IJCESEN

ISSN : 2149-9144

International Journal of Computational and Experimental Science and ENgineering

Volume: 9 - Issue: 1 - 2023

ijcesen@gmail.com

dergipark.org.tr/en/pub/ijcesen

Journal Info	
Web	dergipark.org.tr/en/pub/ijcesen
E-mail	ijcesen@gmail.com
ISSN	2149-9144
Frequency	March-July-November
Founded	2015
Journal Abbreviation	IJCESEN
Language	English-Turkish

Indexing/Abstracting Databases

TOGETHER WE REACH THE GOAL

Table of Contents

International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 9-No.1 (2023) pp. 1-5
http://dergipark.org.tr/en/pub/ijcesen

ISSN: 2149-9144

Research Article

A Novel Shape Descriptor for Object Recognition

Elif Ebru ÇAKI[1]*, Celal Onur GÖKÇE[2]

[1]Afyon Kocatepe Üniversity, Mechatronics Engineering Department, 03204, Afyonkarahisar-Turkiye
* **Corresponding Author** : **Email:** elif_ebru1999@hotmail.com - **ORCID:** 0000-0002-2225-5675

[2]Afyon Kocatepe Üniversity, Software Engineering Department, 03204, Afyonkarahisar-Turkiye
Email: cogokce@aku.edu.tr - **ORCID:** 0000-0003-3120-7808

Article Info:

DOI: 10.22399/ijcesen.1202300
Received : 20 November 2022
Accepted : 17 January 2023

Keywords:

Shape Descriptor
Object Recognition
MNIST

Abstract:

In this study a novel shape descriptor for object recognition is proposed. As a preprocessing stage, Canny edge detection is applied to input images. Output of Canny edge detector, namely edge image, is sampled and various number of points are selected. Chosen points are input to the new shape descriptor. Proposed shape descriptor is composed of deviations from average range and average angle. Shape descriptor is used as a feature extractor output of which is fed to linear classifier. Linear classifier is trained using pseudo-inverse and gradient descent techniques. Full MNIST dataset is used to test the system and results are reported.

1. Introduction

Object recognition is an important subfield of image processing. Several algorithms and methods are developed in this field. One of the most successful fast recognizers is that of Viola Jones [1].
The aim of this study is fast and accurate recognition of objects. Recognition of object means detecting the class of object [2]. In order to classify the image different features can be utilized. Traffic signs, medical images, handwritten character recognition, face recognition and fingerprint recognition are some variants of object recognition. Phases of object recognition process is shown in Figure 1. Image is converted to matrix data structure. In order to be processed by classifier a preprocessing is applied. After feature extraction classification is done.

2. Material and Methods

2.1 MNIST Dataset

One of the mostly used datasets in literature is MNIST (Modified National Institue of Standarts) dataset. It is composed of 60.000 training samples and 10.000 test samples. Each sample is a picture of handwritten digit between 0 and 9. Each digit picture has 28x28 dimension, with 784 pixels of gray levels. Some examples from the MNIST dataset are given in Figure 2 [3].

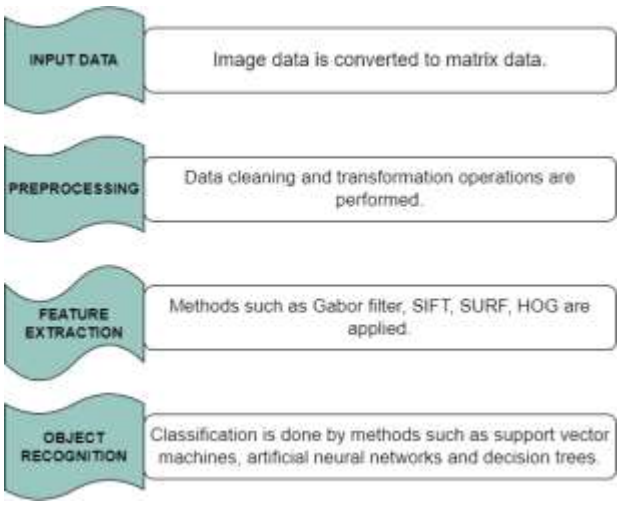


Figure 1. *Object recognition and classification phases*

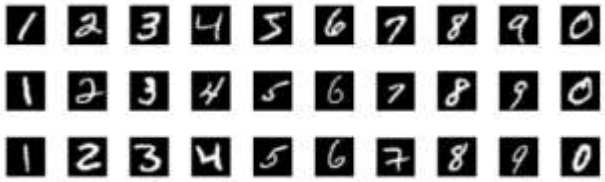


Figure 2. *MNIST dataset examples [3].*

2.2 Proposed Algorithm

Output of Canny edge detector, namely edge image, is sampled using a sub-optimal algorithm [4]. In sub-optimal sampling algorithm, all points are sorted with respect to x coordinates first and y coordinates after. Nearly equally spaced points are sampled within acceptable units of computation time. After sub-optimal algorithm, optimal algorithm is applied to representative points. In optimal algorithm, initially four edge points are selected as east, west, north and south. After these four points, subsequent points are selected optimally that are furthest from every point. Euclidean distance is used between two points $P = (x_1, x_2, \dots, x_n)$ ve $Q = (y_1, y_2, \dots, y_n)$ as shown in Equation 1 [5]. Optimal algorithm is illustrated in Figure 3.

$$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (1)$$

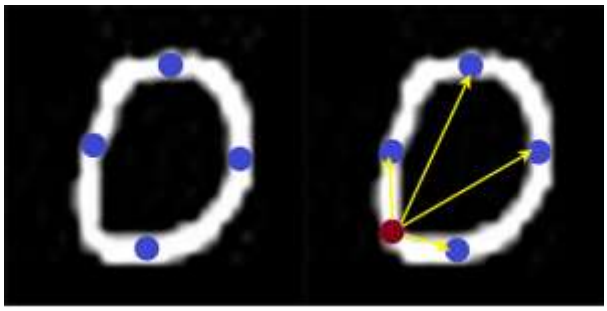


Figure 3. *Optimal algorithm illustrated*

After finding representative subset of edge points, center of mass is found. Distance of each point from center of mass is calculated using Equation 2. Angle of each point to center of mass is calculated using Equation 3. Average distance to center of mass and average angle from center of mass is found and represented as r_m and Θ_m, respectively. For the final descriptor, deviation from average distance and average angle is calculated and normalized as shown in Equation 4 and Equation 5. Distance and angle of representative points are illustrated in Figure 4.

$$r = \sqrt{(x - x_m)^2 + (y - y_m)^2} \qquad (2)$$

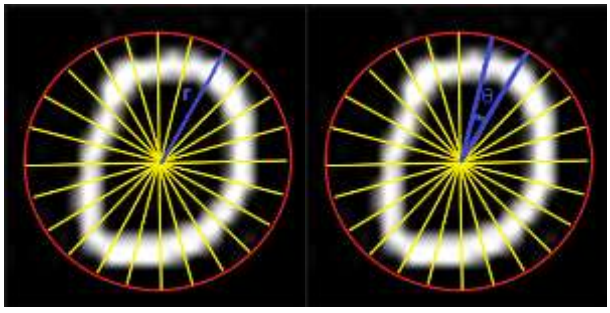


Figure 4. *Distance and angle of representative points*

$$\theta = arctan(\frac{y - y_m}{x - x_m}) \qquad (3)$$

$$\%\Delta r = \frac{r - r_m}{r_m} x100 \qquad (4)$$

$$\%\Delta\theta = \frac{\theta - \theta_m}{\theta_m} x100 \qquad (5)$$

Since the performance of simple algorithm is low, a histogram based method is added. Histogram based data is entered as input to neural network classifier. As future work, optimal thresholds for histograms will be investigated.

2.3 Histogram Based Method

In image processing, a histogram is a bar graph representation of the color distribution of an image. In the graph, the x-axis gives the gray pixel values and the y-axis gives the number of pixels in that value. There is a Cumulative histogram by obtaining the histogram of the image. The Cumulative histogram contains the values obtained from the sum of each value of the histogram with the previous value and itself. It is normalized by dividing the obtained value by the total number of pixels. Thus, the image is improved by scattering the distribution in pixel values on average.

Let the variable r be the pixel value of the image, that is, r is in the range of $[0,\ L-1]$. $r = 0$ is black and $r = L - 1$ is white. Since our images are 8 bits in our study, the maximum pixel value is 255. In this case, $T(r)$ gives the total number of transform function and s pixel values.

$$s = T(r) \quad 0 \le r \le L - 1 \qquad (6)$$

The original data taken from the dataset we used between Figure 5 and Figure 14, the data with the specified 20 points, the angle and distance data, and the histograms of the angle-distance data are given.

Table 1. *Results for various number of neurons*

Number of neurons	Test accuracy	F1 score	Total Loss
128	67.39	0.674	0.968
256	67.9	0.679	0.966
512	69.41	0.694	0.934
1024	69.91	0.699	0.942
2048	68.61	0.686	0.988

Table 2. *Results for various number of epochs*

Number of Epochs	Test accuracy	F1 score	Total Loss
25	69.91	0.699	0.942
50	68.98	0.69	1.08
75	67.75	0.678	1.318
100	66.98	0.67	1.488

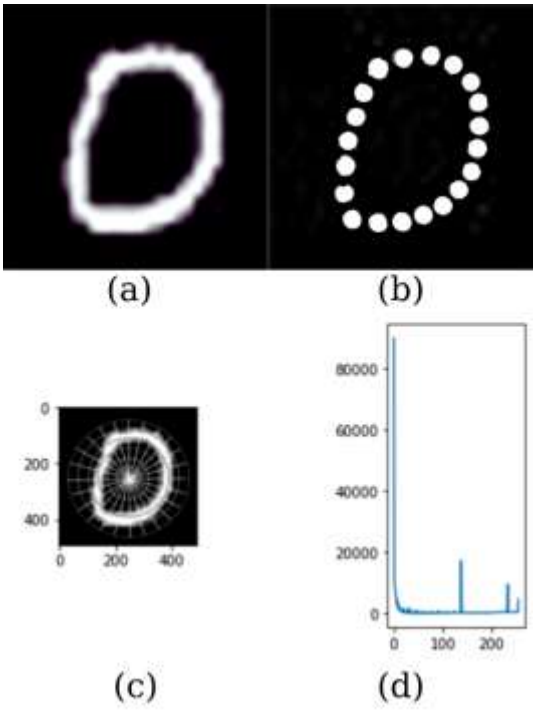


Figure 5. *MNIST dataset- number 0, (a) original data, (b) the data in which it determines the 20 point, (c) data determined by angle and distance, (d) histogram of data*

c

MNIST veri seti 0 sayısı

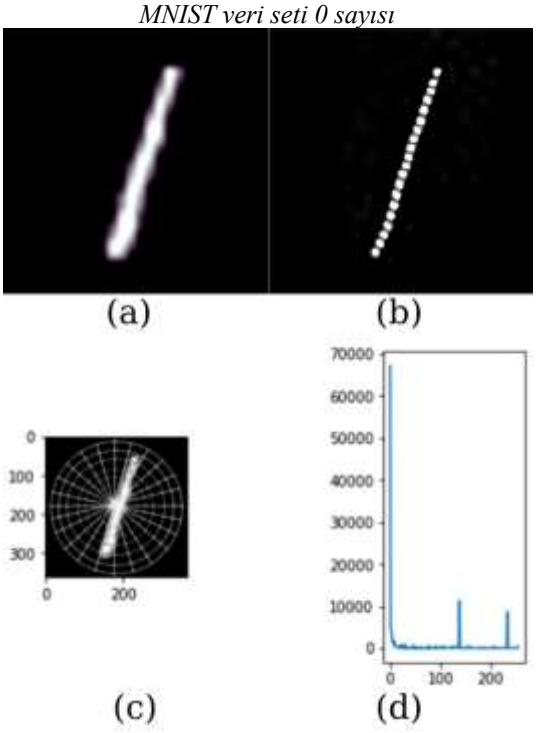
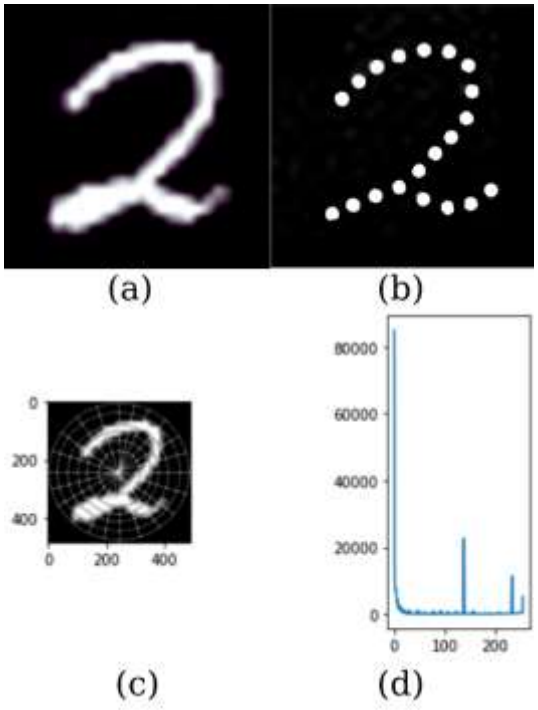


Figure 6. *MNIST dataset- number 1, (a) original data, (b) the data in which it determines the 20 point, (c) data determined by angle and distance, (d) histogram of data*



c

Figure 7. *MNIST dataset- number 2, (a) original data, (b) the data in which it determines the 20 point, (c) data determined by angle and distance, (d) histogram of data*

c

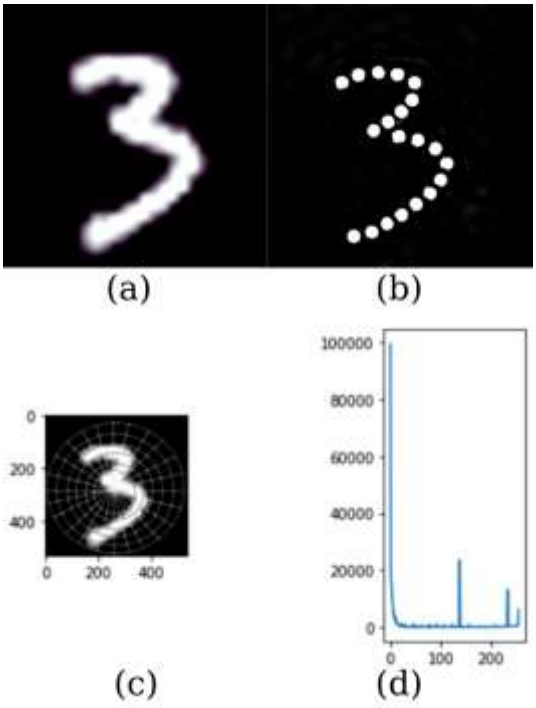


Figure 8. *MNIST dataset- number 3, (a) original data, (b) the data in which it determines the 20 point, (c) data determined by angle and distance, (d) histogram of data*

c

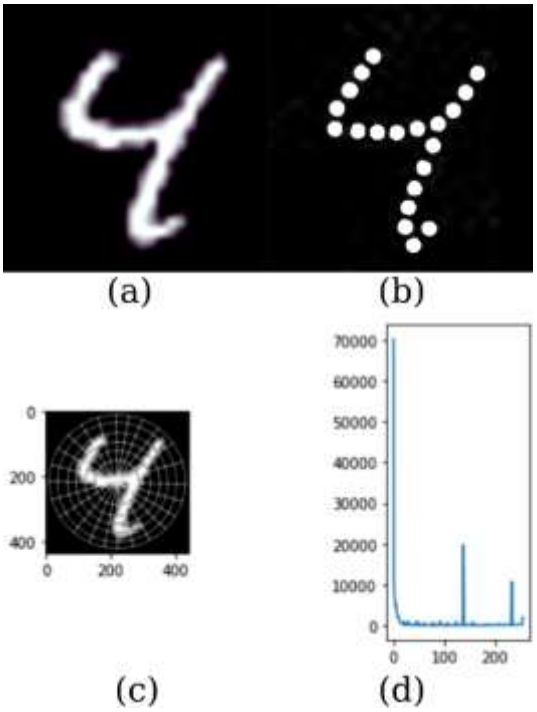


Figure 9. *MNIST dataset- number 4, (a) original data, (b) the data in which it determines the 20 point, (c) data determined by angle and distance, (d) histogram of data c*

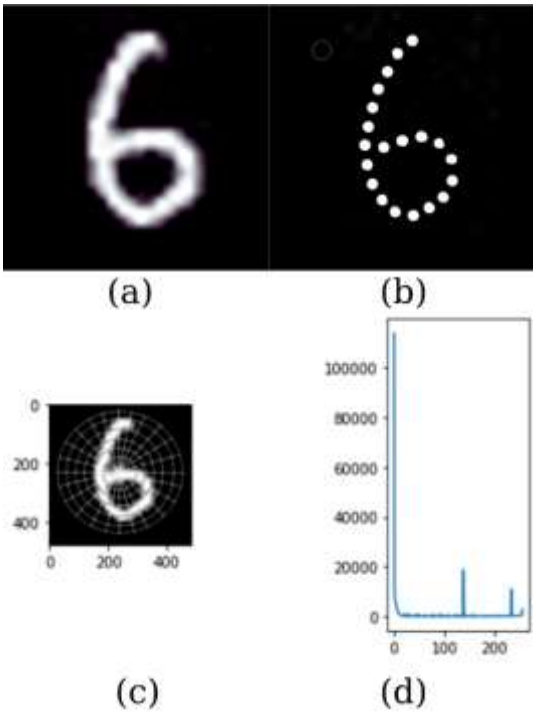


Figure 11. *MNIST dataset- number 6, (a) original data, (b) the data in which it determines the 20 point, (c) data determined by angle and distance, (d) histogram of data c*

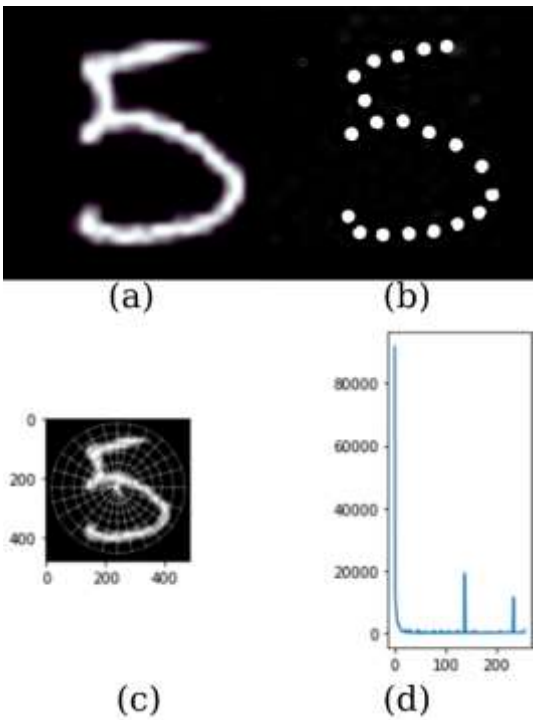


Figure 10. *MNIST dataset- number 5, (a) original data, (b) the data in which it determines the 20 point, (c) data determined by angle and distance, (d) histogram of data c*

