'Ni_j' is the importance of node j, w_j is the weighted number of samples reaching node j, C_j is the impurity value of the node j, $left_{(j)}$ is the child node from left split on node j and $right_{(j)}$ is the child node from right split on node j.

3. Results

3.1.1 Statistical metrics

The dataset was passed onto machine learning algorithms like LDA, CART, KNN, SVM and RF. The number of resamples employed was 10. The mean accuracy of each classifier was calculated, and the obtained mean accuracies were 77.5% (LDA), 87% (CART), 96% (KNN), 93.5% (SVM) and 96% (RF). RF and KNN models were selected as optimal models with

higher accuracy and represented as dot plots (Table 2 & Figure 3).

$$\text{Average Accuracy} = \frac{1}{|D|}\sum_{1}^{|D|}\frac{\text{xor}(yi,y\ )}{|L|}$$

|D| the number of samples and |L| the number of labels, and y_i is the actual label, y^ the predicted label.

Response y_i and covariates x_i for i=1...n, and Loss function is L. The NIR rate of a model f is the average loss of f over all combinations of yi and xi is given as

$$\text{NIR} = \frac{1}{n^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathcal{L}(\text{yi, f(xj)})$$

Table 2. Mean accuracy of each classifier

Classifier	Mean Accuracy
LDA	77.5%
CART	87%
KNN	96%
SVM	93.5%
RF	96%

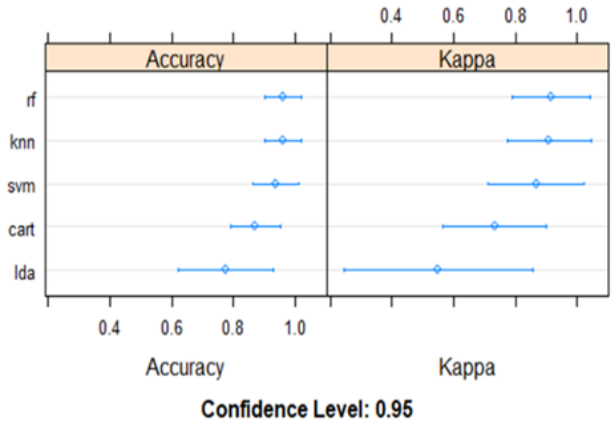


Figure 3. Dot plot

The original dataset was split into training and test data. The type of random forest used in this model is classification with 500 trees. The number of variables attempted at each split is equal to 3. The Out of Bag (OOB) estimate of error rate for training data from all the variables is 12.5%. Confusion matrix using training dataset was represented in a table. The prediction was made using the rf model produced using training data over factor variable (WQ_2LN). The prediction accuracy obtained was 100% with 24 observations ('1') and 16 observations ('2'). This accuracy was 100% obtained at 95% CI within 0.9119 and 1. The No Information Rate (NIR) was 0.6, and the p-value of accuracy is greater than NIR (1.337e-09). The positive class obtained was '1' with a sensitivity and specificity of 1. The positive and negative predictive value was equal to 1. The prevalence and detection rate were 0.6. The detection prevalence is 0.6 with a balanced accuracy of 1.0. The prediction was made using the rf model passed over test data. The accuracy obtained was 100% at 95% CI within 0.7941 and 1. The NIR value is 0.5, and the p-value of accuracy was greater than NIR (1.526e-05). The specificity and sensitivity were 1, with positive and negative prediction values equal one. The positive class obtained was '1'. The variable importance observed was in the order of EC > Cl > NO_3 > Mg > pH > TA > TH > Ca > HCO_3 > Na > SAR > SO_4 > RSC >F. The mean decrease Gini values reflected EC (3.583) as a top player in determining the accuracy. The confusion matrices of training and test data are shown in Tables 3, 4, 5 and 6. Variable importance and mean Gini decrease is shown in Table 7 and 8.

The original dataset was split into training and test data. To avoid confusion, the factor variable (WQ_2LN) was assigned labels 'Yes' and 'No.' The method used in this process was repeated cross-validation with 10-fold and three repeats. The accuracy obtained using training data was 92.5% at k=5 (Table 9). The confusion matrix resulted in an accuracy of 100% at 95% CI with 0.7151 and 1. The NIR value was 0.634, and the p-value of accuracy is more significant than NIR (0.00693). The value of sensitivity and specificity is 1. The positive and negative prediction value obtained was equal to 1. The prevalence value is 0.3636, and the detection rate was 0.3636. The balanced accuracy obtained was 1. The positive class obtained was 'No.' The test data was allowed to run on this model, and a confusion matrix was obtained. The accuracy obtained with test data was 100% at 95% CI with 0.7151 and 1. The NIR value obtained was 0.6364 with a p-value of accuracy greater than NIR (0.00693). The sensitivity, specificity, positive and negative predictive value was 1. The prevalence and detection rate were equal to 0.3636. The balanced accuracy obtained was equal to 1. The positive class is 'No.' For further analysis, the training data was subjected to tuneLength of 20 with 'center' and 'scale' pre-processing (Table 10 & 11).

Table 3. Confusion Matrix (Training Dataset)

	Reference	
Prediction	**1**	**2**
1	24	0
2	0	16

Table 4. RF-Training Dataset statistics

Accuracy	100%
95% CI	(0.9119, 1)
No Information Rate	60%
P-Value [Acc > NIR]	1.337e-09

Table 5. Confusion Matrix (Test Data)

	Reference	
Prediction	1	2
1	8	0
2	0	8

Table 6. RF-Test Dataset statistics

Accuracy	100%
95% CI	(0.7941, 1)
No Information Rate	50%
P-Value [Acc > NIR]	1.526e-05

Table 7. Variable importance

	Overall
HCO_3	1.02559
Ca	1.306406
Cl	2.005265
EC	3.583079
F	0.324502
Mg	1.783165
Na	0.703416
NO3	1.877962
pH	1.538661
RSC	0.616339
SAR	0.662882
SO4	0.657861
TH	1.307576
TA	1.411797

Table 8. Mean decrease Gini

	Mean DecreaseGini
HCO_3	1.0255904
Ca	1.3064055
Cl	2.0052647
EC	3.5830793
F	0.3245019
Mg	1.7831649
Na	0.7034163
NO_3	1.8779617
pH	1.5386611
RSC	0.6163392
SAR	0.6628815
SO4	0.6578606
TH	1.3075757
TA	1.4117973

Table 9. Accuracy of KNN classifier (Training data)

k	Accuracy	Kappa
5	0.925	0.840909
7	0.908333	0.807576
9	0.895	0.771212
11	0.863333	0.706061
13	0.876667	0.742424
15	0.841667	0.659091
17	0.843333	0.660606
19	0.843333	0.660606
21	0.803333	0.578788
23	0.796667	0.563636

Table 10. Confusion Matrix (Test data)

	Reference	
Prediction	No	Yes
No	4	0
Yes	0	7

Table 11. Statistics (KNN)

Accuracy	100%
95% CI	(0.7151, 1)
No Information Rate	63%
P-Value [Acc > NIR]	0.00693

4. Discussion

Machine learning (ML) tools are used in several studies across domains with a high accuracy rate. This study assumes that ML tools can predict the values of the groundwater quality variables with both accuracy and prediction. This work compared five machine learning algorithms under classification mode. Two of the five algorithms provided higher accuracy in predicting the groundwater quality variables. The data was split into training and test data, and their respective accuracies were good. The groundwater surveys are always expensive, and ML tools can predict accurate values for the points that are unknown or yet to be explored. This study can be extended to surface water quality parameters and propagation of the pollutants. The area selected for this study is not conducive to regular groundwater surveys due to topographic inconvenience. It calls for a need to use ML algorithms.