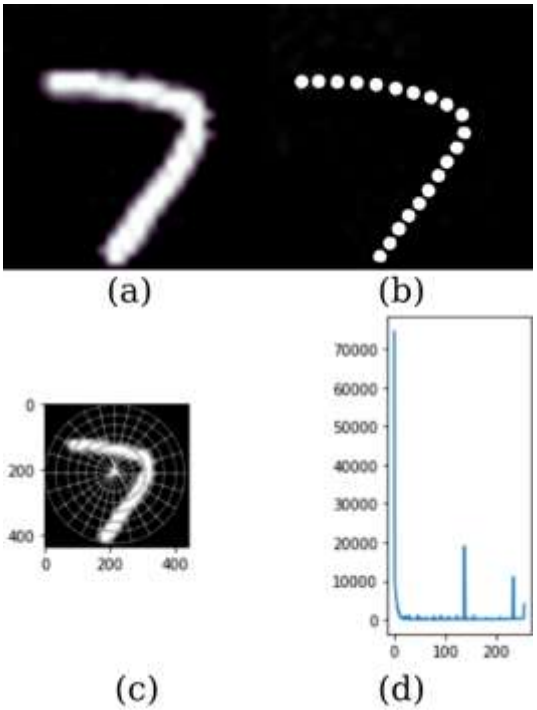


Figure 12. *MNIST dataset- number 7, (a) original data, (b) the data in which it determines the 20 point, (c) data determined by angle and distance, (d) histogram of data c*

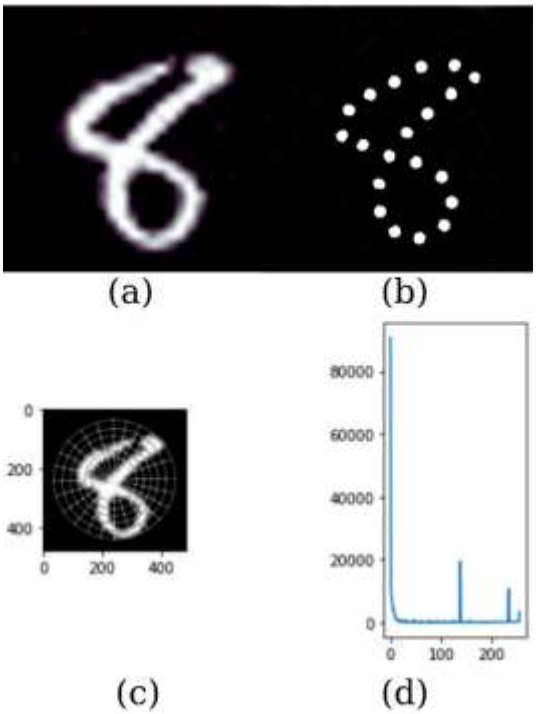


Figure 13. *MNIST dataset- number 8, (a) original data, (b) the data in which it determines the 20 point, (c) data determined by angle and distance, (d) histogram of data c*

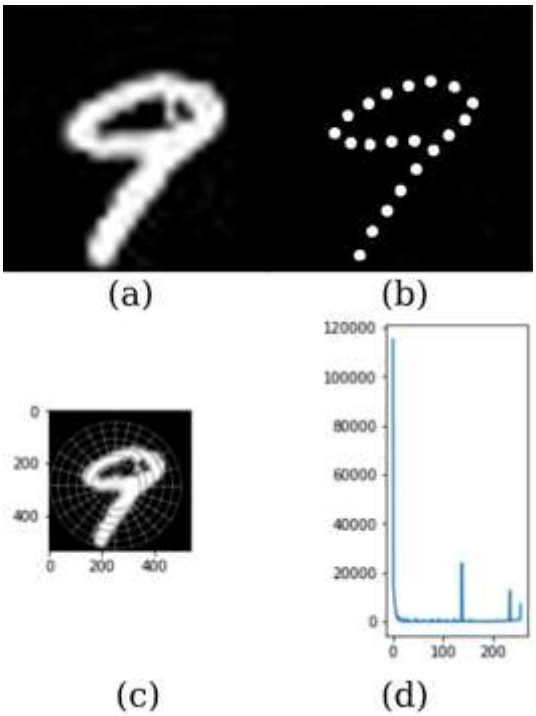


Figure 14. *MNIST dataset- number 9, (a) original data, (b) the data in which it determines the 20 point, (c) data determined by angle and distance, (d) histogram of data*

3. Results and Discussions

A fast shape descriptor is proposed in this study. MNIST dataset is used. A success of 69,91% is obtained using ReLU activation function with 1024 neurons and 25 epochs. Results for various numbers of neurons is given in Table 1. Results for various number of epochs is given in Table 2.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1]Viola, P. ve Jones, M. (2001, Aralık). Basit özelliklerin artırılmış bir dizisini kullanarak hızlı nesne algılama. Bilgisayarla görü ve örüntü tanıma üzerine 2001 *IEEE bilgisayar topluluğu konferansının Bildirilerinde. CVPR 2001* (Cilt 1, s. II). IEEE.

[2]V. V. Nabiyev, Artificial Intelligence. *Sözkesen Matbaacılık*, Ankara, 2005

[3]Wu, M. ve Zhang, Z. (2010). Mnist veri setini kullanarak el yazısı rakam sınıflandırması. *Ders projesi CSE802: Örüntü Sınıflandırma ve Analizi* .

[4]Sonugur, G., Çaki, E. E., Akan, S. A., & Gökçe, C. O. Gerçek Zamanlı İnsan Davranışı Anlamaya Doğru: Optimal-Altı Bir Şekil Tanımlayıcı. *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, *22*(4); 769-777.

[5]Taşcı, E., & Onan, A. (2016). K-en yakın komşu algoritması parametrelerinin sınıflandırma performansı üzerine etkisinin incelenmesi. *Akademik Bilişim*, 1(1); 4-18.

International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 9-No.1 (2023) pp. 6-10
http://dergipark.org.tr/en/pub/ijcesen

ISSN: 2149-9144

Research Article

Assessment of Gamma Ray Shielding Properties for Skin

Roya Boodaghi MALIDARRE[1*], Huseyin Ozan TEKIN[2,3], Kadir GUNOGLU[4], Hakan AKYILDIRIM[5]

[1]Payame Noor University, Tehran-Iran
* **Corresponding Author Email:** roya_boodaghi@yahoo.com - **ORCID:** 0000-0003-4505-7900

[2]Medical Diagnostic Imaging Dep., College of Health Sciences, University of Sharjah, Sharjah, United Arab Emirates
[3]Istinye University, Faculty of Engineering and Natural Sciences, Computer Engineering Dep., Istanbul 34396, Turkiye
Email: tekin765@gmail.com - **ORCID:** 0000-0002-0997-3488

[4]Isparta Applied Science University, 32200, Isparta-Turkiye
Email: kadirgunoglu@isparta.edu.tr - **ORCID:** 0000-0002-9008-9162

[5]Suleyman Demirel University, 32200, Isparta-Turkiye
Email: hakanakyildirim@sdu.edu.tr - **ORCID:** 0000-0001-5723-958X

Article Info:

DOI: 10.22399/ijcesen.1247867
Received : 05 February 2023
Accepted : 05 March 2023

Keywords

Radiation
Skin
Shielding

Abstract:

Gamma ray is uncharged radiation type and having high energy it can ionize any atom and thus can damage human cells. Because of this harmful effect cell should be protected. Besides developing new alternative to lead and lead based materials, it should be interesting to obtain shielding properties of skin. This paper presents a results on the shielding properties of skin.

1. Introduction

Radiation is due to the natural or artificial sources and it is used in wide variety of fields from medicine to many different commercial facilities [1-5]. On the other hand there is a negative effects of the radiation on the human cell and thus the researchers focused on this subject in order to set a limit the exposure. Radiation dosimetry have been developed for this purposes and namely time-distance and shielding were set as radiation protection rules. The conventional shielding materials for several years are lead, tungsten, and other heavy elements. Besides these materials which have a high absorption rate, researcher developed new alternative materials due to negative aspect of this conventional materials [6-20]. In order to obtain cell from harmful radiation effect, the shielding character of the skin itself should be known. Thus in this paper gamma ray shielding properties have been obtained using Phy-X/PSD code.

2. Materials and Methods

The gamma ray shielding properties have been calculated using Phy-X/PSD code [21]. This is done by obtaining linear attenuation coefficients (LAC), mean free path (mfp), half-value length (HVL), thenth value length (TVL).
The LAC is defined as the probability of gamma ray interaction with materials and given as in equation 1.

$$\mu = n\sigma \qquad (1)$$

where μ is LAC, n is the atomic numbers per volume and σ is the cross section.
With the help of the LAC, the mfp, HVL and TVL were calculated. The mfp shows the path where there will be no any interaction and it is given as in equation 2.

$$mfp = \frac{1}{\mu} \qquad (2)$$

The HVL is defined as the length where gamma ray did not interact with the substrate reduced to half its initial value and it is given as in equation 3.

$$HVL = \frac{Ln(2)}{\mu} \qquad (3)$$

The TVL is defined as the length where gamma ray did not interact with the substrate reduced to tenth its initial value and it is given as in equation 4.

$$TVL = \frac{Ln(10)}{\mu} \qquad (4)$$

3. Results and Discussions

The gamma shielding properties of skin have been investigated. For this purposes the obtained LAC has been shown in Fig.1. It can be seen from this figure that the interaction of gamma rays with medium is energy dependent. Using LAC the mfp have been obtained and shown in Fig.2 where it can be seen that the mfp is energy

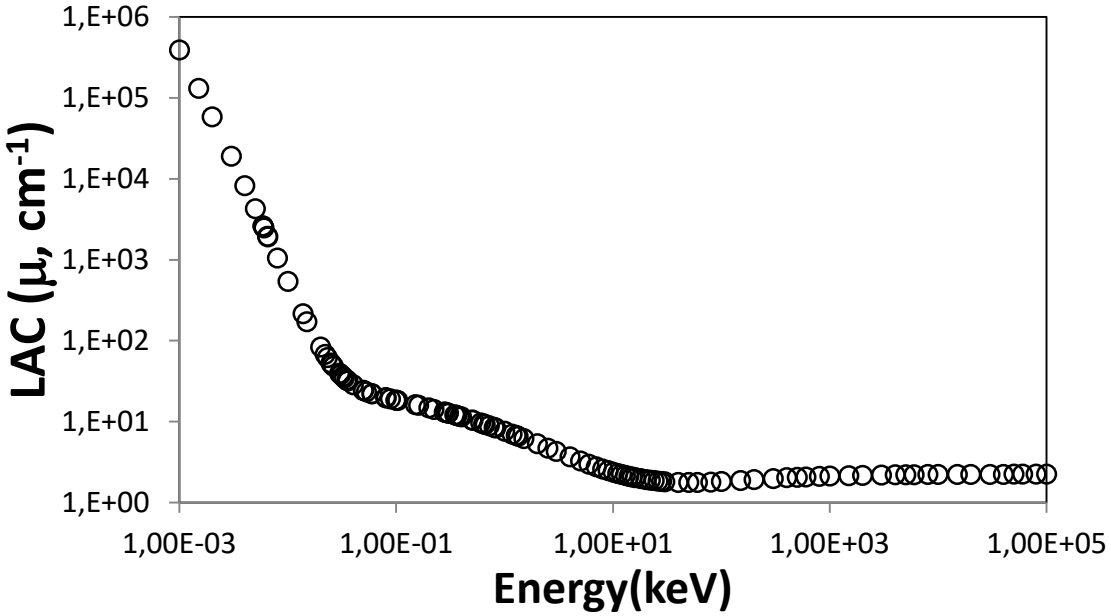


Figure 1. *Obtained LAC as a function of gamma ray energies*

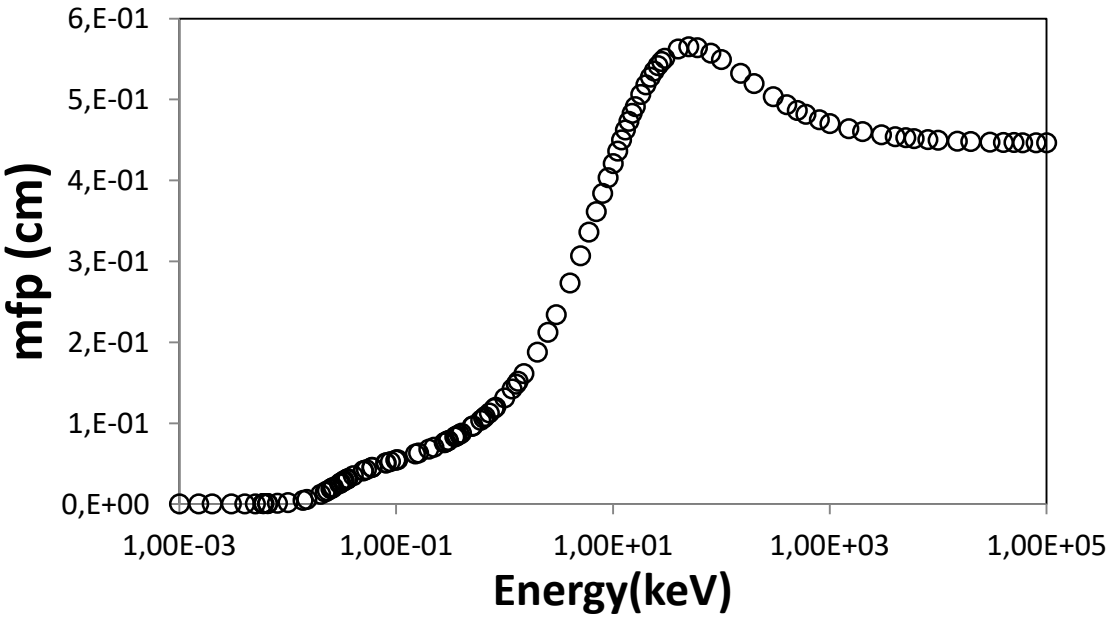


Figure 2. *Obtained mfp as a function of gamma ray energies*

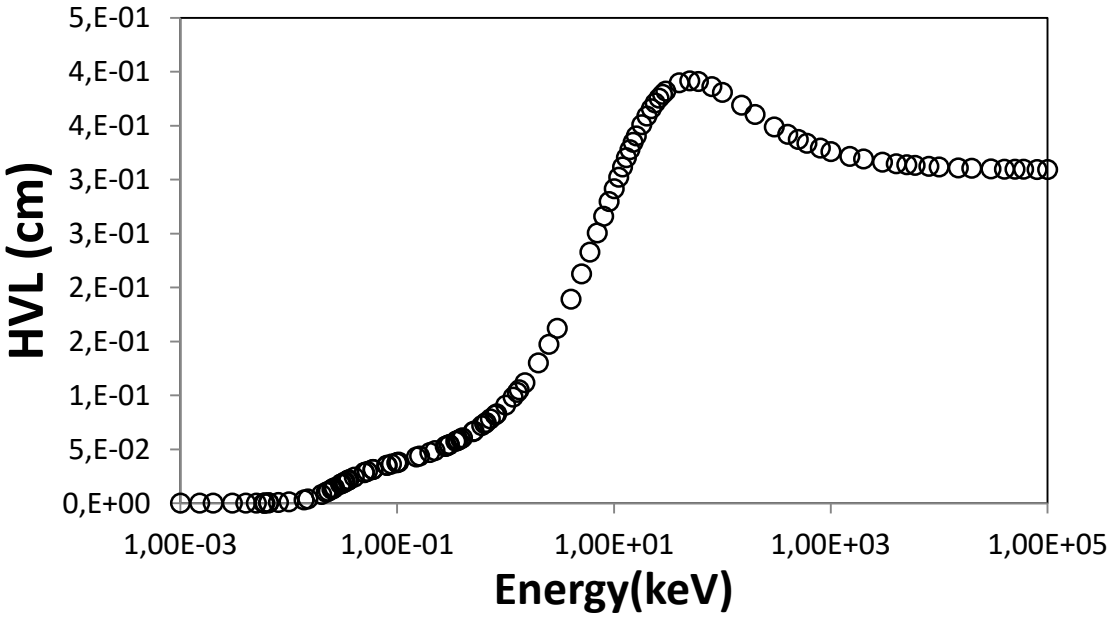


Figure 3. *Obtained HVL as a function of gamma ray energies*

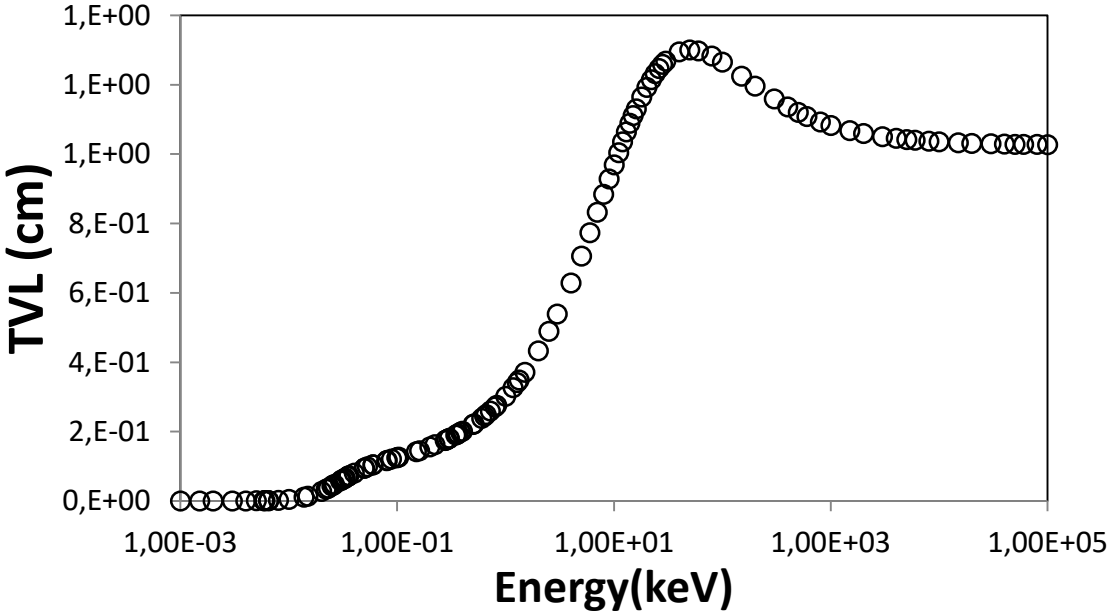


Figure 4. *Obtained TVL as a function of gamma ray energies*

dependent. One of the crucial characteristics for defining the shielding effectiveness is HVL values and obtained results were illustrated as a function of photon energy in Fig. 3. It can be seen from this figure that a similar distribution with the mfp has been obtained. Moreover, in order to confirm the validity shielding properties of skin the TVL values were also obtained. The obtained results were displayed in Fig.4 where it is also seen a similar behave with mfp and also HVL.

4. Conclusions

In the current study, various parameters have been obtained for radiation shielding properties of skin. The shielding performances of the skin in the study were examined for the energy range of 10^{-3}-10^{5} keV. It can be concluded from this work that the energy increases, the value of the linear attenuation coefficient decreased.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have

appeared to influence the work reported in this paper

- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available n request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1]Iskender Akkurt. (2009). Effective atomic and electron numbers of some steels at different energies. *Annals of Nuclear. Energy*. 36(11-12):1702-1705. DOI:10.1016/j.anucene.2009.09.005

[2]Aljawhara H. Almuqrin, M.I. Sayyed, Ashok Kumar, B.O. El-bashir, I. Akkurt. (2021). Optical, mechanical properties and gamma ray shielding behavior of TeO2-Bi2O3-PbO-MgO-B2O3 glasses using FLUKA simulation code. *Optical Materials* 113;110900 https://doi.org/10.1016/j.optmat.2021.110900

[3]Iskender Akkurt (2007). Effective Atomic Numbers for Fe–Mn Alloy Using Transmission Experiment *Chinese Phys. Lett.* 24;2812-2814. https://doi.org/10.1088/0256-307X/24/10/027

[4]Ghada ALMisned, Duygu Sen Baykal, Gokhan Kilic, G. Susoy, Hesham M.H. Zakaly, Antoaneta Ene, H.O. Tekin. (2022). Assessment of the usability conditions of Sb2O3-PbO-B2O3 glasses for shielding purposes in some medical radioisotope and a wide gamma-ray energy spectrum. *Applied Rheology*, 32(1);178-189 DOI: 10.1515/arh-2022-0133

[5]Ghada ALMisned, Duygu Sen Baykal, G. Susoy ,Gokhan Kilic, Hesham M.H. Zakaly, Antoaneta Ene, H.O. Tekin. (2022). Determination of gamma-ray transmission factors of WO3–TeO2–B2O3 glasses using MCPX Monte Carlo code for shielding and protection purposes. *Applied Rheology*, 32(1):166-177 DOI: 10.1515/arh-2022-0132

[6]Çelen, Y.Y., Akkurt, İ. & Kayıran, H.F. (2021). Gamma ray shielding parameters of barium tetra titanate (BaTi4O9) ceramic. *J Mater Sci: Mater Electron* 32;18351–18362. https://doi.org/10.1007/s10854-021-06376-6

[7]Akkurt, I., Malidarre, R.B. & Kavas, T. (2021). Monte Carlo simulation of radiation shielding properties of the glass system containing Bi2O3. *Eur. Phys. J. Plus* 136;264. https://doi.org/10.1140/epjp/s13360-021-01260-y

[8]Ural, A. & Kilimci, Z. H. (2021). The Prediction of Chiral Metamaterial Resonance using Convolutional Neural Networks and Conventional Machine Learning Algorithms. *International Journal of Computational and Experimental Science and Engineering*. 7(3);156-163 . DOI: 10.22399/ijcesen.973726

[9]Zarkooshi, A. , Latıf, K. H. & Hawı, F. (2021). Estimating the Concentrations of Natural Isotopes of 238 U and 232Th and Radiation Dose Rates for Wasit Province-Iraq by Gr-460 system. *International Journal of Computational and Experimental Science and Engineering.* 7(3);128-132. DOI: 10.22399/ijcesen.891935

[10]Caymaz, T. , Çalışkan, S. & Botsalı, A. R. (2022). Evaluation of Ergonomic Conditions using Fuzzy Logic in a Metal Processing Plant. *International Journal of Computational and Experimental Science and Engineering*. 8(1);19-24. DOI: 10.22399/ijcesen.932994

[11]Arbouz, H. (2022). Modeling of a Tandem Solar Cell Structure Based on CZTS and CZTSe Absorber Materials. *International Journal of Computational and Experimental Science and Engineering*. 8(1);14-18 . DOI: 10.22399/ijcesen.843038

[12]Çilli, A. , Beken, M. & Kurt, N. (2022). Determination of Theoretical Fracture Criteria of Layered Elastic Composite Material by ANFIS Method from Artificial Intelligence. *International Journal of Computational and Experimental Science and Engineering.* 8(2);32-39. DOI: 10.22399/ijcesen.1077328

[13]Rwashdı, Q. A. A. D. , Waheed, F. , Gunoglu, K. & Akkurt, İ. (2022). Experimental Testing of the Radiation Shielding Properties for Steel. *International Journal of Computational and Experimental Science and Engineering.* 8(3);74-76. DOI: 10.22399/ijcesen.1067028

[14]MALIDARRE, R.B., AKKURT, I., Kavas, T. (2021). Monte Carlo simulation on shielding properties of neutron-gamma from 252Cf source for Alumino-Boro-Silicate Glasses. *Radiation Physics and Chemistry*.186;109540. https://doi.org/10.1016/j.radphyschem.2021.109540

[15]Waheed, F. , İmamoğlu, M. , Karpuz, N. & Ovalıoğlu, H. (2022). Simulation of Neutrons Shielding Properties for Some Medical Materials. *International Journal of Computational and Experimental Science and Engineering*. 8(1);5-8. DOI: 10.22399/ijcesen.1032359

[16]Boodaghi Malidarre, R. , Akkurt, İ. , Gunoglu, K. & Akyıldırım, H. (2021). Fast Neutrons Shielding Properties for HAP-Fe2O3 Composite Materials. *International Journal of Computational and Experimental Science and Engineering*. 7(3);143-145. DOI: 10.22399/ijcesen.1012039

[17]Huseyin Ozan Tekin, Baris CAVLI, Elif Ebru ALTUNSOY, Tugba MANICI, Ceren OZTURK, Hakki Muammer KARAKAS (2018). An Investigation on Radiation Protection and Shielding Properties of 16 Slice Computed Tomography (CT) Facilities. *International Journal of Computational and Experimental Science and Engineering*. 4(2);37 – 40. https://doi.org/10.22399/ijcesen.408231

[18]Sarihan Mucize. (2022). Simulation of gamma-ray shielding properties for materials of medical interest.

Open Chemistry. 20(1);81-87. https://doi.org/10.1515/chem-2021-0118

[19]Şen Baykal, D , Tekin, H , Çakırlı Mutlu, R . (2021). An Investigation on Radiation Shielding Properties of Borosilicate Glass Systems. *International Journal of Computational and Experimental Science and Engineering*. 7(2);99-108. DOI: 10.22399/ijcesen.960151

[20]Oruncak Bekir (2022). Gamma-ray Shielding Properties of Nd2O3 added Iron-Boron-Phosphate based composites. *Open Chemistry* 20(1);237-243 https://doi.org/ 10.1515/chem-2022-0143

[21] Sakar E., Ozpolat O. , Fırat, Alım, B., Sayyed, M.I., Kurudirek, M., (2020). Phy-X/PSD: development of a user friendly online software for calculation of parameters relevant to radiation shielding and dosimetry. *Radiat. Phys. Chem.* 166;108496.

International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 9-No.1 (2023) pp. 11-16
http://dergipark.org.tr/en/pub/ijcesen

ISSN: 2149-9144

Research Article

Using Linear Regression For Used Car Price Prediction

Sümeyra MUTİ[1], Kazım YILDIZ[2,*]

[1]Marmara University, Institute of Pure and Applied Sciences, Department of Computer Engineering, Istanbul, Turkiye
Email: sumeyramuti@marun.edu.tr **ORCID:** 0000-0001-6489-0258

[2] Marmara University, Technology Faculty, Department of Computer Engineering, 34722, Istanbul, Turkiye
* **Corresponding Author Email** : kazim.yildiz@marmara.edu.tr **ORCID:** 0000-0001-6999-1410

Article Info:

DOI: 10.22399/ijcesen.1070505
Received : 09 February 2022
Accepted : 26 February 2023

Keywords

Machine Learning
Car Price Prediction
Linear Regression

Abstract:

Recently, there have been studies on the use of machine learning algorithms for price prediction in many different areas such as stock market, rent a house and used car sales. Studies give information about which algorithm is more successful in price prediction using different machine learning methods. The most commonly used method for price prediction is the linear regression model. In this study, the effectiveness of the linear regression model was examined for used car price prediction. The linear regression model was applied to the data set that includes the features and price information of vehicles in Turkey as the year 2020. As a result, when we selected 1/3 of the data set as the test data, it was observed that the R2 score for the prediction success of model was 73%. To improve the effectiveness of the results the dataset could be extend or preprocessing part be detailed.

1. Introduction

Selling process of a vehicle, the most important issue is to determine the most affordable price without giving a price below the value of the car. If a value is given above the market price, the probability of selling the car will either decrease or the selling time may be longer. In addition, no one, who wants to buy a used car, does not want to own a car by paying more than it is worth. The price of a used vehicle may vary depending on attributes such as the vehicle's model, year, number of kilometers, gear type, damage record, color, and additional comfort features. Using this data on a used car sale site, it is possible to calculate how much a used vehicle can be sold according to its features, using machine learning algorithms.

Machine learning methods are algorithms that process the given data to perform a specified task without being fully programmed, learn with this data and improve itself as a result of learning [1]. There are studies on the use of machine learning algorithms in many different areas such as the detection of COVID-19 cases [2], crop yield analysis [3], visiting time prediction [4], predicting motor vehicle theft [5]. In this study, we will apply the linear regression (LR) method as a machine learning method for used car price prediction. LR is both a statistical method and a machine learning method. It is aimed to create models to identify the relationship between input and output variables [6].

There are many studies using various machine learning techniques in price estimation applications. Home price prediction [7-10], stock price prediction [11-14], stock market prediction [15], bitcoin price prediction [16], price estimate for vacation rentals [17], car price prediction [18-19] are some of them. In this study the main focus is on used vehicle price estimation. Comparison of LR, Artificial Neural Networks (ANN) and Support Vector Machine (SVM) algorithms for price estimation of used truck models is one of these studies [20]. In another study, the success of the LR method for used car price estimation has been examined [21].

In a study for the estimation of house prices, different machine learning algorithms has been applied by analyzing the housing data of 5359 townhouses in Virginia collected by the Multiple Listing Service (MLS) of Metropolitan Regional Information Systems (MRIS). Data collected

through WEKA then classified various machine learning algorithms such as C4.5, RIPPER Naïve Bayesian and AdaBoost. It has been observed that the RIPPER model gives more successful results compared to others[7].

In another home price estimation study, Support Vector Regression (SVR) and LR methods were compared and it was found that LR had lower error rates [22]. The success of RF and LR methods in home price estimation was compared. It was observed that the training data being more than the test data increased the estimation success of the system and it was concluded that the LR method, which includes 80% data training and 20% data testing, gives lower Absolute Error and RMSE values [23]. A model was developed which predicts prices by using both textual and visual data of houses in housing price estimation. In order to determine the luxury level of the house, the KNN, SVM and Convolutional Neural Network (CNN) methods applied on the visual data were compared and it was observed that the CNN method gave the lowest Median Error Rate. It has been observed that the created model is more successful than Zillow's Zestimate method [24]. In another application stock price estimation was made, the decision tree model, multiple regression and random forest algorithms applied to five different stocks tried to predict the closing prices of the stocks the next day and the experimental results of these algorithms are successful [12].

In a study in which LR and SVM algorithms used to predict stock prices, it has been that LR give better results than SVM [15].

There is a study examining the effect of various machine learning algorithms on price estimation for Airbnb, which offers vacation rental services. In this study, LR, tree-based models, Decision Support Machines and ANN methods were applied to get the best results in terms of Mean Square Line, Mean Absolute Error (MAE) and R2 score. Among the methods tested, SVR gave the best result with an R2 score of 69% [17].

In a study conducted for vehicle price estimation, some machine learning algorithms are used and compared for an application that finds the best price that a truck company can give when buying used vehicles from customers. For this study, the previous offers of the company were used as the data set. LR, ANN and SVM were used for price estimation and the most successful results were obtained by applying 90% data splitting method with the SVM algorithm [20].

In another application, LR method was applied for used car price estimation. The data of 5041 used cars, in which 23 features that affect the price (such as brand and model) were selected, were used and the explanatory rate of these features was found to be 89.1%. The data set was divided into two, half of which was used as training data and the other half as test data, and when the results were compared, the prediction success rate was found to be 81.15%. More successful results can be obtained with more training data [21]. In another application for car price estimation, price prediction success was examined by applying various regression techniques on a data set containing information such as fuel type, mileage and model of vehicles collected from an e-commerce site called Avito. As a result, it was revealed that the Gradient Boosting Regression (GBR) method had the highest R2 score and the least MAE value [25].

In this paper, Section 2 describes the methodology and material, the research results was given in section 3. Finally conclusion and future work details were given.

2. Material and Methods

2.1. Dataset

As shown in Table 1, the dataset contains information about 15 attributes of used car in Turkey in 2020, including car brand, date of announcement, vehicle type group (such as Clio), vehicle type (such as 520i Standard), model year, fuel type, gear type, CCM, horsepower, color, body type (such as sedan), from whom it is sold (from the gallery) etc.), condition (used or new), mileage, price [26]. Since the date of the announcement and the information from whom it was sold have no effect on the price or are negligible, they will be removed from the data set. Horse power column has 61% “unknown” data so will not be included in this study. Status information, on the other hand, will be used distinctively since we will only make price estimations on used vehicles, and only data with second-hand status will be used in the data set. Also CCM data will be ignored.

In this study, the LR algorithm was examined for used car price estimation. Classification and LR processes will be performed with Python [27] in the JupyterLab environment.

2.2. Data Preprocessing

Data preprocessing is used to improve the performance of machine learning methods, especially when it comes to classification. The dataset may initially be noisy and inconsistent. With cleaning activities, such as removing outliers and correcting noisy values, the data set is made suitable for machine learning applications [28].

Table 1. *Sample of the data set used for car price prediction*

Date Annou	Brand	Vehicle Type Group	Vehicle Type	Model Year	Fuel Type	Gear Type	CCM	Horse Power	Color	Body Type	From Who	State	Km	Price
27.05. 2020	Jaguar	XF	2.0 D Prestige Plus	2017	Diesel	Autom atic	1801-2000 cc	176-200 HP	Navy Blue	Hatachback 5 Door	Galery	Used	26100	634500
16.06. 2020	Acura	CL	-	2015	Diesel	Semi-automa tic	1301-1600cc	101-125 HP	Blue	Sedan	Owner	Used	127000	151500
14.06. 2020	Acura	CL	2.2	1194	Benzine/ LPG	Manuel	1301-1600cc	101-125 HP	Turquoise	Sedan	Owner	Used	175000	19750
11.06. 2020	Acura	CL	-	2013	Diesel	Manuel	1301-1600cc	76-100 HP	Brown	Sedan	Owner	Used	325	52000
11.06. 2020	Acura	CL	2.2	2010	Diesel	automa tic	1801-2000cc	151-175 HP	White	Sedan	Owner	used	207000	148750

Incorrectly added data should be removed from the data set, as it will reduce the probability of the system guessing correctly. Similarly, removing data that has no effect on the estimation from the data set will increase the success of the algorithm. Outliers, which are far from other data in the data set, are also among the factors that reduce the success of the algorithm. It is possible to reduce these negative effects with data preprocessing.

We removed the information that was not useful to us, such as "Unknown", "-" from the data set. Then, we detected the outlier data of “price” attribute and removed from the dataset. We found that the price difference was large in two vehicles of the “Jaguar” brand with the same features, and we removed the low price from the dataset. We observed that the accuracy of the prediction decrease when the “price” value rises above 200000, so we also removed the data from the dataset with the “price” value above 200000. Since some mileage information is too low for a used vehicle, we removed lines below 1000 km from the dataset.

We converted all our data into numerical values in order to use machine learning algorithms. After these processes, the numbers of data features in the dataset are shown in Figure 3 and the relationship of the features with each other is shown in Figure 1.

3. Results and Discussions

We partition our dataset as 1/3 test data and train it with LR model. The relationship between the estimated price values and the actual price values in Figure 2 shows that the LR model is suitable for price prediction. The blue solid line shows the regression line and the red dots around it show the intersection of the estimated values and the real values. The closer the red dots are to the blue line, the more accurate the algorithm is making predictions.

As a result, we found that the R2 score [29] which gives the relationship between the real and estimated values of the LR model we applied to our data set was 0.73. While our model success was around 0.62 at the beginning, we observed that our model success reached this value with the data preprocessing process.

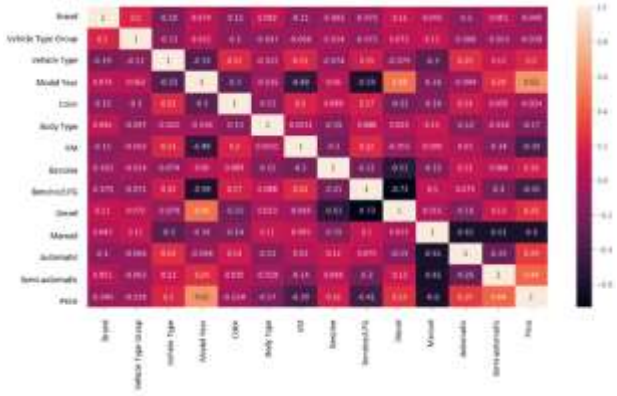


Figure 1. *Correlation of attributes*

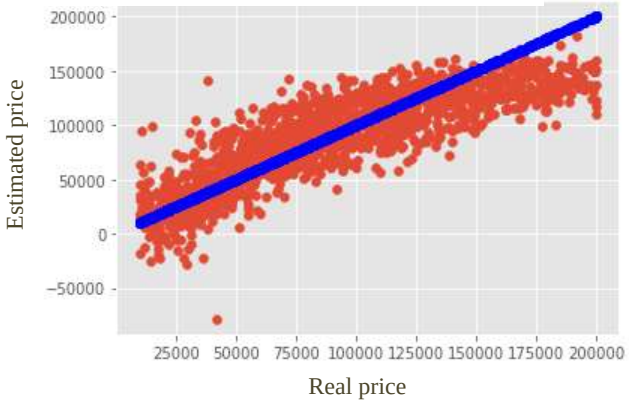


Figure 2. *Correlation between estimated and real price values*

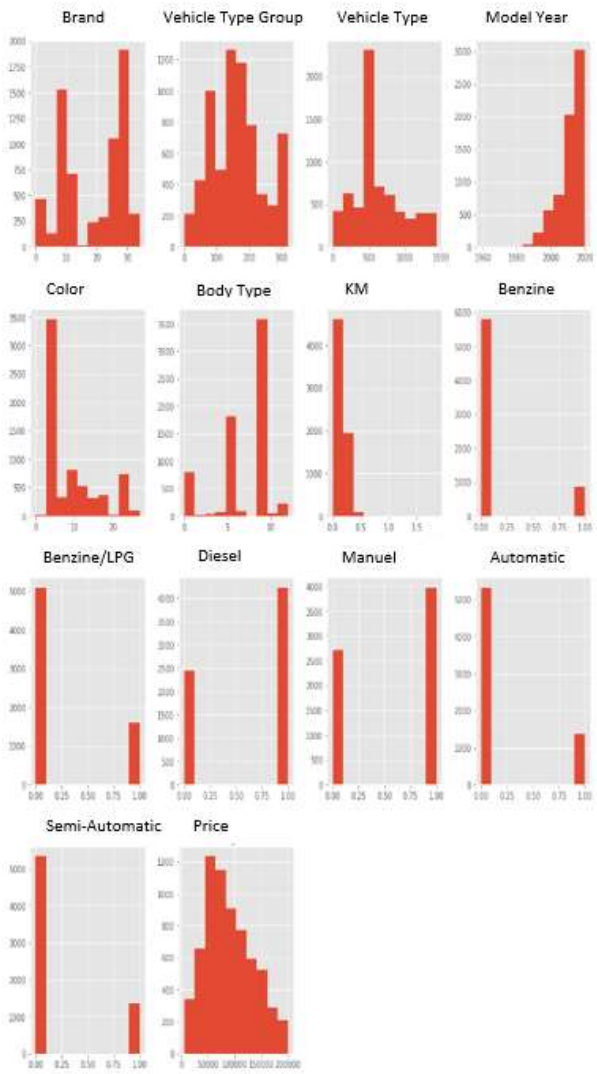


Figure 3. *Amount of attributes in the dataset*

Table 2 shows the comparison of the estimated price values of our model with the actual price values. Accordingly, we see that the difference between model estimates and actual values is quite close in some areas, and this difference increases in some areas. For example, the estimated price of the vehicle whose actual price is 56750 in line 6 is found 60003, and a successful estimate was obtained with a difference of approximately 3000 liras. On the other hand, there is a difference of approximately 40000 TL between the estimate and the actual value in the 2nd line. A more detailed examination of the data preprocessing section or the use of a more suitable data set can increase the success of the model.