5. Conclusion

As the prediction and modeling are always based on the data availability, it is often buoyant in most areas for several reasons. Artificial intelligence, Machine learning and Geostatistics can help us in filling the gap in hydrological research. Though interpolation serves this immediate purpose, whatever is left in prediction studies can be easily satisfied with ML tools. There is a need to explore many aspects of groundwater in this area, and it is expected that machine learning can be added to the methodologies. The deep learning model backed with the neural networks are used in understanding several aspects of groundwater and the pressure of population on it. Since groundwater is being over exploited for obvious reasons, this study might aid the researchers in developing integrated machine learning and AI models in saving water for present and future generations.

The second and subsequent lines of each bibliography should be indented 0.5 cm inward as shown in this text.

Thesis should be written as Master's Thesis or Doctoral Thesis in the reference list.

Acknowledgement

The author would like to acknowledge Central Groundwater Board, Government of India for making data available.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Aytaç, E. (2020). Unsupervised learning approach in defining the similarity of catchments: Hydrological response unit-based k-means clustering, a demonstration on Western Black Sea Region of Turkey. International Soil and Water Conservation Research, 8(3), 321–331. https://doi.org/10.1016/j.iswcr.2020.05.002
2. Singha, S., Pasupuleti, S., Singha, S. S., Singh, R., & Kumar, S. (2021). Prediction of groundwater quality

using efficient machine learning technique. Chemosphere, 276. https://doi.org/10.1016/j.chemosphere.2021.130265
3. Bilali, A., Taleb, A., & Brouziyne, Y. (2021). Groundwater quality forecasting using machine learning algorithms for irrigation purposes. Agricultural Water Management, 245. https://doi.org/10.1016/j.agwat.2020.106625
4. Yenugu, S. R., Vangala, S., & Badri, S. (2020a). Groundwater quality evaluation using GIS and water quality index in and around inactive mines, Southwestern parts of Cuddapah basin, Andhra Pradesh, South India. HydroResearch, 3, 146–157. https://doi.org/10.1016/j.hydres.2020.11.001
5. Brindha, K., Pavelic, P., Sotoukee, T., Douangsavanh, S., & Elango, L. (2017). Geochemical Characteristics and Groundwater Quality in the Vientiane Plain, Laos. Exposure and Health, 9(2), 89–104. https://doi.org/10.1007/s12403-016-0224-8
6. Reddy, B. M., V.Sunitha, M.Prasad, Reddy, Y. S., & Reddy, M. R. (2019). Evaluation of groundwater suitability for domestic and agricultural utility in semi-arid region of Anantapur, Andhra Pradesh State, South India. Groundwater for Sustainable Development, 9, 100262. https://doi.org/10.1016/j.gsd.2019.100262
7. Datta, P. S., & Tyagi, S. K. (1996). Major Ion Chemistry of Groundwater in Delhi Area: Chemical Weathering Processes and Groundwater Flow Regime. Journal of Geological Society of India (Online Archive from Vol 1 to Vol 78), 47(2), 179–188.
8. Raju, N. J. (2007). Hydrogeochemical parameters for assessment of groundwater quality in the upper Gunjanaeru River basin, Cuddapah District, Andhra Pradesh, South India. Environmental Geology, 52(6), 1067–1074. https://doi.org/10.1007/s00254-006-0546-0
9. Ramakrishna Reddy, M., Janardhana Raju, N., Venkatarami Reddy, Y., & Reddy, T. V. K. (2000). Water resources development and management in the Cuddapah district, India. Environmental Geology, 39(3), 342–352. https://doi.org/10.1007/s002540050013
10. Sreedevi, P. D. (2004a). Groundwater Quality of Pageru River Basin, Cuddapah District, Andhra Pradesh. Journal of Geological Society of India (Online Archive from Vol 1 to Vol 78), 64(5), 619–636.
11. Bedi, S., Samal, A., Ray, C., & Snow, D. (2020). Comparative evaluation of machine learning models for groundwater quality assessment. Environmental Monitoring and Assessment, 192(12), 776. https://doi.org/10.1007/s10661-020-08695-3
12. Mosavi, A., Hosseini, F. S., Choubin, B., Abdolshahnejad, M., Gharechaee, H., Lahijanzadeh, A., & Dineva, A. A. (2020). Susceptibility Prediction of Groundwater Hardness Using Ensemble Machine Learning Models. Water, 12(10), 2770. https://doi.org/10.3390/w12102770
13. Sajedi-Hosseini, F., Malekian, A., Choubin, B., Rahmati, O., Cipullo, S., Coulon, F., & Pradhan, B. (2018). A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. Science of The Total Environment, 644, 954–962. https://doi.org/10.1016/j.scitotenv.2018.07.054
14. Agrawal, P., Sinha, A., Kumar, S., Agarwal, A., Banerjee, A., Villuri, V. G. K., … Pasupuleti, S. (2021). Exploring Artificial Intelligence Techniques for Groundwater Quality Assessment. Water, 13(9), 1172. https://doi.org/10.3390/w13091172
15. Tamiru, H., & Wagari, M. (2021). Comparison of ANN model and GIS tools for delineation of groundwater potential zones, Fincha Catchment, Abay Basin, Ethiopia. Geocarto International, 0(0), 1–19. https://doi.org/10.1080/10106049.2021.1946171
16. Naghibi, S. A., Pourghasemi, H. R., & Abbaspour, K. (2018). A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS. Theoretical and Applied Climatology, 131(3), 967–984. https://doi.org/10.1007/s00704-016-2022-4
17. Golkarian, A., Naghibi, S. A., Kalantar, B., & Pradhan, B. (2018). Groundwater potential mapping using C5.0, random forest, and multivariate adaptive regression spline models in GIS. Environmental Monitoring and Assessment, 190(3), 149. https://doi.org/10.1007/s10661-018-6507-8
18. Acar, E., & Özerdem, M. S. (2020). On a yearly basis prediction of soil water content utilizing sar data: A machine learning and feature selection approach. Turkish Journal of Electrical Engineering & Computer Sciences, 28(4), 2316–2330. Retrieved from https://online-journals.tubitak.gov.tr/publishedManuscriptDetails.htm?id=27563
19. Acar, E., Ozerdem, M. S., & Ustundag, B. B. (2019). Machine Learning based Regression Model for Prediction of Soil Surface Humidity over Moderately Vegetated Fields. 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), 1–4. 8820461 https://doi.org/10.1109/AgroGeoinformatics.2019.
20. Al-Adhaileh, M. H., & Alsaade, F. W. (2021). Modelling and prediction of water quality by using artificial intelligence. Sustain., 13. https://doi.org/10.3390/su13084259
21. https://indiawris.gov.in/wris/#/GWQuality
22. http://cgwb.gov.in/GW-data-access.html
23. Districts, India, 2016—University of Texas Libraries GeoData. (n.d.). Retrieved November 21, 2021, from https://geodata.lib.utexas.edu/catalog/stanford-sh819zz8121
24. Yenugu, S. R., Vangala, S., & Badri, S. (2020b). Monitoring of groundwater quality for drinking purposes using the WQI method and its health implications around inactive mines in Vemula-Vempalli region, Kadapa District, South India. Applied Water Science, 10(8), 202. https://doi.org/10.1007/s13201-020-01284-2
25. Sreedevi, P. D. (2004b). Groundwater quality of Pageru River basin, Cuddapah District, Andhra Pradesh. Journal of Geological Society of India, 64.
26. Castro, C. L., & Braga, A. P. (2013). Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data. IEEE Transactions on Neural Networks and Learning