4. Conclusions

LR method is a suitable model for price estimation, as features such as brand, km, year, gear type of used vehicles directly affects their price. We observed that the success of the model increased when we removed the non-logical values in the data set.

Table 2. *Real price vs estimated price values*

	Real value	Prediction
0	4800	31320.27
1	145000	105211.80
2	49250	64534.34
3	118000	133167.16
4	103500	108646.10
5	56750	60063.20
6	104000	113513.86
7	32900	38927.42
8	142950	135422.77
9	60000	67108.96
10	81650	91284.90
11	122500	125503.39
12	93000	100478.09
13	54900	70027.75
14	93500	120581.50
15	52950	83542.02
16	140500	131251.74
17	65500	109028.45
18	103500	124890.43
19	155000	109971.27

While our R^2 score was around 0.62 in the first stage, it increased to 0.73 when we removed the data that reduced the predictive ability of the model in the data preprocessing section. As seen in Figure 2, the relationship between estimated and actual price values shows that LR is suitable for used car price estimation. More appropriate datasets can be used to obtain better results.

In order to examine the success of LR with more up-to-date data, a study can be done by pulling the appropriate data from used car buying sites.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.

- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available n request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] I. E. Naqa and M. J. Murphy. (2015). What Is Machine Learning ?. *Machine Learning in Radiation Oncology* 3;11 DOI 10.1007/978-3-319-18305-3_1.

[2] N. S. Özen, S. Saraç and M. Koyuncu, (2021). COVID-19 Vakalarının Makine Öğrenmesi Algoritmaları ile Tahmini: Amerika Birleşik Devletleri Örneği. *European Journal of Science and Technology (EJOSAT)*. 22;134-139 DOI: 10.31590/ejosat.855113.

[3] F. F. Haque, A. Abdelgawad, V. P. Yanambaka and K. Yelamarthi. (2020). Crop Yield Analysis Using Machine Learning Algorithms. *IEEE 6th World Forum on Internet of Things (WF-IoT)*. pp. 1-2, DOI: 10.1109/WF-IoT48130.2020.9221459.

[4] I. Hapsari, I. Surjandari and K. (2018). Visiting Time Prediction Using Machine Learning Regression Algorithm. *6th International Conference on Information and Communication Technology (ICoICT)*. pp. 495-500, DOI: 10.1109/ICoICT.2018.8528810.

[5] N. Nafi'iyah and K. F. Mauladi. (2021). Linear Regression Analysis and SVR in Predicting Motor Vehicle Theft. *International Seminar on Application for Technology of Information and Communication (iSemantic)*, pp. 54-58 DOI: 10.1109/ISEMANTIC52711.2021.9573225.

[6] Ms Kavita and P. Mathur. (2020). Crop Yield Estimation in India Using Machine Learning. *5th International Conference on Computing Communication and Automation (ICCCA)*, pp. 220-224. DOI: 10.1109/ICCCA49541.2020.9250915.

[7] J. K. Bae and B. Park. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert systems with applications*. 42(6);2928-2934, DOI: 10.1016/j.eswa.2014.11.040.

[8] A. Varma, A. Sarma, S. Doshi and R. Nair. (2018). House Price Prediction Using Machine Learning and Neural Networks. *Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 1936-1939, DOI: 10.1109/ICICCT.2018.8473231.

[9] I. Imran, U. Zaman, M. Waqar, A. Zaman. (2021) Using Machine Learning Algorithms for Housing Price Prediction: The Case of Islamabad Housing Data. *Soft Computing and Machine Intelligence*, 1(1);11-23.

[10] B. Jia (2021). Computer mathematical statistics applied in the housing price investigation through machine learning and linear regression model. *International Conference on Data Science and Computer Application (ICDSCA)*, pp. 769-772, DOI: 10.1109/ICDSCA53499.2021.9650136.

[11] C. K.-S. Leung, R. K. MacKinnon and Y. Wang. (2014). A machine learning approach for stock price prediction. *IDEAS '14: Proceedings of the 18th International Database Engineering & Applications Symposium*, pp. 274-277, DOI:10.1145/2628194.2628211.

[12] Z. D. Akşehir and E. Kılıç, (2019). Makine Öğrenmesi Teknikleri ile Banka Hisse Senetlerinin Fiyat Tahmini. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 12(2);30.

[13] M. Nikou, G. Mansourfar, and J. Bagherzadeh (2019). Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management*, 26(4)164-174, DOI: 10.1002/isaf.1459.

[14] W. Lu, W. Ge, R. Li and L. Yang. (2021). A Comparative Study on the Individual Stock Price Prediction with the Application of Neural Network Models. *International Conference on Computer Engineering and Artificial Intelligence (ICCEAI)*, pp. 235-238, DOI: 10.1109/ICCEAI52939.2021.00046,

[15] B. Panwar and P. J. Gaurav Dhuriya. (2021). Stock Market Prediction Using Linear Regression and SVM. *International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pp. 629-631, DOI: 10.1109/ICACITE51222.2021.9404733.

[16] V. Siddhi, S. Valecha and M. Shreya. (2018). Bitcoin price prediction using machine learning. *20th International Conference on Advanced Communication Technology (ICACT)*, pp. 144-147, DOI: 10.23919/ICACT.2018.8323676.

[17] P. R. Kalehbasti, L. Nikolenko and H. Rezaei. (2021). Airbnb Price Prediction Using Machine Learning and Sentiment Analysis. *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*, pp. 173-184, DOI: 10.1007/978-3-030-84060-0_11.

[18] E. Gegic, B. Isakovic, D. Keco, Z. Masetic and J. Kevric, (2019). Car Price Prediction using Machine Learning Techniques. *TEM Journal*, 8(1)113, DOI: 10.18421/TEM81-16.

[19] S. Selvaratnam, B. Yogarajah, T. Jeyamugan and N. Ratnarajah, (2021). Feature selection in automobile price prediction: An integrated approach. *International Research Conference on Smart Computing and Systems Engineering (SCSE)*, 4;106-112, DOI: 10.1109/SCSE53661.2021.9568288.

[20] Namlı, E., Gül, E., and Ünlü, R., (2019). Fiyat Tahminlemesinde Makine Öğrenmesi Teknikleri ve Doğrusal Regresyon. *Konya Mühendislik Bilimleri Dergisi*, 7;806-821, DOI: 10.36306/konjes.654952.

[21] Çelik, Ö and Osmanoğlu, U.Ö, (2019). Prediction of The Prices of Second-Hand Cars. *European Journal of Science and Technology (EJOSAT)*, 16;77-83, DOI: 10.31590/ejosat.542884.

[22] Amaresh, V., Singh, R. R., Kamal, R., & Kulkarni, A. (2022). Linear Regression Models based Housing

Price Forecasting. *2022 International Conference on Industry 4.0 Technology (I4Tech)* (pp. 1-5). IEEE.

[23] Septianingrum, S. A., Dzikri, M. A., Soeleman, M. A., Pujiono, P., & Muslih, M. (2022). Performance Analysis of Multiple Linear Regression and Random Forest for an Estimate of the Price of a House. *International Seminar on Application for Technology of Information and Communication (iSemantic)* (pp. 415-418). IEEE.

[24] Nouriani, A., & Lemke, L. (2022). Vision-based housing price estimation using interior, exterior & satellite images. *Intelligent Systems with Applications*, 14;200081.

[25] M. Hankar, M. Birjali and A. Beni-Hssane (2022). Used Car Price Prediction using Machine Learning: A Case Study. *11th International Symposium on Signal, Image, Video and Communications (ISIVC)*, El Jadida, Morocco, pp. 1-4, doi: 10.1109/ISIVC54825.2022.9800719.

[26] KAGGLE (10.10.2021), https://www.kaggle.com/alpertemel/turkey-car-market-2020.

[27] G. V. Rossum, Python Development Team (2020). Python Tutorial Release 3.8.1 *The Python Software Foundation.*

[28] L. Moreira, C. Dantas, L. Oliveira, J. Soares and E. Ogasawara, (2018). On Evaluating Data Preprocessing Methods for Machine Learning Models for Flight Delays. *International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8, DOI: 10.1109/IJCNN.2018.8489294.

[29] D. Chicco, M. J. Warrens and G. Jurman, (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Peer J Computer Science*, 7; DOI: 10.7717/peerj-cs.623.

International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 9-No.1 (2023) pp. 17-19
http://dergipark.org.tr/en/pub/ijcesen

ISSN: 2149-9144

Research Article

Disturbance Rejection Performance Comparison of PSO and ZN Methods for Various Disturbance Frequencies

Celal Onur GÖKÇE[1]*, Volkan DURUSU[2], Rıdvan ÜNAL[3]

[1] Afyon Kocatepe University, Software Engineering Department, Afyonkarahisar, Turkiye
***Corresponding Author**: **Email:** cogokce@aku.edu.tr - **ORCID:** 0000-0003-3120-7808

[2] Afyon Kocatepe University, Electrical and Electronics Engineering Department, Afyonkarahisar, Turkiye
Email: durusuvolkan@gmail.com - **ORCID:** 0000-0001-7726-5435

[3] Afyon Kocatepe University, Electrical and Electronics Engineering Department, Afyonkarahisar, Turkiye
Email: runal@aku.edu.tr - **ORCID:** 0000-0001-6842-7471

Article Info:

DOI: 10.22399/ijcesen.1202255
Received : 10 November 2022
Accepted : 17 January 2023

Keywords

PID control
Ziegler-Nichols
Particle Swarm Optimization

Abstract:

In this study Proportional-Integral-Derivative (PID) control of brushed DC Motor is analysed. The parameters of the PID controller are tuned with two different approaches, namely Ziegler-Nichols (ZN) and Particle Swarm Optimization (PSO). The system is tested under sinusoidal disturbance of varying frequencies in order to evaluate and compare disturbance rejection performances. It is shown that PSO approach has clearly higher performance compared with ZN approach for all disturbance frequencies. Simulations are done using Python programming language with trapezoid rule for differentiation and integration. Comments are done on results and future study is planned.

1. Introduction

DC motor is one of the most widely used electrical machines in industry. Speed control of DC motor has several applications including robotics, CNC and automation and electrical vehicles. Controlling speed of DC motor with no load is relatively simple problem which is studied quite exclusively. In real applications, DC motor is run with varying amount and types of loads. The logic lying behind PSO is to search the parameter space by considering the most successful points found. K. Khandani, A. A. Jalali and M. Alipoor designed PID controllers for time delay systems. PSO technique has been used to obtain optimal parameters of 2-DOF PID controller and compared with Genetic Algorithm (GA). Results has been shown that PSO has more performance than GA [1]. Krohling and Rey designed a PID controller for optimal disturbance rejection. The method which is used for study has handled the methodology as a constraint optimization problem. Servo motor has used for simulation and GA approach has used for optimize parameters [2]. Ibrahim, Hassan and Shomer compared Bacterial Foraging (BF) technique and PSO for determining the optimal parameters of PID controller in speed control of brushless dc motor. The proposed technique is shown that PSO method has more performance than BF especially in dynamic performance of the system [3]. Song, Xiao and Xu designed a random vibration PSO- Gravitational Searh Algorithm (GSA) based Fuzzy PI controller for brushless dc motor control. Results has given in study both simulation and experimental [4]. In this study, sinusoidal disturbance is used as a load to brushed dc motor for speed control. PSO and ZN methods are compared for various frequencies of disturbance and it is seen that PSO clearly has more performance than that of ZN especially in higher frequencies.

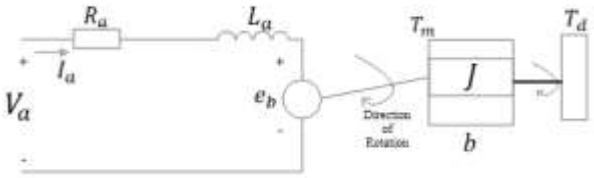


Figure 1: *Dc Motor Electrical and Mechanical Model*

Illustrative diagram of DC motor is shown in Figure 1. According to Kirchhoff, the sum of the voltages dissipated in a closed circuit is equal to the sum of the source voltages.

$$V_a = R_a i_a + L_a \frac{di}{dt} + e_b \quad (1)$$

In a conductor moving in a magnetic field, a reverse induced voltage of magnitude inversely proportional to the speed of movement occurs.

$$e_b = K_b w \qquad (2)$$

A force proportional to the magnitude of the current is applied to a current-carrying conductor while in a magnetic field.

$$T_m = K_m i_a \qquad (3)$$

Newton's Second Law of Motion is used to derive the mathematical model of the mechanical part.

$$T_m - T_d = J\frac{dw}{dt} + bw \quad (4)$$

Equation 5 is obtained when the transfer function of equation 4 is found.

$$\frac{w(s)}{T_m(s) - T_d(s)} = \frac{K_m}{1 + K_m J s} \quad (5)$$

Equation 6 is obtained when the transfer function of Equation 1 is found.

$$\frac{i_a(s)}{V_a(s) - e_b(s)} = \frac{1}{R_a + L_a s} \qquad (6)$$

The block diagram obtained for the DC Motor using Equations 5 and 6 is shown in Figure 2.

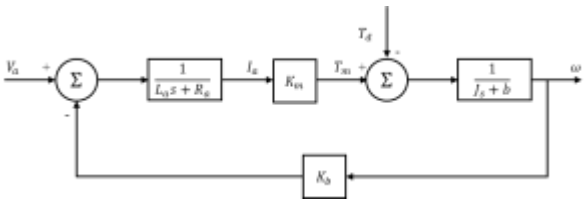


Figure 2: *DC Motor Block Diagram*

2. Material and Methods

PSO is a successful optimization technique, which is shown in literature [5], used especially for problems those are very hard or impossible to be solved analytically [6,7].

In this study position of each particle is a three-dimensional vector including Kp, Ki and Kd of the PID controller. For simulation Python programming language is used with Spyder IDE. Discrete time derivatives are calculated using trapezoid rule. DC motor parameters are taken realistically. Sinusoidal disturbance of various frequencies is given and sinusoidal reference input is used.

3. Results and Discussions

The motor has been tested under different loads and frequencies for performance evaluation. Results for various disturbance frequencies are shown in Figure 3, Figure 4, Figure 5 and Figure 6.

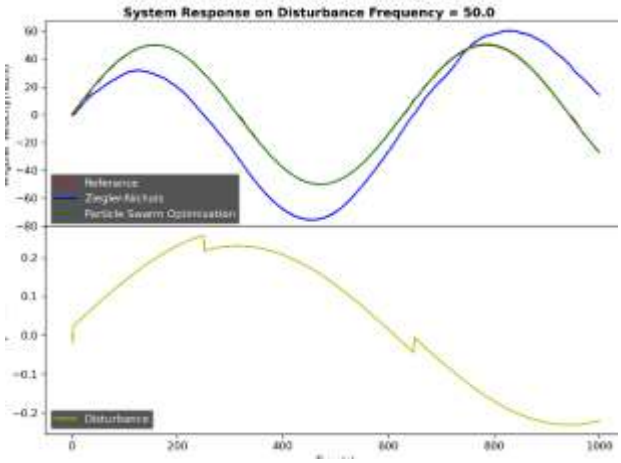


Figure 3: *System Response Under Disturbance Frequency = 50*

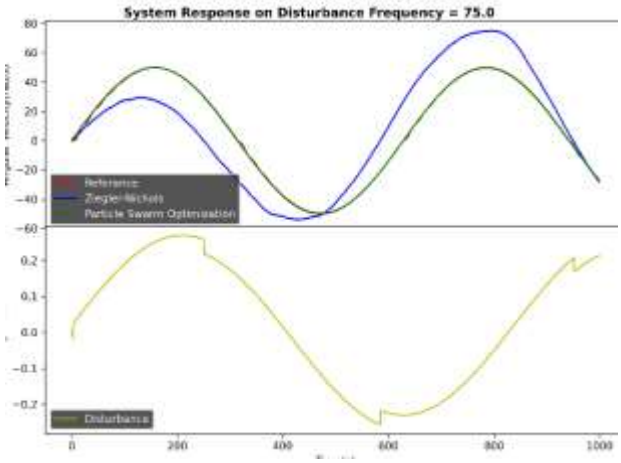


Figure 4: *System Response Under Disturbance Frequency = 75*

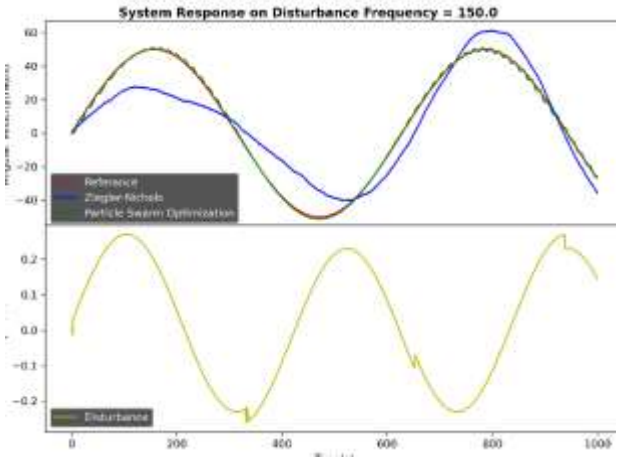


Figure 5: *System Response Under Disturbance Frequency = 150*

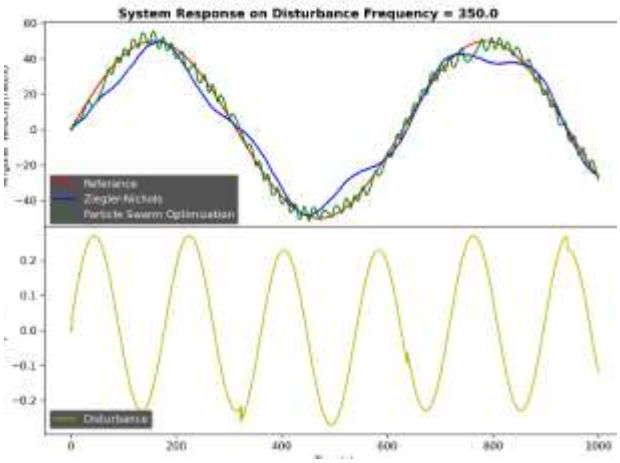


Figure 6: *System Response Under Disturbance Frequency = 350*

4. Conclusions

It is clearly seen that PSO has better disturbance rejection performance especially in higher disturbance frequencies. One of the reasons for the high success of the PSO algorithm in sinusoidal reference can be explained as the ease in which the reference value changes more slowly and the motor follows the reference. As a result, in all amplitudes and frequencies, the PSO algorithm outperformed the ZN method

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] K. Khandani, A. A. Jalali and M. Alipoor, (2009). Particle Swarm Optimization based design of disturbance rejection PID controllers for time delay systems. *IEEE International Conference on Intelligent Computing and Intelligent Systems*, pp. 862-866, doi: 10.1109/ICICISYS.2009.5358043.

[2] R. A. Krohling and J. P. Rey, (2001). Design of optimal disturbance rejection PID controllers using genetic algorithms. *IEEE Transactions on Evolutionary Computation*, 5(1);78-82. doi: 10.1109/4235.910467.

[3] H.E.A.Ibrahima, F.N.Hassan, Anas O.Shomer. (2014) Optimal PID control of a brushless DC motor using PSO and BF techniques. *Ain Shams Engineering Journal*, 2(5);391-398.

[4] Baoye Song, Yihui Xiao and Lin Xu (2020). Design of fuzzy PI controller for brushless DC motor based on PSO–GSA algorithm, *Systems Science & Control Engineering*, 8(1);67-77. DOI: 10.1080/21642583.2020.1723144.

[5] Yazgan H, Yener F, Soysal S, Gür A, (2019). Comparison Performances of PSO and GA to Tuning PID Controller for the DC Motor, *Sakarya University Journal of Science*, 23(2); 162-174.

[6] Ziegler J G, Nichols N B, 1942, Optimum Settings for Automatic Controllers, *Transactions of the American Society of Mechanical Engineers*, 64(11);759-765.

[7] Akyol S, Alataş B, (2012). Current Swarm Intelligence Optimization Algorithms. *Nevşehir University Journal of Graduate School of Natural and Applied Sciences*, 1(1);36-50.

International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 9-No.1 (2023) pp. 20-28
http://dergipark.org.tr/en/pub/ijcesen

ISSN: 2149-9144

Research Article

Exergy Analyses of Vehicles Air Conditioning Systems for Different Refrigerants

Arzu KEVEN*

Golcuk Vocational School, Kocaeli University, Kocaeli, Turkey
* **Corresponding Author** : **Email:** arzu.keven@kocaeli.edu.tr - **ORCID:** 0000-0003-0040-9167

Article Info:

DOI: 10.22399/ijcesen.1258770
Received : 02 March 2023
Accepted : 12 March 2023

Keywords

Refrigerants
Performance
Exergy

Abstract:

There are a limited number of studies in the literature that include detailed exergy analysis of vehicle air conditioning systems. In this study, in order to increase the performance of the air conditioning system in vehicles, a detailed exergy analysis has been made with the assumption that different refrigerants are used. R-134A, R-E245cb2, R-404A, R-1234ze(Z), R-161, R-1234zd(E), R-513A, R-1234ze(E) and R-1234yf has been chosen as the refrigerant. In the analysis, a comparison has been made by considering the environment, performance and safety values. While the COP values of the cycles increase with increasing evaporator temperatures, the COP values decrease at increasing condenser temperatures. On the other hand, exergy efficiency decreases with increasing evaporator and condenser temperatures. Also, it is aimed to evaluate all the elements of a vehicle air conditioning system with exergy analysis..

1. Introduction

Air conditioning and refrigeration is getting more important in our daily life and industry. In our daily life air conditioning and refrigeration are used in home, vehicles and other cooling demands. In industry, cooling, refrigeration and cryogenic applications are used widely. Many studies are carried out about the performance and exergy analyses of these applications [1, 2, 3]. There is an air conditioning system operating according to the vapor compression refrigeration cycle in the vehicles. and the compressor is powered directly from the internal combustion engine. Many studies are carried out to reduce losses and emissions in vehicle air conditioning systems. In the literature, studies on vehicle air conditioning have generally focused on improving driver comfort conditions, using different refrigerants, optimizing air conditioning elements and working conditions [4]. In addition to these studies, there are studies on refrigerants such as R-152a, R-430a, R-290, which can be an alternative to R-13a refrigerant in vehicle air conditioning systems [5-11]. Many studies have been conducted to determine the appropriate design of a bus air conditioner [12-14]. From the experimental studies, it has been obtained that the R-513A and 410-a refrigerant has higher performance than R-134a the refrigerants [15,16]. Also, Unal [17] is studied on the thermodynamic analysis of the vapor-compression refrigeration system with a two-phase ejector. The air conditioning system used in vehicles needs energy to operate. The energy source used in vehicles is fuel. Chemical energy is converted into mechanical work by burning fuel in the vehicle engine. Again, the alternator in the vehicle generates electrical energy with the mechanical drive it receives from the engine. All systems and elements in the vehicle consume mechanical or electrical energy during their operation. The place with the highest energy consumption in the air conditioning system is the compressors. Power consumption values must be determined in order to determine the net effect of the compressor in the air conditioning system on the engine. The refrigerants used in the air conditioning system also have a decisive effect on the efficient and effective operation of the air conditioning system. In this study, in order to increase the performance of bus air conditioners, detailed thermodynamic calculations of the cycles were done by using R-134A, R-E245cb2, R-404A, R-1234ze(Z), R-161, R-1234zd(E), R-513A, R-1234ze(E) and R-1234yf as the refrigerants. In the analyses, also a comparison has been studied by taking into account the safety values, environment and performance. The environmental friendliness of the analyzed refrigerants is also important. Thus, in this study, it is aimed to recommend energy saving and environmentally friendly refrigerants in air

conditioning systems. The results have been given in tables and graphs.

2. Material and Methods

2.1. Description of the vapor-compression refrigeration system (VCRS)

In the vapor compression mechanical cooling system, the refrigerant compressed in the compressor enters the condenser as vapor, as can be seen in Figure 1. In the condenser, the refrigerant condenses by giving off heat to the environment. Then, the refrigerant as a liquid enters in the throttling valve, and in the evaporator absorbs the heat of the cooling medium and cools the medium. The cycle repeats as the refrigerant in saturated vapor state from the evaporator, and goes to the compressor.

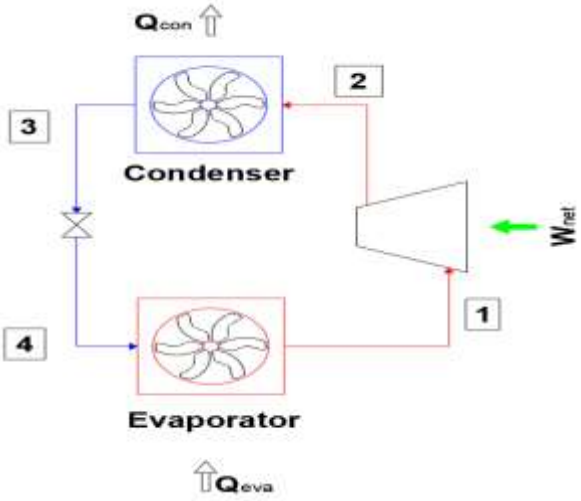


Figure 1. *The vapor-compression refrigeration system (VCRS)*

The main purposes of the use of vehicle air conditioning systems can be listed as heating, cooling and dehumidification processes. Vehicle air conditioning systems basically operate on a vapor compression refrigeration cycle [18]. An ideal vehicle air conditioning (AAC) system provides the appropriate temperature and humidity for human comfort [19].

2.2. Thermodynamic modelling

The equations, used in system analysis are given below.

The mass balances:

$$\sum \dot{m}_{in} = \sum \dot{m}_{out} \quad (1)$$

The energy balance:

$$\sum \dot{Q} - \sum \dot{W} = \sum H_{out} - \sum H_{in} \quad (2)$$

The performance coefficients of the analyzed cycles are calculated as follows:

$$COP_{vapour-comp} = Q_{eva}/W_{comp} \quad (3)$$

The exergy balance for the analyzed cycles is as follows [20].

$$\sum m_{in}\, e_{in} - \sum m_{out}\, e_{out} + \sum Q(1 - T/T_0 - \sum W - E_D = 0 \quad (4)$$

Specific exergy is;

$$e = (h - h_0) - T_0(s - s_0) \quad (5)$$

The exergy efficiency of the CRS

$$\eta_{ex} = \text{Exergy in product/Exergy of fuel} \quad (6)$$

3. Results and Discussions

3.1. Safety characteristics

The world is currently dealing with pollution and climate change. Also, there are a real need for a solution of a sustainable developments in industrial sectors. Pollution levels are skyrocketing as the automobile industry expands daily [14]. In this context, fluids that may be suitable for vehicle air conditioners have been examined in this study. The refrigerants physical, safety and environmental data analyzed in our study are given in Table 1 [21]. It has been seen in this study that R-1233ze(Z) refrigerant is the best refrigerant compared to other refrigerants according to the performance data of these refrigerants. In addition, environmental and safety data values are also good compared to other refrigerants, but due to the flammability of this refrigerant, necessary safety measures are needed. The same is true for R-1234ze(E) and R-1234yf refrigerant. Although the ODP value of the R-E245cb2 refrigerant is zero, its GWP value is high. Although the ODP values of R-404A, R-134A and R-513A refrigerants are 0, they have high GWP values, which is the biggest disadvantage of these refrigerants.

3.2. Effect of evaporator temperature on performance parameters

For refrigerants selected as R-134A, R-E245cb2, R-404A, R-1234ze(Z), R-161, R-1234zd(E), R-513A, R-1234ze(E) and R-1234yf, detailed analyses have been made according to the first and second laws of thermodynamics for different evaporator and condenser temperatures. Analyses results according to different evaporator temperatures are given in Figure 2-11. The operating conditions accepted for the analyses are T_{con}= 313 K, and the refrigeration capacity is 26 kW. Figure 2 shows that the COP values increase as the evaporator temperatures increase. Among the analyzed refrigerants, the

highest COP value belongs to R-1234ze(Z) refrigerant, followed by R-1234zd(E) and R-161 refrigerants, respectively. It has been observed that the lowest COP value belonged to the R-404A refrigerant.

Table 1. *Physical, environmental and safety data of analyzed refrigerants.*

Refrig.	Mol. weight (g/mol)	Critical Temp. (ºC)	ASHRAE 34 Safety	ODP	GWP
R-134A	102.03	101.10	A1	0	1430
R-161	48.06	102.2	A1	0	12
R-404A	97.60	72	A1	0	3900
R-513A	108.40	96.50	A1	0	573
R-1234ze(E)	114.04	-19.12	A2	0	4
R-1234yf	114	94.70	A2	0	4
R-1233zd(E)	130.50	154	A1	0	7
R-1234ze(Z)	114.04	150.1	A2	0	1
R-E245cb2	150.04	133.663	-	0	680

In addition, R-1234ze(Z) is classified as A2, which has some flammability while the component R-1234zd(E) and R-161 have a flame-retardant effect. Thus, R-1234zd(E) and R-1234ze(Z) refrigerants can be recommended as an alternative to R134a in vehicles.

As the evaporator temperatures increase, the compressor work required for the cooling system decreases (Figure 3). In Figure 3, it is also seen that the highest compressor work belongs to the R-404A refrigerant, and the lowest compressor work belongs to the R-1234ze(Z) refrigerant at increasing evaporator temperatures. Figure 4 shows the variation of the mass flow rate ratio with increasing evaporator temperature. In all refrigerants, the mass flow rate decreases with increasing evaporator temperatures. The highest and lowest mass flow rate ratios consist at R-404A and R-161 refrigerants. In addition, Figure 5 shows the heat values released from the condenser decreases with increasing evaporator temperatures. While the highest heat values thrown from the condenser belong to the R404A and R-1234yf refrigerant, it is seen that the lowest value belongs to the R-1234ze(Z) refrigerant (Figure 5). In Figure 6, the second law efficiency values of the analyzed refrigerants are given for different evaporator temperatures. The exergy efficiency decreases with increasing evaporator temperatures. It has been observed that the highest second law efficiency value belonged to the R-1234ze(Z) refrigerant, while the lowest efficiency value obtained at the R-404A refrigerant. It is also seen from the figure that the second law efficiency of the R-1234zd(E) refrigerant is 4.68% lower than

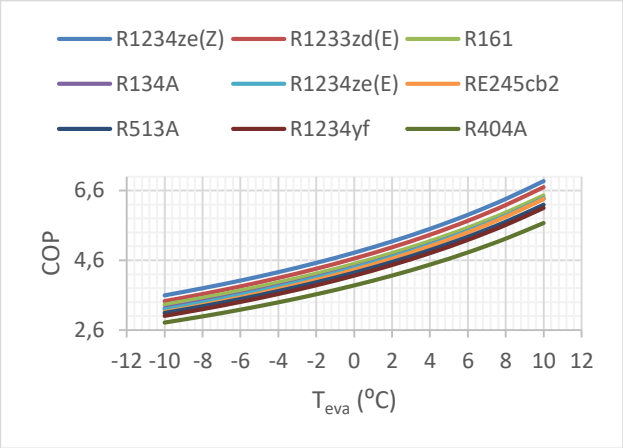


Figure 2. *Variation of the COP with evaporator temperatures of the refrigerants.*

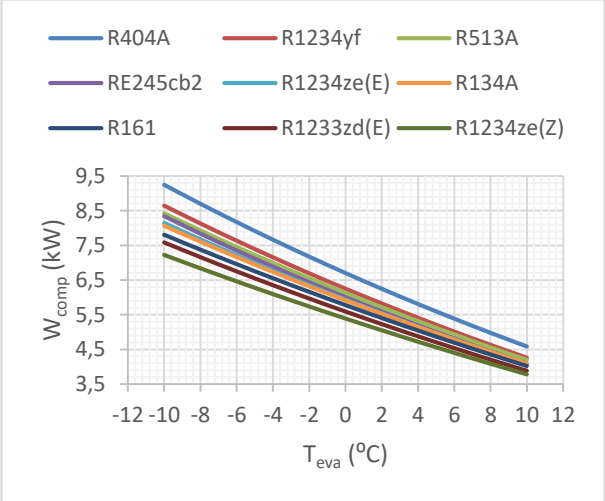


Figure 3. *Variation of the W_{comp} with the evaporator temperatures of the refrigerants*

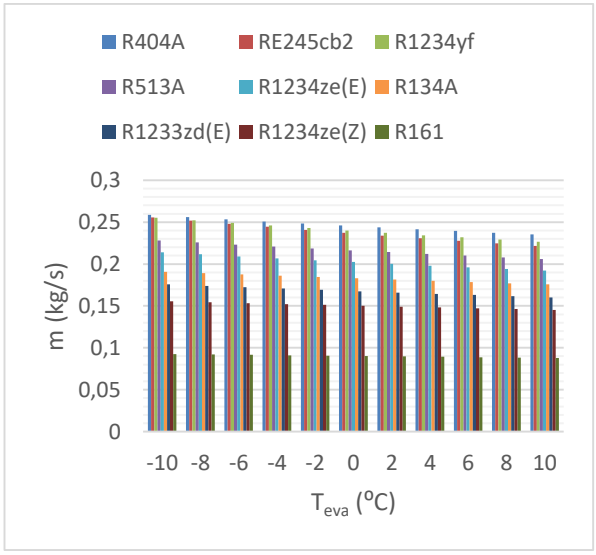


Figure 4. *Variation of the mass flow rate ratio with the evaporator temperatures of the refrigerants.*

the R-1234ze(Z) refrigerant. The total exergy destruction value of the system in the analysis performed at different evaporator temperatures is given in Figure 7. The highest exergy destruction

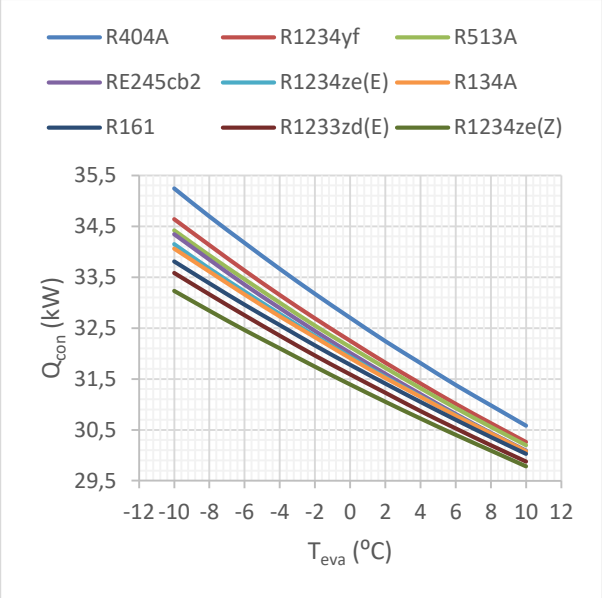


Figure 5. *Variation of the Q_{con} with the evaporator temperatures of* the *refrigerants.*

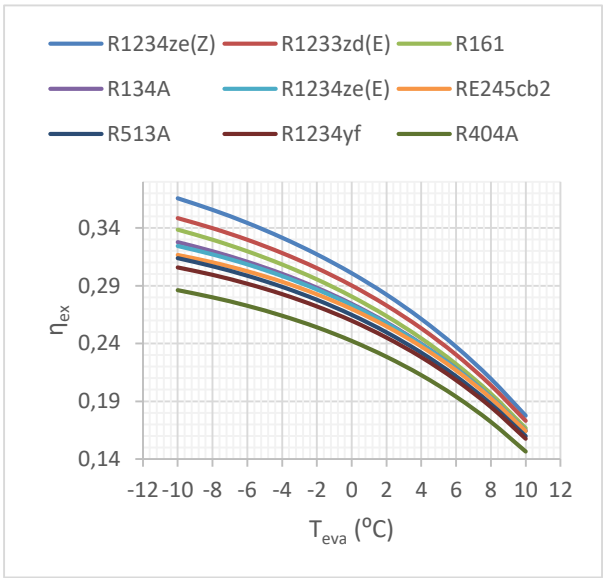


Figure 6. *Variation of the η_{ex} with the evaporator temperatures of the refrigerants.*

values obtained from R-404A, the lowest values were obtained for R-1234ze(Z). In the analysis, the total exergy destruction value of each system element performed at different evaporator temperatures is shown in from Figure 8 to Figure 11. In all refrigerants, exergy destruction decreases with increasing evaporator temperatures (Figure 8). The highest exergy destruction values belong to R-404A, followed by R-1234zd(E), R-1234yf and R-161 in the evaporator. The lowest values have been obtained for R-1234ze(Z) (Figure 8). In Figure 9, similarly, the exergy destruction decreases with increasing evaporator temperatures for all the refrigerants in the condenser.