Systems, 24. https://doi.org/10.1109/TNNLS.2013.2246188
27. Collins, R., & Jerkins, A. (1996). The impact of agriculture land use on stream chemistry in the middle Hills of the Himalayas, Nepal. Journal of Hydrology, 185. https://doi.org/10.1016/0022-1694(95)03008-5
28. Ako, A. A., Eyong, G. E. T., Shimada, J., Koike, K., Hosono, T., Ichiyanagi, K., … Roger, N. N. (2014). Nitrate contamination of groundwater in two areas of the Cameroon Volcanic Line (Banana Plain and Mount Cameroon area). Applied Water Science, 4(2), 99–113. https://doi.org/10.1007/s13201-013-0134-x
29. Cateni, S., Colla, V., & Vannucci, M. (2014). A method for resampling imbalanced datasets in binary classification tasks for real-world problems. Neurocomputing, 135. https://doi.org/10.1016/J.NEUCOM.2013.05.059
30. Ajmera, T. K., & Goyal, M. K. (2012). Development of stage discharge rating curve using model tree and neural networks: An application to Peachtree Creek in Atlanta. *Expert Systems with Applications*, 39. https://doi.org/10.1016/j.eswa.2011.11.101
31. Zhou, Z. H., & Liu, X. Y. (2006). Training cost-sensitive neural networks with methods addressing the class imbalance problem. IEEE Transactions on Knowledge and Data Engineering, 18. https://doi.org/10.1109/TKDE.2006.17
32. Zhang, C., Tang, Y., Xu, X., & Kiely, G. (2011). Towards spatial geochemical modelling: Use of geographically weighted regression for mapping soil organic carbon contents in Ireland. Applied Geochemistry, 26.
33. Cunningham, P., & Delany, S. J. (2021). k-Nearest Neighbour Classifiers—A Tutorial. Conference Papers. https://doi.org/10.1145/3459665
34. Celestino, A. E. M., Cruz, D. A. M., Sánchez, E. M. O., & Reyes, F. G. (n.d.). Groundwater Quality Assessment: An Improved Approach to K-Means Clustering, Principal Component Analysis and Spatial Analysis: A Case Study. Retrieved from https://core.ac.uk/display/156977871
35. Biau, G. (2012). Analysis of a Random Forests Model. Journal of Machine Learning Research, 13(38), 1063–1095. Retrieved from http://jmlr.org/papers/v13/biau12a.html
36. Hastie, T., Tibshirani, R., & Friedman, J. (2009). Random Forests. In T. Hastie, R. Tibshirani, & J. Friedman (Eds.), The Elements of Statistical Learning: Data Mining, Inference, and Prediction (pp. 587–604). New York, NY: Springer. https://doi.org/10.1007/978-0-387-84858-7_15
37. Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. Pattern Recognition Letters, 27(4), 294–300. https://doi.org/10.1016/j.patrec.2005.08.011

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

A critical evaluation of maximum power point tracking techniques for PV systems working under partial shading conditions

Fuad Alhaj Omar [*1], **Nihat Pamuk** [2], **Ahmet Afşin Kulaksız** [3]

[1] *Zonguldak Bülent Ecevit University, Department of Electric and Energy, Türkiye*
[2] *Zonguldak Bülent Ecevit University, Department of Electrical and Electronics Engineering, Türkiye*
[3] *Konya Technical University, Department of Electrical and Electronics Engineering, Türkiye*

Keywords
Photovoltaic system
Maximum power point tracking
Partial shading conditions
Global tracking
Artificial intelligence

Research Article
DOI: 10.31127/tuje.1032674

Received: 05.12.2021
Accepted: 13.03.2022
Published: 01.04.2022

Abstract

Photovoltaic (PV) energy is a promising source of renewable energy which is sturdy and environmentally friendly. PV generation systems, once installed, produce electricity from solar irradiance without emitting greenhouse gases. To maximize the output power of PV systems, the maximum power point tracking system has been employed (MPPT). The MPPT constitutes a fundamental part of PV systems. In recent years, a large number of MPPT techniques have been proposed. This paper is set up to critically review some of the proposed maximum power point tracking (MPPT) techniques to handle the emergence of multiple MPPs in PV panel characteristics due to the partial shading conditions (PSCs). To define the working principle and the pros and cons of the different proposed techniques clearly and sequentially, they are divided into three groups as follows: conventional MPPT techniques, improved MPPT techniques and artificial intelligence- based MPPT techniques to deal with PSCs. The paper also critically summarizes the findings in terms of their performance in capturing the global maximum power point (GMPP) for PV systems operating under PSCs.

1. Introduction

Nowadays, as a result of fossil fuel depletion, the importance of renewable energy has reached an unprecedented height. The photovoltaic (PV) systems are considered one of the most distinctive systems among the resources of renewable energy because they have many merits, for example, availability, low maintenance, environmental friendliness and a longer lifespan [1]. As a result of these advantages, the PV systems are developing rapidly throughout the world, where the PV energy generation shows a significant development compared to other types of renewable energy sources [2].

However, despite all the successive improvements in the PV industry, Solar cells still have some drawbacks such as high manufacturing cost, low efficiency, degradation of the cells and the fact that the initial investment cost of the solar system is high compared to the traditional fossil fuel systems. Moreover, the nonlinear characteristic of the solar cell (shown in Fig.1) depends on several factors such as the irradiance level and ambient temperature and that have limited the global utilization of the PV system [3] . To overcome these drawbacks and increase the efficiency of the PV system, Maximum Power Point Tracking (MPPT) has been considered as the main solution [4]. The MPP is the point on the current-voltage (I-V) curve that indicates the maximum power that a solar panel can produce under certain climatic conditions as shown in Fig. 1 [5] [6]. Maximum power point tracker (MPPT): an electronic device which continuously searches for the MPP of a PV panel and then makes the operating point of the system at the MPP. In another word, the main purpose of MPPT is to oblige the PV system to operate at a point where maximum efficiency level is obtained. Yet, the strong dependence of the PV system upon the atmospheric conditions makes extracting the maximum available power from its nonlinear characteristics more difficult. To handle these issues, many of MPPT techniques have been proposed to make the PV power generation system operate at the optimal point. The proposed techniques

* **Corresponding Author**
*(fuad2081@hotmail.com) ORCID ID 0000-0001-5969-2513
(nihatpamuk@beun.edu.tr) ORCID ID 0000-0001-8980-6913
(aakulaksiz@ktun.edu.tr) ORCID ID 0000-0003-3216-8185

Cite this article
Alhaj Omar, F., Pamuk, N., & Kulaksız, A. A. (2023). A critical evaluation of maximum power point tracking techniques for PV systems working under partial shading conditions. Turkish Journal of Engineering, 7(1), 73-81

vary in several aspects such as convergence speed, complexity, cost, implementation hardware, sensors required and range of effectiveness [7]. These techniques can be divided into conventional MPPT techniques, improved MPPT techniques and artificial intelligence-based MPPT techniques to handle the PSCs. In the following sections, we will provide a reference study on the most important proposed techniques, along with an account of their pros and cons.

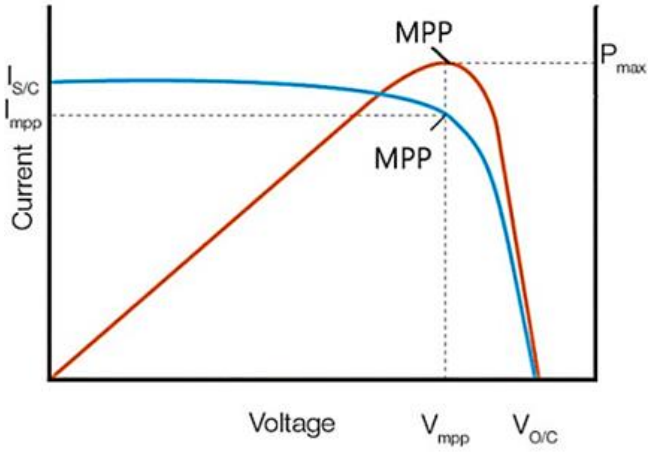


Figure 1. I-V and P-V curve of a PV panel

Where:
V_{MPP}: the voltage at the maximum power point.
I_{MPP}: the current at the maximum power point.
I_{SC}: the short circuit current.
V_{OC}: the open-circuit voltage.

2. Conventional MPPT Techniques

Several conventional MPPT techniques have been proposed and there has been considerable research on them [8] [9]. Among the conventional techniques mentioned in these publications are perturbation and observation (P&O) [10], [11], hill climbing (HC) algorithm [12], fractional open circuit voltage and short circuit current methods [13], switching ripple correlation [14], sliding mode control [15], incremental conductance (IC) [16] [17], constant voltage [18] and some other techniques [19]. Amongst the conventional MPPT techniques, the P&O and HC algorithms are considered the most popular ones due to their ease of implementation [20]. The working principle of both algorithms depends on changing the control parameter by a constant value and exploring whether the MPP has been captured or not. The direction of changing the control parameter is determined based on the increase in the power produced. Besides, the IC algorithm is considered one of the most used conventional algorithms due to its high performance. It is based on computing the differential of the PV power to PV voltage to determine the location of the operating point, where the differential is zero at the MPP [21].

Although these techniques can catch the optimal point of the PV panels under uniform solar irradiance conditions, they fail to capture the GMPP of the PV panels operating under PSCs. Under PSCs, the PV panel characteristics exhibit many local maximum power points (LMPPs) and one global MPP (see Fig.2) due to the use of the bypass diodes to handle the hot spot phenomena [22]. In this situation, to accurately capture the GMPP, and avoid trapping in one of the LMPPs, optimized techniques are required.

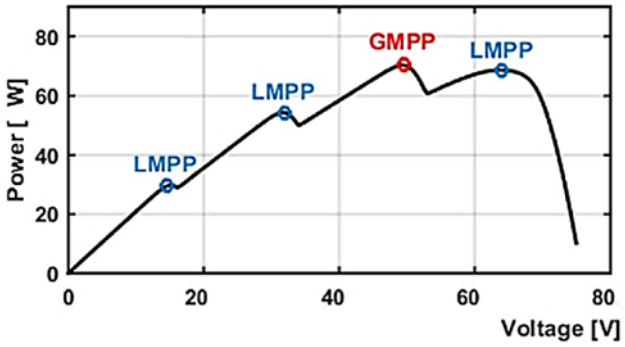


Figure 2. Partially shaded PV panel curve

3. Improved MPPT Techniques to Deal with PSCs

To overcome the drawbacks of the conventional MPPT techniques and handle the PSCs issue, researchers have developed new MPPT techniques, some of which are addressed here. The method introduced in Ref. [23] is an improved IC algorithm. This method is able to capture the GMPP by distinguishing all MPPs in the P-V characteristic. However, a wide range of data, which are the electrical parameters of the PV panels, is required. In Ref. [24], to achieve the GMPPT, an extremum seeking control is proposed. Notwithstanding that this technique has a rapid dynamic response, it requires more information about the electrical parameters of the PV panels. The technique proposed in [25] improved the transient response of the PV system and increased the tracking speed by scanning the entail P-V characteristic of the PV array. However, implementation of this technique requires a high-performance processor, which makes the overall PV system cost- inefficient. The proposed algorithm in [26] is based on the open-circuit voltage of the PV array to estimate the possible GMPP, then the P&O algorithm is employed to track the GMPP. Although this algorithm has a good dynamic tracking performance, it could fail in finding the GMPP under complex PSCs. In Ref. [27] & [28], to decrease the tracking time, a beforehand calculating of the regions of all MPPs is proposed, which, however, requires a great deal of data related to the structure of the PV array which is difficult to obtain or predict. In Ref. [29], a distributed MPPT (DMPPT) algorithm is proposed in order to compensate for the energy loss due to PSCs. This technique requires an increase in the number of converters used; however, this will increase the cost of the whole system. To achieve GMPPT, a two-stage method is outlined in Ref. [30], while Ref. [31] describes a Fibonacci sequence-based technique. However, they were not able to find GMPP in all cases. In Ref. [32], additional sensors are required to implement the algorithm proposed which increase the implementation cost. In Ref. [33], to detect the PSC, a method based on calculating the solar irradiance at two different points is proposed. The aim of using this technique is to avoid unnecessary searching for GMPP. The research in Ref.

[34] introduced innovation technique with the comparison of power tracked in incremented duty cycle during GMPPT process. Although effective results were obtained in this research, the scan and sampling processes throughout the P-V characteristic are still required. In Ref. [35], the authors demonstrated that if the P-V curve has multiple MPPs, the algorithm may be unable to identify the true MPP. In Ref. [36] & [28], to improve the performance of the conventional MPPTs, the authors proposed using the "skipping" mechanism. In this approach, specified voltage intervals will not be scanned because the algorithm has already detected that GMPP is not located in those divisions. In this approach, the tracking speed is improved as a result of reducing the search space. Both maximum power trapezium algorithm (MPT) [36], and voltage window search algorithm (VWS) [28], are based on skipping mechanism. In MPT algorithm, during convergence to the GMPP, the sample of current and voltage are used to update the following voltage references. Although convergence speed is improved, the MPT algorithm possibly does not find GMPP if the chosen voltage step is incorrect.

In Ref. [37], the search-skip-judge approach (SSJ) is proposed. In this method, the section dividing points (SDP) are determined using the short-circuit current values of the shaded modules, which in turn determine the unnecessary voltage periods, so that they can be overshooted. However, according to Ref. [35], if the GMPP is located far at the other end of the P-V curve, the convergence speed of the SSJ will be low. In Ref. [38], to detect the shading, light-detecting resistors (LDRs) were positioned on each PV module. Once the LDRs are shaded, the method becomes active for defining shaded modules; otherwise, the method is ineffective. In Ref. [39], for better detecting shaded modules, authors proposed using PV modules' respective analog sensors to collect their electrical properties. However, the limitations of analog sensing design with respect to the PV modules were not properly discussed. Furthermore, such data can be collected without using analog sensors.

Considering the high cost of analog sensors, authors in Ref. [40] & [41] proposed a technique based on the measured voltage (VPV) and measured current (IPV). The technique starts scanning the P-V curve and its operation initiates from the far-right region, and the scanning process stops if the theoretical power of the next section is less than the practical power of the current section. However, this method suffers if two peaks are quite away from each other. In another technique proposed in ref [42], the power slope of each section is examined to determine the GMPP. Even though this technique presented good performance, it cannot distinguish between regular and PSCs. Arrow Sudoku puzzle pattern which is one of the reconfiguration techniques is studied for PV system under PSCs in ref [43]. To implement such techniques, programmable power switches are required. However, solid-state power switches are expensive while mechanical power switches are slow and may degrade over time. In ref [44], the authors proposed a technique to estimate the location of GMPP by using the voltage in various sub-assemblies of PV modules through determining the irradiance levels in the sub-assemblies. Although the technique introduced good performance under PSCs, implementing this technique requires additional sensors which makes the system more complex and expensive. In Ref. [45] authors introduced a GMPP tracking method by dividing the voltage range of the PV array, and then samples are taken in each section to narrow down the search space and determine the optimal section in which GMPP occurs. However, this technique does not guarantee to find the GMPP under all partial shading patterns; furthermore, many samples are needed at different points to compare with predefined constants to determine any change in irradiance pattern.