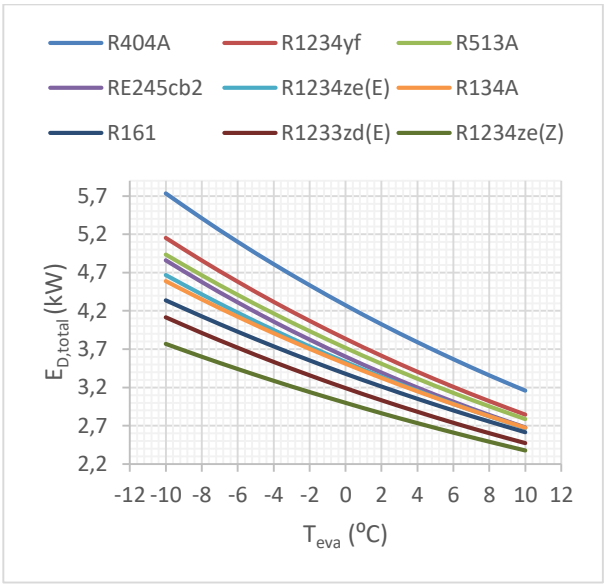


Figure 7. *Variation of the $E_{D,total}$ with the evaporator temperatures of the refrigerants.*

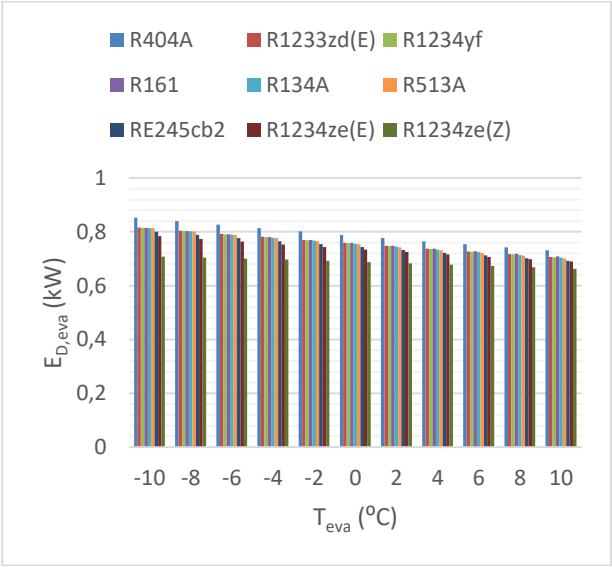


Figure 8. *Variation of the $E_{D,eva}$ with the evaporator temperatures of the refrigerants.*

Figure 9 illustrates that the highest exergy and the lowest destruction values in the condenser belong to R-161 and R-1234zd(E), respectively. In addition, it is seen from the Figure 9 that the highest exergy destruction values belong to the refrigerant R-404A, 1234ze(Z) and R-134A. In Figure 10, it is seen that the highest exergy destruction values in the compressor belong to the R-404A refrigerant. On the other hand, the lowest exergy destruction values were obtained for R-1233ze(Z). It shows that the highest and the lowest exergy destruction values in the throttling valve have been obtained in R-404A and R-1233ze(Z), respectively (Figure 11). In all the refrigerants, the exergy destruction decreases with the increasing evaporator temperatures.

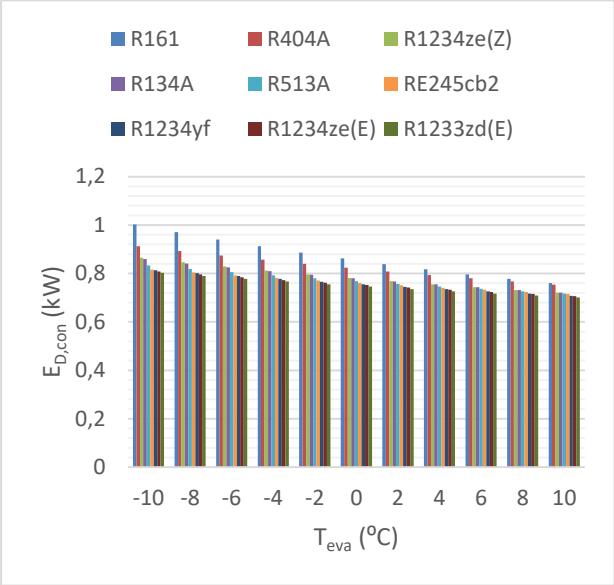


Figure 9. *Variation of the $E_{D,con}$ with the evaporator temperatures of the refrigerants.*

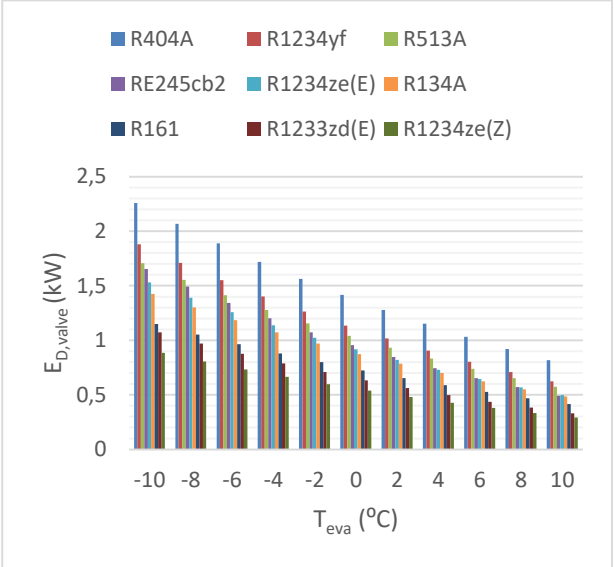


Figure 11. *Variation of the $E_{D,valve}$ with the evaporator temperatures of the refrigerants.*

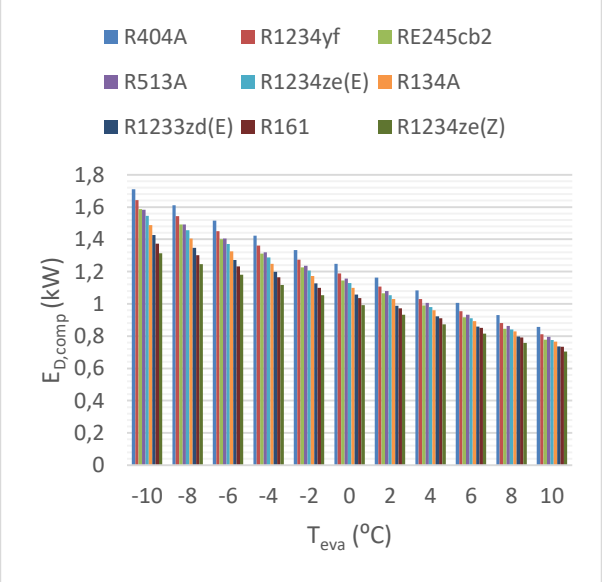


Figure 10. *Variation of the $E_{D,comp}$ with the evaporator temperatures of the refrigerants.*

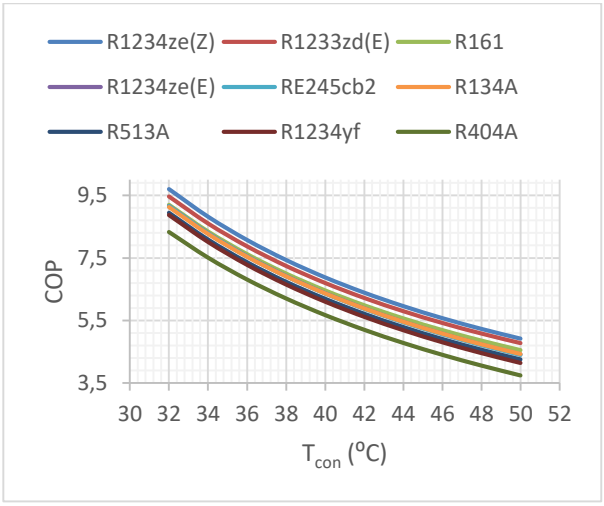


Figure 12. *Variation of the COP with the condenser temperatures of the refrigerants.*

3.3. Effect of condenser temperature on performance parameters

Analysis results according to different evaporator temperatures are given from Figure 12 to Figure 21. The operating conditions accepted for the analysis were T_{eva}= 283 K, and the refrigeration capacity was 26 kW. It is seen this the COP values decrease as the condenser temperatures increase (Figure 13). While the maximum COP value was obtained from the R-1234ze(Z) refrigerant among the analyzed refrigerants, it has seen that the lowest COP value was obtained from R-404A refrigerant. In Figure 13, the highest compressor work at increasing the condenser temperatures belongs to R-404A refrigerant, followed by R-1234yf and R-513A refrigerants, respectively. It is seen that the lowest compressor work belongs to the R-1234ze(Z) refrigerant under the same operating conditions. Figure 14 shows the variation of the mass flow rate ratio with increasing condenser temperature. The highest and lowest mass flow rates occur when using R-404A and R-161. As the condenser temperature increases, the mass flow rates of the refrigerants also increase. In Figure 15, it is seen that the heat values released from the condenser increase according to the increasing condenser temperatures for all the refrigerants. It is seen that the highest heat values from the condenser belong to the R-404A, R-1234yf and R-513A refrigerants, respectively. The exergy efficiency of the refrigerants examined in the analysis decreases with increasing the condenser temperature.

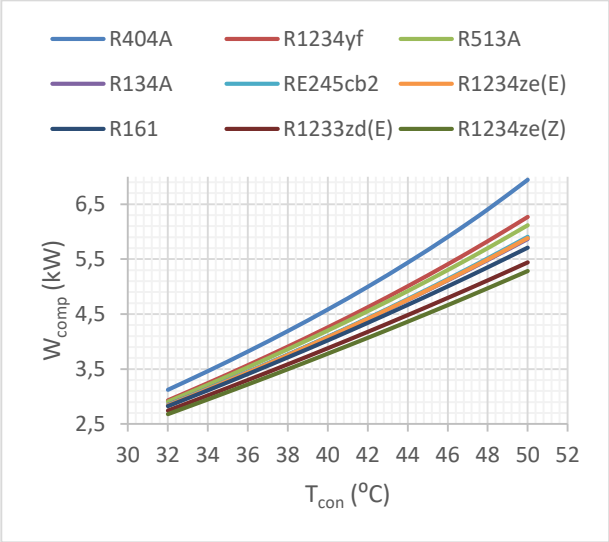


Figure 13. *Variation of the W_{comp} with the evaporator temperatures of the refrigerants.*

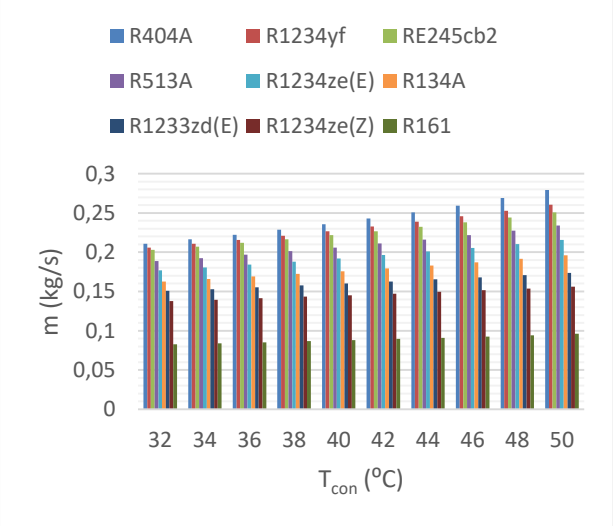


Figure 14. *Variation of the mass flow rate ratio with the condenser temperatures of the refrigerants.*

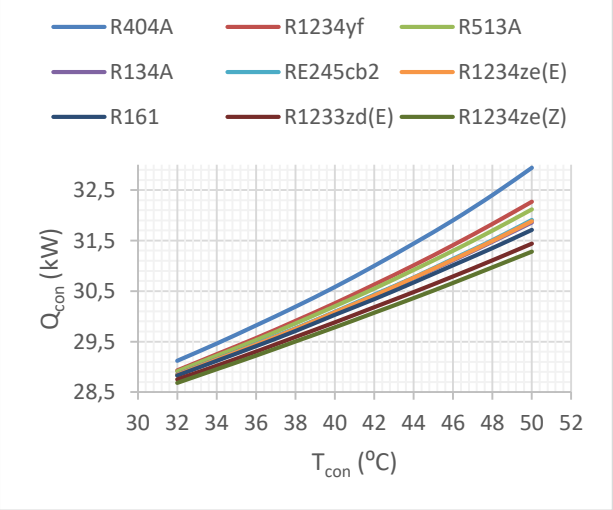


Figure 15. *Variation of the Q_{con} with the condenser temperatures of the refrigerants.*

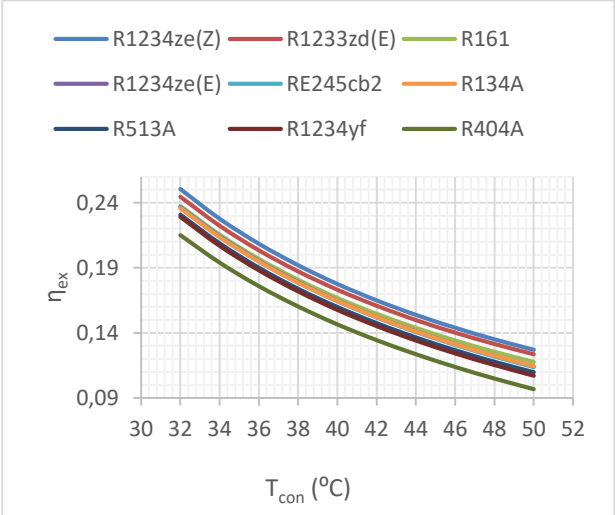


Figure 16. *Variation of the η_{ex} with the condenser temperatures of the refrigerants.*

In Figure 16, it is seen that the highest second law efficiency value belongs to R-1234ze(Z) refrigerant, while the lowest efficiency value belongs to R-404A refrigerant at increasing the condenser temperatures. Total exergy destruction is very important in exergy analysis as it provides information about the energy quality degradation. The total exergy destruction value of the system in the analysis performed at different condenser temperatures is shown in Figure 17. As the condenser temperature increases, the total exergy destruction also increases. In Figure 17, the highest exergy destruction values at increasing condenser temperatures belong to R-404A, R-1234yf and R-513A refrigerants, respectively. It is seen that the lowest exergy destruction belongs to the R-1234ze(Z) refrigerant. The total exergy destruction value of each components performed at different condenser

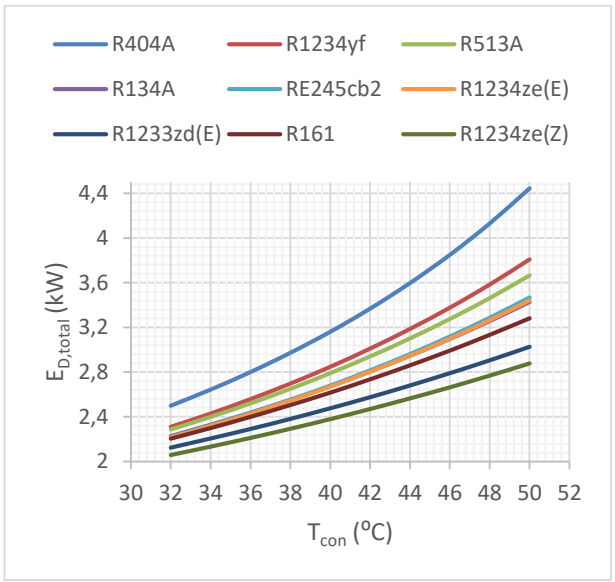


Figure 17. *Variation of the $E_{D,total}$ with the condenser temperatures of the refrigerants*

temperatures is shown from Figure 18 to Figure 21. Figure 18 shows that the highest exergy destruction values belong to R-404A, followed by R-161 and R-1234zd(E) in the evaporator. The lowest values have been obtained for R-1234ze(Z). The highest exergy and the lowest destruction values in the condenser belong to R-404A and R-1234zd(E), respectively (Figure 19). As the condenser temperature increases, the exergy destruction values in the

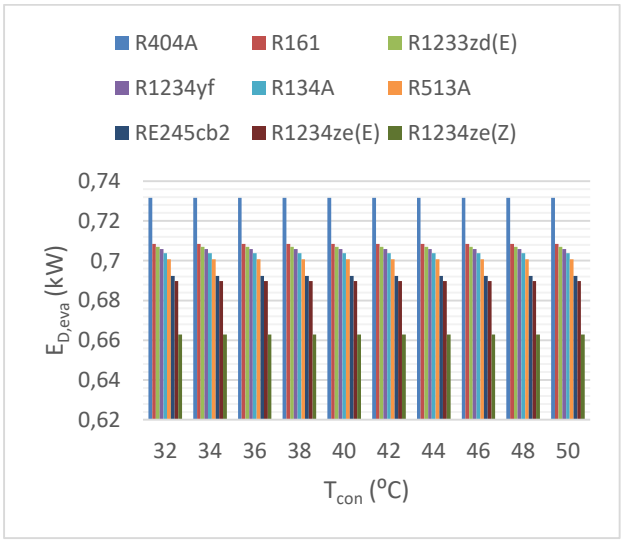


Figure 18. *Variation of the $E_{D,eva}$ with the condenser temperatures of the refrigerants.*

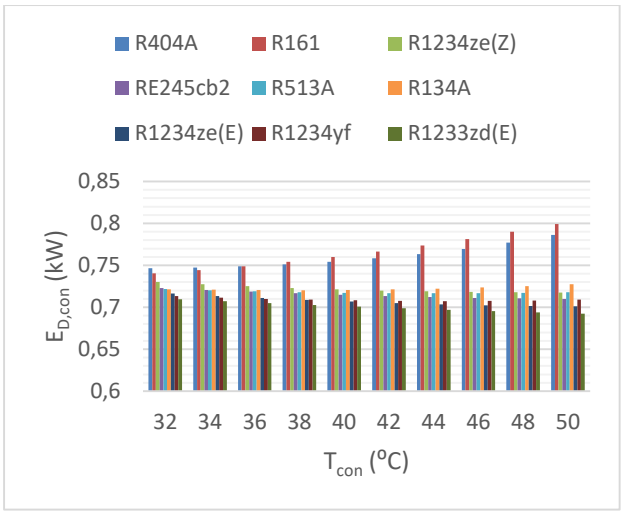


Figure 19. *Variation of the $E_{D,con}$ with the condenser temperatures of the refrigerants.*

compressor also increase. (Figure 20). In Figure 20, it is seen that the highest exergy destruction values in the compressor belong to the R-404A refrigerant. The other highest exergy destruction values belong to R-1234yf and R-513A refrigerants, respectively. On the other hand, the lowest exergy destruction values have been obtained for R-1233ze(Z). It shows that the highest and the lowest exergy destruction values in the hrottling valve were obtained in R-404A and R-1233ze(Z), respectively (Figure 21).