4. Artificial Intelligence-Based MPPT Techniques to Deal with PSCs

To date, many artificial intelligence- based MPPT techniques have been proposed to address the PSCs issue. In this subsection, several pieces of research are summarized to provide a comparative study of the most frequently adopted AI-based MPPT techniques.

To achieve global MPPT, in Ref. [46] & [47] particle swarm optimization (PSO) and in Ref. [48] the genetic algorithm (GA) were proposed. Although PSO is simple in implementation and able to find the optimal point, it is computationally intensive and time-consuming which gradually reduces search accuracy. As well GA is complex and requires a long computation time to capture the GMPP. In Ref. [49], a modified PSO method is proposed. In this algorithm, due to the use of many variables with a random number, more iterations are required to find GMPP perfectly, and this makes the convergence speed to GMPP lower than PSO. In the last years, the ant colony algorithm and simulated annealing algorithm have also been proposed to achieve the GMPPT [50]. Modified Java MPPT algorithms were proposed in Ref. [51] & [52], and this algorithm does not require specific parameters. According to the experimental results in Ref. [53] & [50], for increasing the tracking speed of the AI-based MPPT algorithms, suitable initial parameters are required. By assigning the initial values of AI-based MPPT algorithms, the region containing the GMPP can be determined and thus, the tracking speed can be improved [54], [55], [56]. Numerous strategies in identifying the region of the GMPP are documented in Ref. [54], [55] and [56]; however, they required a great deal of information about PV panels characteristics, and this may increase the complexity of implementing these algorithms, given the need to test devices before implementation. In Ref. [57], a fuzzy logic (FL)- based algorithm is proposed. In this algorithm, the slope of the P-V characteristic is the input while the value of the duty cycle is the output. Seven membership functions are used for both the input and the output, and the system has a set of seven rules. In this technique, the whole P-V curve is scanned and the value of the duty cycle which is corresponding to GMPP is stored. In Ref. [58], another FL based MPPT with 7 membership functions and 49 rules is proposed. In this research, error in power and change in error are considered. These FL based MPPT techniques introduced high performance rather than conventional techniques; however, they require more time to converge to the optimal MPP due to large search space.

In addition, MPPT methods are implemented depending on many metaheuristic techniques including cuckoo search [59], grey wolf optimization (GWO) [60], and bat algorithm (BA) [61]. To acquire the benefits of two techniques together, hybrid optimizations were proposed in some papers [60], [62]. In these proposals, two different techniques are used together under different operating conditions. Using these techniques in MPPT applications was evaluated in Ref. [63]. However, the inherent problem in metaheuristic techniques is that they cannot determine the optimal initial value of the duty cycle. As a consequence, they are unable to improve the convergence speed and dynamic variation of the shading patterns. In Ref. [64] BA was proposed as an MPPT of PV systems supplying synchronous reluctance motors. The author tested both BA and PSO techniques under the same conditions and the collected results were compared. BA presented better performance than PSO; however, the problems related to the PSCs, or the dynamic variation of shading patterns were not discussed. In Ref. [65] & [66], firefly algorithm (FA), which is one of the nature-inspired techniques is proposed. It has the advantages of simple computational processes with low power disturbance. Notwithstanding, under PSCs, it may not find the GMPP, and be trapped in one of the LMPPs due to the inherent over-attraction issue [67]. In Ref. [68], to increase the MPPT speed, a modified firefly MPPT algorithm was proposed. The author proposed simplifying the movement rules of fireflies to increase the convergence speed to GMPP. Even though this problem was mitigated by the modified FA, the convergence speed is still slow [69]. Although simplified FA (SFA), which was proposed in Ref. [65], has better convergence speed, it has lower accuracy in capturing GMPP. To deal with this a hybrid approach was proposed in Ref. [70]. The modified FA is used to identify the GMPP, then the P&O algorithm is employed to track it. The results demonstrated that the proposed approach exhibits high performance in tracking the GMPP and improves the convergence speed. In Ref. [71], artificial neural network (ANN) is used to track GMPP under PSCs. However, this technique needs prior training on the PV modules being used. In Ref. [72], General Regression Neural Network (GRNN) was proposed to find the GMPP under PSCs. This technique showed good performance in capturing GMPP; however, implementing such a technique requires a sophisticated embedded system; moreover, the convergence speed to GMPP is slow. In Ref. [73], to achieve GMPPT, a hybrid technique was proposed, which is based on a scanning technique combined with a fuzzy logic technique. In this technique, the entail P-V curve is scanned to capture the GMPP. However, the method on which the scanning depends was not clear. Furthermore, since the observation stating the increasing and decreasing trends when moving far from GMPP cannot be true under all conditions, it was disagreed by [74]. In the technique proposed in Ref. [74], the process of scanning the P-V curve and determining the GMPP was achieved by using a switch in the boost converter for executing the short-circuiting. Regardless of the approach seems excellent, a DSP processor and high-speed ADC are required for sampling the entire data, which increases the implementation cost. Additionally, if the scanning is performed repeatedly under a frequent change in irradiance, then discharging of energy stored in the input capacitor will reduce the efficiency of the system.

A combination between P&O algorithm and the genetic algorithm was proposed in Ref. [54] & [75]. In this combination, the performance of the P&O algorithm was improved in PSCs. Practically, the genetic algorithm identifies the location of GMPP; thereafter, the P&O algorithm tracks the GMPP of the system. Although the performance of the P&O algorithm is improved, it requires much time to track GMPP. Both memory self-organizing incremental neural network (M−SOINN) technique and adaptive neuro-fuzzy inference system (ANFIS) technique were proposed for fault detection [76] & [77]. However, this strategy requires fine-tuning of the parameters, which increases the complexity of programming since more powerful hardware is required to implement such methods. This in turn, increases the implementation cost.

Advanced intelligent algorithms such as Ant Colony Optimization (ACO) [78], Artificial Bee Colony (ABC) [79], Grey Wolf Optimization (GWO) [80] and Simulated Annealing (SA) [81] were adopted to achieve GMPPT. Although these optimizations presented good performance in identifying and capturing GMPP under PSCs, complex processes are needed to fine- tuning of their parameters. Furthermore, to implement such algorithms, sophisticated embedded systems are required, which in turn increases the implementation cost. In addition, these algorithms require a significant number of samples to identify the GMPP which reduces the convergence speed.

It seems clear from the literature that the algorithms using two phases to achieve GMPP are the most effective. Firstly, the section containing the GMPP is identified, then one of the conventional algorithms such as P&O or IC is used to track the optimal point [39] & [70]. Including most of the above details, a detailed analysis of MPPT techniques is given in Table 1.

5. Conclusion

This paper provided a summary of the proposed MPPT techniques to deal with the formation of multiple MPPs in the PV panel characteristic as a result of the PSCs. The information mentioned in the previous studies was used to identify the features of the proposed techniques, and to evaluate their performance in reaching the GMPP. In short, all the proposed techniques were able to deal with the multiple MPPs condition and to determine and reach the GMPP. Depending on this study, researchers can identify the appropriate technique to use in their applications. The weighting in selecting the appropriate technique is related to the features that are important in the application, for example, when a high convergence speed is required, it is preferable to use the skipping mechanism technique or modified firefly algorithm. When simplicity and flexibility of the implementation are the most important, it is more likely to use PSO, while if the accuracy is the most important, it is preferable to use ABC or GWO.