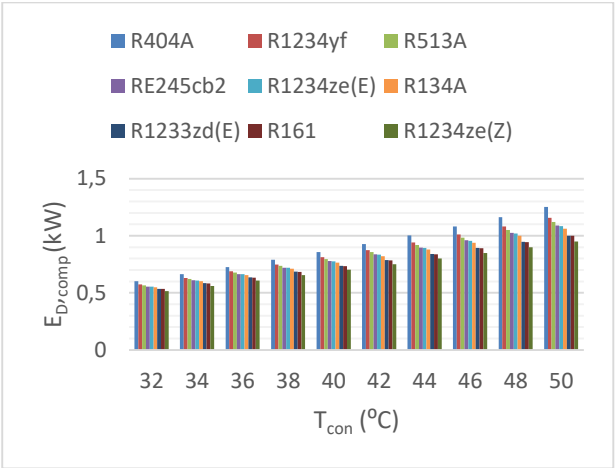


Figure 20. *Variation of the $E_{D,comp}$ with the condenser temperatures of the refrigerants.*

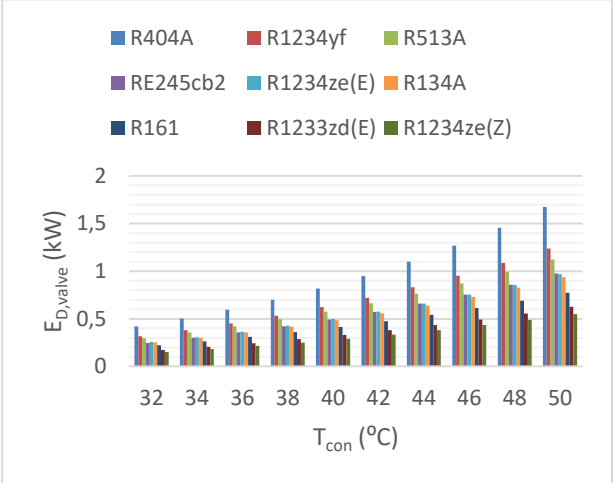


Figure 21. *Variation of the $E_{D,valve}$ with the condenser temperatures of the refrigerants.*

4. Conclusions

The air conditioning system used in vehicles needs energy to operate. The energy source used in vehicles is fuel. Chemical energy is converted into mechanical work by burning fuel in the vehicle engine. Again, the alternator in the vehicle generates electrical energy with the mechanical drive it receives from the engine. All systems and elements in the vehicle consume mechanical or electrical energy during their operation. The place with the highest energy consumption in the air conditioning system is the compressors. Power consumption values must be determined in order to determine the net effect of the compressor in the air conditioning system on the engine. The refrigerants used in the air conditioning system also have a decisive effect on the efficient and effective operation of the air conditioning system. In this study, detailed thermodynamic analyses of different refrigerants has been carried out in order to increase the performance of an air conditioning system operating according to the vapor compression refrigeration cycle in

vehicles. For refrigerants selected as R-134A, R-E245cb2, R-404A, R-1234ze(Z), R-161, R-1234zd(E), R-513A, R-1234ze(E) and R-1234yf, detailed analyses have been made according to the first and the second laws of thermodynamics for different evaporator and condenser temperatures. In the study, the comparison has been made by taking account the safety, the environment and the performance values of the refrigerants.

In the analysis it was understood that for different evaporator temperatures, the highest COP value belongs to the R-1234ze(Z) refrigerant, followed by R-1234zd(E) and R-161 refrigerants, respectively. It has been observed that the lowest COP value belonged to the R-404A refrigerant. It is seen that the COP values decrease as the condenser temperatures increase. Among the analyzed refrigerants, the highest COP value belonged to R-1234ze(Z) refrigerant, while the lowest COP value belonged to R-404A refrigerant.

In the study, the second law efficiency values of the analyzed refrigerants have been calculated for different evaporator temperatures. It has been observed that the highest second efficiency value belonged to the R-1234ze(Z) refrigerant, while the lowest efficiency value belonged to the R-404A refrigerant. The exergy efficiency of all refrigerants decreases with the increasing of the condenser temperature. At increasing the condenser temperatures, the highest second law efficiency value belongs to R-1234ze(Z) refrigerant, while the lowest efficiency value belongs to R-404A refrigerant.

Total exergy destruction is very important in exergy analysis as it provides information about the energy quality degradation. The aim of the detailed exergy analyses in this study is to improve the cycle performance by finding exactly where the actual losses occur and how these losses can be reduced. In the analyses made for different evaporator temperatures, the highest exergy destruction values have been obtained for R-404A, while the lowest values have been obtained for R-1234ze(Z) according to the total exergy destruction values of the system. The highest exergy destruction values at increasing the condenser temperatures belong to R-404A, R-1234yf and R-513A refrigerants, respectively. It is seen that the lowest exergy destruction belongs to the R-1234ze(Z) refrigerant.

Among these refrigerants, R-1233ze(Z) has been found to be the best refrigerant compared to other refrigerants according to the environmental and the safety data. However, due to the flammability of these refrigerants, the necessary safety precautions are needed. The same is true for the R-1234ze(E) and the R-1234yf refrigerant. Although the ODP value of the R-E245cb2 refrigerant is zero, its GWP value is high. Although the ODP values of R-404A, R-134A and R-513A refrigerants are zero, they have high GWP values, which is the biggest disadvantage of these refrigerants.

The compressor, which is one of the main components of the air conditioning system, is driven by a pulley connected to the engine, which puts an additional load on the engine and therefore increases fuel consumption. The selection of the refrigerants with good performance, environmental and safety features in the vapor compression cooling system will make vehicles even more advantageous. As a result of the increasing the efficiency of the air conditioners, the size of the components used in the system can be reduced and the fuel consumption can be decreased.

Author Statements:

- The authors declare that they have equal right on this paper.
- The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- The authors declare that they have nobody or no-company to acknowledge.

Nomenclature

BACS	Bus Air Conditioning System
COP	Coefficient of performance
e	specific exergy (kJ kg^{-1})
E	exergy flow rate (kW)
h	enthalpy (kJ kg^{-1})
$\dot{m}$	mass flow rate (kg s^{-1})
P	pressure (kPa)
$\dot{Q}$	heat flow rate (kW)
s	specific entropy (kJ kg-1 K-1)
T	temperature (K or ^{o}C)
$\dot{W}$	compressor work (kW)

Greek Letters

η	*e*fficiency

Subscripts

com	compressor
con	condenser
eva	evaporator
D	exergy destruction
in	inlet
o	ambient
total	total system

References

[1] Karaali, R., (2016). Thermodynamic Analysis of a Cascade Refrigeration System.' *Acta Phys. Pol. A* 130;101-106. DOI: 10.12693/APhysPolA.130.101

[2] Karaali, R., (2016). Exergy Analysis of a Combined Power and Cooling Cycle. 130 *Acta Phys. Pol. A* 130(1);209-213. DOI: 10.12693/APhysPolA.130.209

[3] Musa Atasbak, Arzu Keven, and Rabi Karaali, (2022). Exergy analyses of two and three stage cryogenic cycles. *Applied Rheology*. 32:190–204. https://doi.org/10.1515/arh-2022-0134

[4] Ozgoren, M., Kahraman, Ali, Solmaz, O., (2011). Determination of Hourly Performance of Automobile Air Conditioner Using Different Refrigerants for İzmir Province. *X. National Plumbing Engineering Congress*, İzmir.

[5] Zhang, N.Y., (2022). Performance evaluation of alternative refrigerants for R134a in automotive air conditioning system. *Asia-Pac J Chem Eng*.; 17: e2732.

[6] Aized, T., Hamza, A., (2019). Thermodynamic Analysis of Various Refrigerants for Automotive Air Conditioning System. *Arabian Journal for Science and Engineering*, 44:1697–1707. https://doi.org/10.1007/s13369-018-3646-8

[7] Papasavva, S., Moomaw, W., (2014). Comparison between HFC-134a and Alternative Refrigerants in Mobile Air Conditioners using the GREEN-MAC-LCCP© Model. *15th International Refrigeration and Air Conditioning Conference at Purdue*, July 14-17.

[8] Mohanraj, M., Andrew Pon Abraham, J. D., (2022). Environment friendly refrigerant options for automobile air conditioners: a review. *Journal of Thermal Analysis and Calorimetry*,147:47–72. https://doi.org/10.1007/s10973-020-10286-w

[9] Poongavanam, G., Sivalingam, V., Prabakaran, R., Salman, M., Kim, S. C., (2021). Selection of the best refrigerant for replacing R134a in automobile air conditioning system using different MCDM methods: A comparative study. *Case Studies in Thermal Engineering*, 27;101344. https://doi.org/10.1016/j.csite.2021.101344

[10] Zhang, N., Dai, Feng, Y Li, B., (2022). Study on environmentally friendly refrigerant R13I1/R152a as an alternative for R134a in automotive air conditioning system. *Chinese Journal of Chemical Engineering*, 44: 292–299. https://doi.org/10.1016/j.cjche.2021.02.028

[11] Meng, Z., LIU, Y., Wang, D., GAO, L., Yan, J., (2021). Refrigerating Fluid with a Low Global Warming Potential For Automotive Air Conditioning Systems in Summer. *Thermal Science*, 25(2B);1443-1451. DOI:10.2298/TSCI200308045M

[12] Unal, S. (2017). An Experimental Study on a Bus Air Conditioner to Determine its Conformity to Design and Comfort Conditions. *Journal of Thermal Engineering*, 3(1);1089-1101. https://doi.org/10.18186/thermal.277288

[13] Suh, I., Lee, M., Kim, J., Uh, S., Won, J. (2015). Design and experimental analysis of an efficient HVAC (heating, ventilation, air-conditioning) system on an electric bus with dynamic on-road wireless charging. *Energy*, 81;262-273. https://doi.org/10.1016/j.energy.2014.12.038

[14] Sagar Vashisht, S., Rakshit, D. (2021). Recent advances and sustainable solutions in automobile air conditioning systems. *Journal of Cleaner Production*, 329;129754. https://doi.org/10.1016/j.jclepro.2021.129754

[15] Yan, M., He, H., Jia, H., Li, M., Xue, X., (2017). Model predictive control of the air-conditioning system for electric bus. *Energy Procedia*, 105;2415 – 2421.

[16] Andrew, J.D., Abraham, P., Mohanraj, M., (2019). Thermodynamic performance of automobile air conditioners working with R430A as a drop-in substitute to R134a. *Journal of Thermal Analysis and Calorimetry*, 136:2071–2086. DOI:10.1007/s10973-018-7843-1

[17] Unal, Saban. (2015). Determination of the ejector dimensions of a bus air-conditioning system using analytical and numerical methods. *Applied Thermal Engineering*, 90;110-119. https://doi.org/10.1016/j.applthermaleng.2015.06.090

[18] Janotkova, E., Pavelek, M., (2006). News Trend in the field of Automobil Air Conditioning. https://www.aiha.org/aihce06/handouts/janotkova.pdf (Created 2006.05.03).

[19] Bilgili, M., Ediz Cardak, E., Aktas, A.E. (2017). Thermodynamic Analysis of Bus Air Conditioner Working with Refrigerant R600a. *European Mechanical Science*, 1(2): 69-75. https://doi.org/10.26701/ems.321874

[20] Bejan, A., Tsatsaronis, G., Moran, M., (1996). *Thermal design and optimization*, New York: John Wiley and Sons Inc.

[21] Calm J. M. and Hourahan G. C., (2007). Refrigerant Data Update, *HPAC Engineering.* 79;50-64.

International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 9-No.1 (2023) pp. 29-35
http://dergipark.org.tr/en/pub/ijcesen

ISSN: 2149-9144

Research Article

Investigation of the Parameters of Non-Cylindrical Ice Load on Power Transmission Lines

Ali AJDER*

Yildiz Technical University, Faculty of Electrical and Electronics, Department of Electrical Engineering, 34220, Istanbul-Turkey

* **Corresponding Author**: **Email:** aliajder@yildiz.edu.tr - **ORCID:** 0000-0001-9411-4452

Article Info:

DOI: 10.22399/ijcesen.1260707
Received: 06 March 2023
Accepted: 15 March 2023

Keywords

AutoCAD
Ice load
Ground wire
Non-cylindrical ice,
Phase conductor

Abstract:

Ice load on transmission lines is a critical factor that affects their cost and operation. National standards specify how ice load is considered in the design of power lines and poles. These standards generally use empirical relations that assume that the ice load on each phase accumulates uniformly and cylindrically. However, field tests and fault records show that the actual ice load on conductors is often not cylindrical due to altitude, wind strength and direction, and terrain topography. This study firstly defines several parameters to describe asymmetrical ice load. This load can cause additional vertical force on the line, conductor swing angle deviation, and sag changes. Since empirical equations are only valid for cylindrical ice load, the cross-sectional shape of the conductor must be transferred to millimeter paper, and calculations performed using one of several numerical integral methods. The coefficients for asymmetric ice are calculated in kg/m (N/m) using an AutoCAD model in the numerical study.

1. Introduction

National standards and regulations define the design principles of power transmission lines. For example, DIN VDE 0210 (Germany), ÖVE-L11 (Austria), LeV (Switzerland), NBR 5422 (Brazil), GB50545 (China), NESC (USA), KTP 18 (Albania) and Electricity Network Regulation, Electricity High Current Facilities Regulation (Turkey) are some of these national standards/regulations. For each country, different criteria are determined according to transmission line designs, considering the existing conditions [1]. The IEC 80626 is the international standard for the design of transmission lines [2]. Ice load is one of the most important parameters considered in mechanical calculations of power transmission lines [3]. The ice load accreted on the phase conductors and the protection wire causes forces in the vertical direction, leading to the selection of large-scale cross members, diagonals, and vertical brackets and increasing the cost of poles. Even if the transmission line design is well designed, the extra ice loads over time can cause conductors to break and cause energy loss [4]. Even if the line does not rupture, when the wind force acts on the conductor while it is ice-loaded, the conductor's oscillation angle changes, resulting in unexpected problems [5,6].

Ice load (ice cover) accreted on conductors can be formed in different structures such as hoarfrost, crystalline hoarfrost, frost, crystalline ice, and snow load formed as a result of the adhesion of wet snow on the conductor during the sudden change of weather. Another critical parameter for ice load on transmission lines is the icing speed. According to TS IEC 60826 Standard, the water content in the air, wind speed, the average volume of ice particles, ambient temperature, and dimensions of the icy object are defined as factors affecting the icing speed.

Below -10 °C, the water composition in the air decreases, so icing of overhead lines is generally not observed. However, an ice load of 8 kg/m has been recorded in Switzerland even at air temperatures below -20 °C (with strong winds). TS IEC 60826 Standard recommends that the ice load be reflected in the design of the lines with the help of measurements, statistical data on wind and air temperature, and mathematical models.

The ice load should be obtained from measurements taken from conductors representing the line [7]. These measurement methods are defined in IEC

61774 [8]. Ice accretion models from different studies can help with ice data measurement, but verification with actual results is required [9]. The terrain effect is critical for ice accumulation calculations [10]. Since the terrain structure dramatically affects the icing mechanism, transferring data from one area to another causes inaccurate measurements. As a result, icing data obtained from metering stations like or near the line route are required for transmission line designs [2].

TS IEC/TR2 61774 describes icing models in the literature. These models represent model characteristics such as the type of climatic data required and the form of predicted accreted ice.

Accordingly, in the transparent ice model, the thickness of ice accumulation on power transmission lines is estimated using a correction factor based on the diameter of the icy conductor updated at regular intervals of 1 hour. This correction factor is inspired by another study on wind tunnels [11].

In the hoarfrost model, the icing intensity on an object is estimated based on factors such as wind speed, precipitation rate, and temperature. The ice load is calculated by considering the effects of changing the diameter of ice accumulation [12].

In the sleet model, only sleet accumulation is predicted based on factors such as precipitation, snowfall, wind speed, and air temperature. The model uses observed data to determine the average amount of snow thickness and density. Snowmelt is also considered due to factors such as dispersion and reflection.

In the clear ice, hoarfrost, and sleet model, all three types of icing are predicted based on climatic data and the surface characteristics on which the ice is accreted. Ice accretion rate, accretion efficiency, and other factors are calculated to determine the mass and thickness of ice on the conductors. The model updates the conductor diameter and surface properties at each time step to simulate ice formation on transmission lines [13].

The ice and sleet model was developed for simulation studies on power transmission lines. Three-dimensional, time-dependent mathematical models for predicting cylindrical elongation and axial expansion consider the effect of ice/snow load and aerodynamic moment on the rotation and torsion of the conductor, as well as the position of the accretion area along the wind direction [14].

The geometry of ice loads, which are highly effective at high altitudes above sea level with extreme icing and simultaneous storms, can deviate from the cylindrical shape of the field. This study summarizes the calculation of cylindrical ice loads and provides practical definitions and mathematical relations for non-cylindrical ice loads.

2. Material and Methods

2.1 Calculation of Ice Load on Power Transmission Lines

There is no analytical relation for calculating ice load on transmission lines; generally, empirical relations have been developed as a function of conductor diameter $d\ (mm)$ using historical operating information and meteorological and topographical data in countries.

According to EN 50341-3 Standard in European countries, ice load $g_i\ (N/m)$ is formulated as a function of conductor diameter $d\ (mm)$ under some conditions as in other countries. In some cases, it is defined directly. In all these relations, ice is assumed to be cylindrical, symmetrical, and uniformly collected in phase conductors. According to this assumption, the weight of the ice load in the conductor $g_i\ (kg/m)$, the ice diameter $d_i\ (mm)$, and the ice density $\rho_i\ (kg/dm^3)$ are related in Equation 1:

$$d_i = \sqrt{[d^2 + 1274 \times g_i \times \rho_i^{-1}\,]} \qquad (1)$$

The $icing\ rate\ (kg/m/h)$ is the rate at which ice (kg/m) accretes on the conductor for 1 hour. There are empirical relations between $icing\ rate$ and amount of precipitation $y\ (mm/h)$, air temperature t (°C), and wind speed $v\ (km/h)$ given in Equation 2-5 [15].

$$icing\ rate = 0.00061 + 0.00245y \qquad (2)$$

$$icing\ rate = e^{at} - e^{[b(t+c)]} + d \quad \{a, b, c, d: empirical\ coefficients\} \qquad (3)$$

$$icing\ rate = 0.01 - 0.00208v \quad \{parallel\ wind\} \qquad (4)$$

$$icing\ rate = -0.005 + 0.00095v \quad \{steep\ wind\} \qquad (5)$$

If the ice loads $g_{i_1}\ (kg/m)$ and $g_{i_2}\ (kg/m)$ of a conductor at T_1 and T_2 are known, the icing rate at time $T = T_1 - T_2\ (h)$ is calculated by Equation 6.

$$icing\ rate = \frac{g_{i_2} - g_{i_1}}{T_2 - T_1} \quad (kg/m/h) \qquad (6)$$

In Equation 6, minus sign (-) for the icing rate at period T indicates that the ice is in the process of melting. Figure 1 shows a perspective view of the ice load accumulated on the phase conductor. In addition to the accumulation of ice load on the phase conductors, it can also accumulate on the protection wires. These different scenarios should be considered separately in ice load analysis.