Table 1. Analysis of MPPT techniques

MPPT technique	Features
• Perturbation and Observation (P&O) • Hill climbing (HC) • Fractional open circuit voltage • Fractional short circuit current • Switching ripple correlation • Sliding mode control • Incremental conductance (IC) • Constant voltage	- Conventional MPPT techniques. - Low complexity. - Able to find MPP under uniform radiation conditions. - Not able to find GMPP under PSCs.
• Improved IC algorithm	- Able to capture the GMPP under PSCs. - A wide range of data, which are the electrical parameters of the PV panels, is required.
• Extremum seeking control technique	- Has a rapid dynamic response. - Requires more information about the electrical parameters of the PV panels.
• Distributed MPPT (DMPPT)	- Compensation for power loss due to PSCs. - Requires an increase in the number of converters
• Fibonacci sequence-based technique	- Low complexity. - Not able to find GMPP in all cases.
• Trapezium algorithm	- Improve the convergence speed. - Possibly does not find GMPP if the chosen voltage step is incorrect.
• Voltage window search algorithm	- High convergence speed. - Not able to find GMPP in all cases.
• Search-skip-judge approach (SSJ)	- Able to capture the GMPP under PSCs. - Low convergence speed.
• Particle swarm optimization (PSO)	- Able to capture the GMPP under PSCs. - Low complexity. - Computationally intensive and time-consuming.
• Genetic algorithm (GA)	- Able to capture the GMPP under PSCs. - Complex and requires a long computation time to capture the GMPP
• Modified PSO method	- Able to capture the GMPP perfectly under PSCs. - Convergence speed to GMPP lower than PSO.
• Ant Colony Optimization (ACO) • Artificial Bee Colony (ABC) • Grey Wolf Optimization (GWO) • Simulated Annealing (SA) • Simulated annealing algorithm	- Able to capture the GMPP under PSCs. - High convergence speed. - Suitable initial parameters are required - Complex processes are needed to fine-tune their parameters.
• Fuzzy logic (FL)	- Able to capture the GMPP under PSCs. - High performance. - Low convergence speed.
• Cuckoo search • Grey wolf optimization • Bat algorithm (BA)	- Able to capture the GMPP perfectly under PSCs. - High convergence speed. - High complexity.
• Firefly algorithm (FA)	- Simple computational processes with low power disturbance. - Low complexity. - May not find the GMPP.
• Modified firefly algorithm	- Able to capture the GMPP under PSCs. - Low complexity. - Medium convergence speed.
• Hybrid algorithm based on modified firefly and P&O algorithms	- Able to capture the GMPP perfectly under PSCs. - High performance. - Good convergence speed.
• Artificial neural network (ANN)	- Able to capture the GMPP under PSCs. - Needs prior training on the PV modules being used.
• General Regression Neural Network (GRNN)	- Able to capture the GMPP perfectly under PSCs. - High performance. - Requires a sophisticated embedded system - Low convergence speed

Author contributions

Fuad Alhaj Omar: Coordination, preparing literature review, Analyzing, Formatting, and Writing. **Nihat Pamuk:** Checking references, Auditing, and Preparing the structure. **Ahmet Afşin Kulaksiz:** Supervising and Writing-Reviewing.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Burrett, R., Clini, C., Dixon, R., Eckhart, M., El-Ashry, M., Gupta, D., ... & Ballesteros, A. R. (2009). Renewable Energy Policy Network for the 21st Century.
2. Jäger-Waldau, A. (2019). PV status report 2019. Publications Office of the European Union: Luxembourg.
3. Cheng, P. C., Peng, B. R., Liu, Y. H., Cheng, Y. S., & Huang, J. W. (2015). Optimization of a fuzzy-logic-control-based MPPT algorithm using the particle swarm optimization technique. Energies, 8(6), 5338-5360.
4. Alhajomar, F., Gokkus, G., & Kulaksiz, A. A. (2019). Rapid control prototyping based on 32-Bit ARM Cortex-M3 microcontroller for photovoltaic MPPT algorithms. International Journal of Renewable Energy Research-IJRER.
5. Mule, S., Hardas, R., & Kulkarni, N. R. (2016, March). P&O, IncCon and Fuzzy Logic implemented MPPT scheme for PV systems using PIC18F452. In 2016 international conference on wireless communications, signal processing and networking (WiSPNET) (pp. 1320-1325). IEEE.
6. Anwer, A. M. O., Omar, F. A., & Kulaksiz, A. A. (2020). Design of a fuzzy logic-based MPPT controller for a PV system employing sensorless control of MRAS-based PMSM. International Journal of Control, Automation and Systems, 18(11), 2788-2797.
7. Mohanty, P., Bhuvaneswari, G., Balasubramanian, R., & Dhaliwal, N. K. (2014). MATLAB based modeling to study the performance of different MPPT techniques used for solar PV system under various operating conditions. Renewable and Sustainable Energy Reviews, 38, 581-593.
8. Rezk, H., & Eltamaly, A. M. (2015). A comprehensive comparison of different MPPT techniques for photovoltaic systems. Solar energy, 112, 1-11.
9. Ram, J. P., Babu, T. S., & Rajasekar, N. (2017). A comprehensive review on solar PV maximum power point tracking techniques. Renewable and Sustainable Energy Reviews, 67, 826-847.
10. Kwan, T. H., & Wu, X. (2017). High performance P&O based lock-on mechanism MPPT algorithm with smooth tracking. Solar Energy, 155, 816-828.
11. Omar, F. A., Gökkuş, G., & Kulaksız, A. A. (2019). Şebekeden bağımsız FV sistemde maksimum güç noktası takip algoritmalarının değişken hava şartları altında karşılaştırmalı analizi. Konya Mühendislik Bilimleri Dergisi, 7(3), 585-594.
12. Elgendy, M. A. (2016, March). Comparative investigation on hill climbing MPPT algorithms at high perturbation rates. In 2016 7th International Renewable Energy Congress (IREC) (pp. 1-6). IEEE.
13. Shebani, M. M., Iqbal, T., & Quaicoe, J. E. (2016, October). Comparing bisection numerical algorithm with fractional short circuit current and open circuit voltage methods for MPPT photovoltaic systems. In 2016 IEEE Electrical Power and Energy Conference (EPEC) (pp. 1-5). IEEE.
14. Paz, F., & Ordonez, M. (2016). High-performance solar MPPT using switching ripple identification based on a lock-in amplifier. IEEE Transactions on Industrial Electronics, 63(6), 3595-3604.
15. Mojallizadeh, M. R., Badamchizadeh, M., Khanmohammadi, S., & Sabahi, M. (2016). Designing a new robust sliding mode controller for maximum power point tracking of photovoltaic cells. Solar Energy, 132, 538-546.
16. Kamarzaman, N. A., & Tan, C. W. (2014). A comprehensive review of maximum power point tracking algorithms for photovoltaic systems. Renewable and Sustainable Energy Reviews, 37, 585-598.
17. Anwer, A. M. O., Omar, F. A., Bakir, H., & Kulaksiz, A. A. (2020). Sensorless Control of a PMSM Drive Using EKF for Wide Speed Range Supplied by MPPT Based Solar PV System. Elektronika ir Elektrotechnika, 26(1), 32-39.
18. Obukhov, S. G., Plotnikov, I. A., & Sheryazov, S. K. (2016, May). Methods of effective use of solar power system. In 2016 2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM) (pp. 1-6). IEEE.
19. Husain, M. A., Tariq, A., Hameed, S., Arif, M. S. B., & Jain, A. (2017). Comparative assessment of maximum power point tracking procedures for photovoltaic systems. Green Energy & Environment, 2(1), 5-17.
20. Liu, F., Kang, Y., Zhang, Y., & Duan, S. (2008, June). Comparison of P&O and hill climbing MPPT methods for grid-connected PV converter. In 2008 3rd IEEE Conference on Industrial Electronics and Applications (pp. 804-807). IEEE.
21. Shah, K. B., & Joshi, L. P. (2013, November). Comparative analysis of incremental conductance base MPPT for multi-string photovoltaic system. In 2013 Nirma University International Conference on Engineering (NUiCONE) (pp. 1-6). IEEE.
22. Seyedmahmoudian, M., Horan, B., Soon, T. K., Rahmani, R., Oo, A. M. T., Mekhilef, S., & Stojcevski, A. (2016). State of the art artificial intelligence-based MPPT techniques for mitigating partial shading effects on PV systems–A review. Renewable and Sustainable Energy Reviews, 64, 435-455.
23. Ji, Y. H., Jung, D. Y., Kim, J. G., Kim, J. H., Lee, T. W., & Won, C. Y. (2010). A real maximum power point tracking method for mismatching compensation in PV array under partially shaded conditions. IEEE Transactions on power electronics, 26(4), 1001-1009.
24. Lei, P., Li, Y., & Seem, J. E. (2011). Sequential ESC-based global MPPT control for photovoltaic array

with variable shading. IEEE Transactions on Sustainable Energy, 2(3), 348-358.
25. Ghasemi, M. A., Foroushani, H. M., & Parniani, M. (2015). Partial shading detection and smooth maximum power point tracking of PV arrays under PSC. IEEE Transactions on Power Electronics, 31(9), 6281-6292.
26. Patel, H., & Agarwal, V. (2008). Maximum power point tracking scheme for PV systems operating under partially shaded conditions. IEEE transactions on industrial electronics, 55(4), 1689-1698.
27. Nguyen, T. L., & Low, K. S. (2010). A global maximum power point tracking scheme employing DIRECT search algorithm for photovoltaic systems. IEEE transactions on Industrial Electronics, 57(10), 3456-3467.
28. Boztepe, M., Guinjoan, F., Velasco-Quesada, G., Silvestre, S., Chouder, A., & Karatepe, E. (2013). Global MPPT scheme for photovoltaic string inverters based on restricted voltage window search algorithm. IEEE transactions on Industrial Electronics, 61(7), 3302-3312.
29. Femia, N., Lisi, G., Petrone, G., Spagnuolo, G., & Vitelli, M. (2008). Distributed maximum power point tracking of photovoltaic arrays: Novel approach and system analysis. IEEE Transactions on Industrial Electronics, 55(7), 2610-2621.
30. Kobayashi, K., Takano, I., & Sawada, Y. (2006). A study of a two-stage maximum power point tracking control of a photovoltaic system under partially shaded insolation conditions. Solar energy materials and solar cells, 90(18-19), 2975-2988.
31. Miyatake, M., Inada, T., Hiratsuka, I., Zhao, H., Otsuka, H., & Nakano, M. (2004, August). Control characteristics of a fibonacci-search-based maximum power point tracker when a photovoltaic array is partially shaded. In The 4th International Power Electronics and Motion Control Conference, 2004. IPEMC 2004. (Vol. 2, pp. 816-821). IEEE.
32. Kazmi, S. M. R., Goto, H., Ichinokura, O., & Guo, H. J. (2009, September). An improved and very efficient MPPT controller for PV systems subjected to rapidly varying atmospheric conditions and partial shading. In 2009 Australasian Universities Power Engineering Conference (pp. 1-6). IEEE.
33. Ahmed, J., & Salam, Z. (2017). An accurate method for MPPT to detect the partial shading occurrence in a PV system. IEEE transactions on industrial informatics, 13(5), 2151-2161.
34. Koutroulis, E., & Blaabjerg, F. (2012). A new technique for tracking the global maximum power point of PV arrays operating under partial-shading conditions. IEEE journal of photovoltaics, 2(2), 184-190.
35. Kermadi, M., Salam, Z., Ahmed, J., & Berkouk, E. M. (2018). An effective hybrid maximum power point tracker of photovoltaic arrays for complex partial shading conditions. IEEE Transactions on Industrial Electronics, 66(9), 6990-7000.
36. Furtado, A. M., Bradaschia, F., Cavalcanti, M. C., & Limongi, L. R. (2017). A reduced voltage range global maximum power point tracking algorithm for photovoltaic systems under partial shading conditions. IEEE Transactions on Industrial Electronics, 65(4), 3252-3262.
37. Wang, Y., Li, Y., & Ruan, X. (2015). High-accuracy and fast-speed MPPT methods for PV string under partially shaded conditions. IEEE Transactions on Industrial Electronics, 63(1), 235-245.
38. Bayod-Rújula, Á. A., & Cebollero-Abián, J. A. (2014). A novel MPPT method for PV systems with irradiance measurement. solar energy, 109, 95-104.
39. Chen, K., Tian, S., Cheng, Y., & Bai, L. (2014). An improved MPPT controller for photovoltaic system under partial shading condition. IEEE transactions on sustainable energy, 5(3), 978-985.
40. Manickam, C., Raman, G. P., Raman, G. R., Ganesan, S. I., & Chilakapati, N. (2016). Efficient global maximum power point tracking technique for a partially shaded photovoltaic string. IET Power Electronics, 9(14), 2637-2644.
41. Manickam, C., Raman, G. R., Raman, G. P., Ganesan, S. I., & Nagamani, C. (2016). A hybrid algorithm for tracking of GMPP based on P&O and PSO with reduced power oscillation in string inverters. IEEE Transactions on Industrial Electronics, 63(10), 6097-6106.
42. Balasankar, R., Arasu, G. T., & Raj, J. C. M. (2017). A global MPPT technique invoking partitioned estimation and strategic deployment of P&O to tackle partial shading conditions. Solar energy, 143, 73-85.
43. Tatabhatla, V. M. R., Agarwal, A., & Kanumuri, T. (2019). Performance enhancement by shade dispersion of Solar Photo-Voltaic array under continuous dynamic partial shading conditions. Journal of cleaner production, 213, 462-479.
44. Kouchaki, A., Iman-Eini, H., & Asaei, B. (2013). A new maximum power point tracking strategy for PV arrays under uniform and non-uniform insolation conditions. Solar Energy, 91, 221-232.
45. Liu, Y. H., Chen, J. H., & Huang, J. W. (2014). Global maximum power point tracking algorithm for PV systems operating under partially shaded conditions using the segmentation search method. Solar Energy, 103, 350-363.
46. Seyedmahmoudian, M., Rahmani, R., Mekhilef, S., Oo, A. M. T., Stojcevski, A., Soon, T. K., & Ghandhari, A. S. (2015). Simulation and hardware implementation of new maximum power point tracking technique for partially shaded PV system using hybrid DEPSO method. IEEE transactions on sustainable energy, 6(3), 850-862.
47. Liu, Y. H., Huang, S. C., Huang, J. W., & Liang, W. C. (2012). A particle swarm optimization-based maximum power point tracking algorithm for PV systems operating under partially shaded conditions. IEEE transactions on energy conversion, 27(4), 1027-1035.
48. Daraban, S., Petreus, D., & Morel, C. (2014). A novel MPPT (maximum power point tracking) algorithm based on a modified genetic algorithm specialized on tracking the global maximum power point in photovoltaic systems affected by partial shading. Energy, 74, 374-388.
49. Efendi, M. Z., Murdianto, F. D., & Setiawan, R. E. (2017, September). Modeling and simulation of MPPT sepie

converter using modified PSO to overcome partial shading impact on DC microgrid system. In 2017 International Electronics Symposium on Engineering Technology and Applications (IES-ETA) (pp. 27-32). IEEE.
50. Lyden, S., & Haque, M. E. (2015). A simulated annealing global maximum power point tracking approach for PV modules under partial shading conditions. IEEE Transactions on Power Electronics, 31(6), 4171-4181.
51. Kumar, N., Hussain, I., Singh, B., & Panigrahi, B. K. (2017). Rapid MPPT for uniformly and partial shaded PV system by using JayaDE algorithm in highly fluctuating atmospheric conditions. IEEE Transactions on Industrial Informatics, 13(5), 2406-2416.
52. Huang, C., Wang, L., Yeung, R. S. C., Zhang, Z., Chung, H. S. H., & Bensoussan, A. (2017). A prediction model-guided Jaya algorithm for the PV system maximum power point tracking. IEEE Transactions on sustainable energy, 9(1), 45-55.
53. Sundareswaran, K., Vigneshkumar, V., Sankar, P., Simon, S. P., Nayak, P. S. R., & Palani, S. (2015). Development of an improved P&O algorithm assisted through a colony of foraging ants for MPPT in PV system. IEEE transactions on industrial informatics, 12(1), 187-200.
54. Lian, K. L., Jhang, J. H., & Tian, I. S. (2014). A maximum power point tracking method based on perturb-and-observe combined with particle swarm optimization. IEEE journal of photovoltaics, 4(2), 626-633.
55. Ishaque, K., & Salam, Z. (2012). A deterministic particle swarm optimization maximum power point tracker for photovoltaic system under partial shading condition. IEEE transactions on industrial electronics, 60(8), 3195-3206.
56. Kollimalla, S. K., & Mishra, M. K. (2014). A novel adaptive P&O MPPT algorithm considering sudden changes in the irradiance. IEEE Transactions on Energy conversion, 29(3), 602-610.
57. Shah, N., & Chudamani, R. (2012, December). A novel algorithm for global peak power point tracking in partially shaded grid-connected PV system. In 2012 IEEE International Conference on Power and Energy (PECon) (pp. 558-563). IEEE.
58. Mahamudul, H., Saad, M., & Ibrahim Henk, M. (2013). Photovoltaic system modeling with fuzzy logic based maximum power point tracking algorithm. International Journal of Photoenergy, 2013.
59. Ahmed, J., & Salam, Z. (2014). A Maximum Power Point Tracking (MPPT) for PV system using Cuckoo Search with partial shading capability. Applied energy, 119, 118-130.
60. Eltamaly, A. M., & Farh, H. M. (2019). Dynamic global maximum power point tracking of the PV systems under variant partial shading using hybrid GWO-FLC. Solar Energy, 177, 306-316.
61. Kaced, K., Larbes, C., Ramzan, N., Bounabi, M., & elabadine Dahmane, Z. (2017). Bat algorithm based maximum power point tracking for photovoltaic system under partial shading conditions. Solar Energy, 158, 490-503.
62. Priyadarshi, N., Padmanaban, S., Maroti, P. K., & Sharma, A. (2018). An extensive practical investigation of FPSO-based MPPT for grid integrated PV system under variable operating conditions with anti-islanding protection. IEEE Systems Journal, 13(2), 1861-1871.
63. Eltamaly, A. M., Farh, H. M., & Al-Saud, M. S. (2019). Grade point average assessment for metaheuristic GMPP techniques of partial shaded PV systems. IET Renewable Power Generation, 13(8), 1215-1231.
64. Oshaba, A. S., Ali, E. S., & Abd Elazim, S. M. (2015). MPPT control design of PV system supplied SRM using BAT search algorithm. Sustainable Energy, Grids and Networks, 2, 51-60.
65. Li, G., Jin, Y., Akram, M. W., Chen, X., & Ji, J. (2018). Application of bio-inspired algorithms in maximum power point tracking for PV systems under partial shading conditions–A review. Renewable and Sustainable Energy Reviews, 81, 840-873.
66. Sundareswaran, K., Peddapati, S., & Palani, S. (2014). MPPT of PV systems under partial shaded conditions through a colony of flashing fireflies. IEEE transactions on energy conversion, 29(2), 463-472.
67. Huang, Y. P., Ye, C. E., & Chen, X. (2018). A modified firefly algorithm with rapid response maximum power point tracking for photovoltaic systems under partial shading conditions. Energies, 11(9), 2284.
68. Teshome, D. F., Lee, C. H., Lin, Y. W., & Lian, K. L. (2016). A modified firefly algorithm for photovoltaic maximum power point tracking control under partial shading. IEEE Journal of Emerging and Selected Topics in Power Electronics, 5(2), 661-671.
69. Wang, H., Wang, W., Zhou, X., Sun, H., Zhao, J., Yu, X., & Cui, Z. (2017). Firefly algorithm with neighborhood attraction. Information Sciences, 382, 374-387.
70. Alhaj Omar, F., & Kulaksiz, A. A. (2021). Experimental evaluation of a hybrid global maximum power tracking algorithm based on modified firefly and perturbation and observation algorithms. Neural Computing and Applications, 33(24), 17185-17208.
71. Gupta, A., Chauhan, Y. K., & Pachauri, R. K. (2016). A comparative investigation of maximum power point tracking methods for solar PV system. Solar energy, 136, 236-253.
72. Chuang, S. J., Hong, C. M., & Chen, C. H. (2016). Design of intelligent control for stabilization of microgrid system. International Journal of Electrical Power & Energy Systems, 82, 569-578.
73. Boukenoui, R., Salhi, H., Bradai, R., & Mellit, A. (2016). A new intelligent MPPT method for stand-alone photovoltaic systems operating under fast transient variations of shading patterns. Solar Energy, 124, 124-142.
74. Kotti, R., & Shireen, W. (2015). Efficient MPPT control for PV systems adaptive to fast changing irradiation and partial shading conditions. Solar Energy, 114, 397-407.
75. Sundareswaran, K., Sankar, P., Nayak, P. S. R., Simon, S. P., & Palani, S. (2014). Enhanced energy output from a PV system under partial shaded conditions through artificial bee colony. IEEE transactions on sustainable energy, 6(1), 198-209.

76. Kow, K. W., Wong, Y. W., Rajkumar, R., & Isa, D. (2018). An intelligent real-time power management system with active learning prediction engine for PV grid-tied systems. Journal of cleaner production, 205, 252-265.
77. Kaid, I. E., Hafaifa, A., Guemana, M., Hadroug, N., Kouzou, A., & Mazouz, L. (2018). Photovoltaic system failure diagnosis based on adaptive neuro fuzzy inference approach: South Algeria solar power plant. Journal of Cleaner Production, 204, 169-182.
78. Jiang, L. L., Maskell, D. L., & Patra, J. C. (2013). A novel ant colony optimization-based maximum power point tracking for photovoltaic systems under partially shaded conditions. Energy and Buildings, 58, 227-236.
79. Soufyane, B. A., Chouder, A., Kara, K., & Silvestre, S. (2015). Artificial bee colony-based algorithm for maximum power point tracking (MPPT) for PV systems operating under partial shaded conditions. Applied Soft Computing, 32, 38-48.
80. Mohanty, S., Subudhi, B., & Ray, P. K. (2015). A new MPPT design using grey wolf optimization technique for photovoltaic system under partial shading conditions. IEEE Transactions on Sustainable Energy, 7(1), 181-188.
81. Lyden, S., & Haque, M. E. (2015). A simulated annealing global maximum power point tracking approach for PV modules under partial shading conditions. IEEE Transactions on Power Electronics, 31(6), 4171-4181.