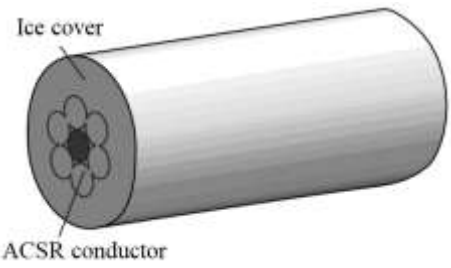


Figure 1. *3D model of cylindrical ice cover in a phase conductor*

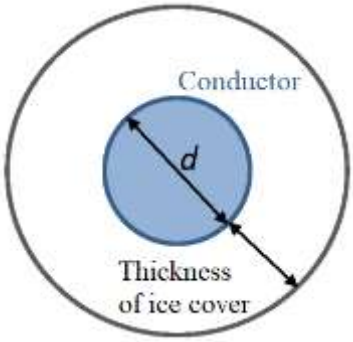


Figure 2. *Thickness of cylindrical icing*

According to TEIAS Project Technical Specifications, the thickness of ice on a conductor δ_b is accepted as 0 for Zone I, 15 (mm) for Zone II, 20 (mm) for Zone III, and 30 (mm) for Zone IV. Figure 2 shows the case where cylindrical icing occurs. Equation 7 is used to calculate the ice thickness.

$$d_i = d + 2 \times ice\ thickness \tag{7}$$

2.2 Definition of Non-Cylindrical (Asymmetric) Ice Load in Power Transmission Lines

2.2.1 Partial Ice Load

While the accumulation of ice is generally considered uniform for the whole conductor, due to the nature of the terrain, the sun's rays incident on the icy conductor may be partially blocked by a mountain, trees, or any other obstacle. As a result, part of the ice cover may melt and fall off, called the 'partial ice load' problem in the literature on power transmission lines.

Assuming uniform icing, the conductor only move in the vertical direction, whereas under partial ice load, there may be slippage in both vertical and horizontal directions. The possibility of partial ice loading needs to be examined in the design of relatively more critical transmission lines passing through areas with a high density of natural obstacles. With symmetrical pole span and uniform ice loading, the 'horizontal tangent point' occurs at the center of the pole span $(a/2)$, whereas with partial ice loading, the horizontal tangent point will shift towards the suspension point according to the moment effect [16].

2.2.2 Non-Cylindrical Ice Load

During the design of transmission lines, ice accretion is assumed to be 'cylindrical'; however, for unforeseen reasons, the ice cover on conductors or protection wires may only sometimes be cylindrical. In such cases, the sample of non-cylindrical ice cover is examined at ice load monitoring stations for analysis. The ice cover sample taken from the conductor is sliced, and cross-sectional photographs are taken for each slice. Perspectives of the conductor and ice cover are obtained according to the averages of the measurement values taken from different points of the x and y axes in the photographs taken. This information is used during the design of the ice cover.

The relation $g_i = k\sqrt{d}\ (kg/m)$ applies to cylindrical icing. Therefore, for asymmetric icing, the cross-sectional photograph of the ice is imported to millimeter paper, and the area of the ice cover is calculated using one of the numerical integration methods (Trapezoidal, Simpson, Durand, et al.) or AutoCAD modelling. The approximate weight (kg/m) is obtained from the sample's length and the density of ice.

Skewness

The skewness in terms of icing because of ice cover forming on a particular surface of the conductor while the other section of the same contour remains bare is defined as in Equation 9.

$$Skewness\ (\varepsilon) = \frac{arc\ length\ of\ bare\ conductor}{perimeter\ of\ conductor} \tag{9}$$

Degree of Deviation from Cylindrical Shape

The degree of deviation $(\boldsymbol{v})$ from the cylindrical shape can be approximated by Equation 10.

$$v \approx \left(\frac{d_{i_{avg}}}{d'_{i_{avg}}} - 1\right) \times 100 \quad (\%) \tag{10}$$

If the ice cover is cylindrical $\boldsymbol{v = 0}$ since the numerator and denominator will be equal in Equation 10.

Shape Coefficient

The average ice thickness $(\overline{\boldsymbol{\delta_i}})$, depending on the short and long diameters of the ice cover, can be obtained by Equation 11:

$$\overline{\delta_i} = \frac{1}{2}\sqrt{d_{i1}d_{i2}} - \frac{d}{2} \quad (mm) \tag{11}$$

In Equation 11, $\boldsymbol{d}\,(\boldsymbol{mm})$ is the diameter of the conductor. Equivalent cylindrical ice thickness can be calculated from Equation 12, where $\{area = area\ of\ ice + area\ of\ conductor\}$

$$\delta_i = \sqrt{\frac{area}{\pi}} - \frac{d}{2} \tag{12}$$

The shape coefficient $(\boldsymbol{\sigma})$, which indicates the amount by which the ice cover deviates from the cylindrical shape, is calculated by dividing Equation 12 by Equation 11.

Degree of Flatness

If there is an elliptical ice cover surrounding the conductor with $\boldsymbol{d_{b_X} > d_{b_Y}}$, the degree of flatness $(\boldsymbol{\alpha})$ can be approximated using Equation 13:

$$\alpha \approx \left(\frac{d_{i_X}}{d_{i_Y}} - 1\right) \times 100 \quad (\%) \tag{13}$$

Degree of Sharpness

If $\boldsymbol{d_{i_Y} > d_{i_X}}$ in the ice cover surrounding the conductor for the shape of an ellipse, then the degree of sharpness $(\boldsymbol{\beta})$ is defined, which is approximated in Equation 14.

$$\beta \approx \left(\frac{d_{b_Y}}{d_{b_X}} - 1\right) \times 100 \quad (\%) \tag{14}$$

Ovality of the Ice Cover

The ovality $(\boldsymbol{\gamma})$ can be calculated from the $\boldsymbol{d_{max}}$ and $\boldsymbol{d_{min}}$ measurements taken from the cross-section of the ice cover as in Equation 15:

$$\gamma = \left|\frac{d_{max} - d_{min}}{(d_{max} + d_{min})/2}\right| \times 100 \quad (\%) \tag{15}$$

If more precise calculations for $v, \sigma, \alpha, \beta, \gamma$ are desired, arithmetic averages can be used by taking different slices from the icy conductor sample.
Numerical integration methods can be used to calculate the average ice load, average wind-exposed lateral area, average degree of deviation of the ice cover from the cylindrical shape, average degree of flatness, and average ice flatness using a scaled image of the non-cylindrical ice cover along the span.

Unbalanced Icing

An unbalanced ice load can be defined as the exposure of phase or bundle conductors forming phases to different ice covers. The degree of unbalanced icing $\boldsymbol{\lambda_a}$ can be calculated for $phase\ a$ using Equation 16.

$$\lambda_a = \left|1 - \frac{g_{i_a}}{\bar{g}_i}\right| \times 100 \quad (\%) \tag{16}$$

In Equation 16, $\overline{\boldsymbol{g_b}}$ is the arithmetic average of the $phases\ a, b, c$ ($phases\ a, b, c$ a', b', c' if the pole is a double circuit). Calculating $\bar{\boldsymbol{g}}_{\boldsymbol{i}}$ with camera recordings taken from different locations on the relevant transmission line is possible.

For single circuit lines

$$\bar{g}_i = \frac{g_{i_a} + g_{i_b} + g_{i_c}}{3} \quad (kg/m) \tag{17}$$

For double circuit lines

$$\bar{g}_i = \frac{g_{i_a} + g_{i_b} + g_{i_c} + g_{i_{a'}} + g_{i_{b'}} + g_{i_{c'}}}{6} \quad (kg/m) \tag{18}$$

3. Results and Discussions

A single-circuit power transmission line with double protection wires is handled for the numerical study. The protection wires are galvanized steel conductors with a cross-section of 95 mm2. The phase conductors are steel-cored aluminum (ACSR) conductors with an outer diameter of 30 mm. Cardinal conductor with 954 MCM cross-section 54 aluminum 7 steel cores has a diameter of 30.42 mm, and Rail conductor with 954 MCM cross-section 45 aluminum 7 steel cores has a diameter of 29.61 mm; therefore, the conductor diameter is approximately 30 mm.
The area of the transmission line is in zone III of the Ice Load Map of Turkey (ice load coefficient, k = 0.3). The density of ice (ρb) is accepted as 0.6 kg/dm3 in the Regulation on Electrical Power Plants; however, ρb = 0.70 kg/dm3 is taken here by the 'Technical Specification for Pole Design for High Voltage Power Transmission Lines of TEIAS' [17]. Figure 3 shows an example of non-cylindrical icing on the phase conductor, and Figure 4 shows an example of non-cylindrical icing on the double protection wires. In Figure 3 (d), the horizontal scale is 1:2000, and the vertical scale is 1:1. Non-cylindrical icing examples in the phase conductor in Figure 3 (a), (b), (c), (d), and non-cylindrical icing examples in the protection wires in Figure 4 (a) and (b) are modelled in AutoCAD with all details. The results of the calculations using the definitions of non-cylindrical ice loads are given in Table 1 for each sample separately.

4. Conclusions

Among other extra loads on power transmission lines, the formation of non-cylindrical / asymmetric ice loads accelerates the aging of phase conductors and protection wires. Moreover, under the ice load, the angle of oscillation of single and bundle conductors changes with the effect of wind The conductor or protection wire exhibits an aerodynamically unstable movement in the span.

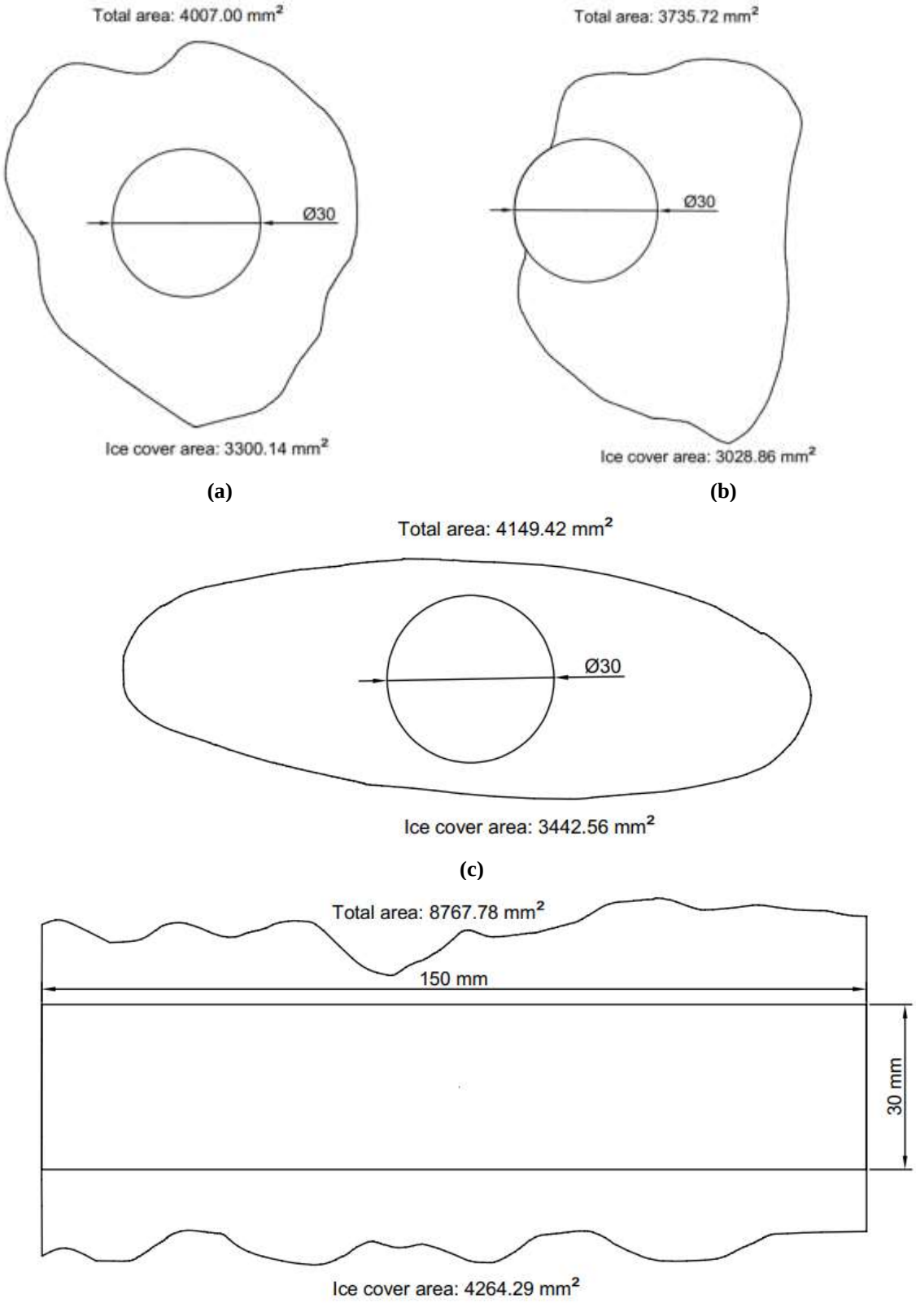
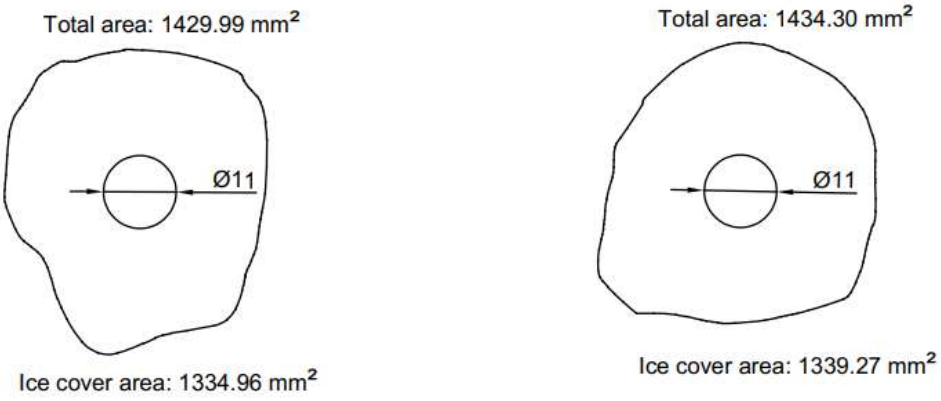


Figure 3 *Cross-sections of the non-cylindrical ice loads in the phases (a,b,c) and lateral field (d)*



(a) **(b)**

Figure 4 *Cross-sections of non-cylindrical ice loads in double protection wires (a, b)*

Table 1 *Values of the defined quantities of non-cylindrical icing*

Values	Figure 1 (a)	Figure 1 (b)	Figure 1 (c)	Figure 1 (d)	Figure 2 (a)	Figure 2 (b)
ε	-	0.26	-	-	-	-
v (%)	–12.7	–9.6	–14.2	-	3	2.8
σ	1.32	-	1.33	-	0.88	0.9
α (%)	-	-	194.4	-	-	-
β (%)	-	-	–66	-	-	-
γ (%)	16.4	-	98.6	-	10.8	4.3
λ_a (%)	-	-	-	-	0.16	-
λ_b (%)	-	-	-	-	-	0.16
F_b'/F_b	-	-	-	3.35	-	-

with asymmetrical icing. These unstable movements in phase conductors may cause phase-to-phase or phase-to-ground faults. The solution methods for such problems are beyond the scope of this study. Ice load on the transmission line increases the stress on the poles and conductors and weakens the foundations of the poles over time. Power transmission lines are the lifelines of power systems; it is crucial to maintain the functionality of power systems and prevent any possible damage by considering the social and economic aspects of electrical energy. In addition to the analysis and calculations, the power system's operation must continuously and reliably access information from the grid and continually monitor the transmission line. Power flow or voltage problems may indicate that there is a problem in the power lines due to ice load. For this purpose, fixed sensors installed on the transmission system or mobile sensors can be used. While a ground-up installation for the existing power grid is difficult and costly, different methods exist to detect snow load instantaneously. For example, helicopters or drones with the necessary equipment could be helpful. In this way, solutions that are fast, economical, and do not require any contact with the existing electricity system can be developed. However, the weather must be clear and not too windy to carry out such operations. A drone or UAV (Unmanned Aerial Vehicle) would be more cost-effective than a helicopter. UAVs equipped with laser scanners, cameras, and aviation security systems can be used for long-distance monitoring. Data obtained with advanced imaging techniques can be used for ice load analysis using image processing methods.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Fikke, S.M., (2008). *Modern meteorology and atmospheric icing*. Dordrecht: Springer.

[2] IEC 60826. (2003) *Design criteria of overhead transmission lines* IEC 60823:2003 (E) 3rd Ed., International Electrotechnical Commission.

[3] Farzaneh, M. (Ed.). (2008). *Atmospheric icing of power networks*. Canada: Springer.

[4] Cai, J., Liu, X., & Zhang, S. (2012). Numerical analysis for galloping of iced quad bundle conductors. *Applied Mechanics and Materials*, 226(228);30–34. DOI:10.4028/www.scientific.net/amm.226-228.30.

[5] Sopper, R., Daley, C., Colbourne, B., & Bruneau, S. (2017). The influence of water, snow and granular ice on ice failure processes, ice load magnitude and process pressure. *Cold Regions Science and Technology*, 139; 51–64. DOI:10.1016/j.coldregions.2017.04.006.

[6] Kermani, M., Farzaneh, M., & Kollar, L. E. (2013). The effects of wind induced conductor motion on accreted atmospheric ice. *IEEE Transactions on Power Delivery*, 28(2);540–548. DOI:10.1109/TPWRD.2013.2244922

[7] Mirshafiei, F., McClure, G., & Farzaneh, M. (2013). Modelling the dynamic response of iced transmission lines subjected to cable rupture and ice shedding. *IEEE Transactions on Power Delivery*, 28(2); 948–954. DOI:10.1109/TPWRD.2012.2233221

[8] IEC/TR2 61774. (1997) *Overhead lines - Meteorological data for assessing climatic loads*, International Electrotechnical Commission.

[9] Wen, Z., Yu, Q., Zhang, M., Xue, K., Chen, L., & Li, D. (2016). Stress and deformation characteristics of transmission tower foundations in permafrost regions along the Qinghai–Tibet Power Transmission Line. *Cold Regions Science and Technology*, 121; 214–225. DOI:10.1016/j.coldregions.2015.06.007

[10] Ma, G., Li, C., Jiang, J., Luo, Y. & Cheng, Y. (2012). A novel optical load cell used in icing monitoring on overhead transmission lines. *Cold Regions Science and Technology*, 71;67–72. DOI:10.1016/j.coldregions.2011.10.013

[11] Lozowski, E. P., Stallabrass J. R., & Hearty P. F. (1983). The icing of an unheated, nonrotating cylinder. Part I: A simulation model. *Journal of Applied Meteorology and Climatology*, 22(12); 2053-2062. DOI:10.1175/1520-0450(1983)022<2053:TIOAUN>2.0.CO;2

[12] Makkonen, L. (1984). Modeling of ice accretion on wires. *Journal of Climate and Applied Meteorology*, 23;929–939. DOI:10.1175/1520-0450(1984)023<0929:MOIAOW>2.0.CO;2

[13] Makkonen, L. (1998). Modeling power line icing in freezing precipitation. *Atmospheric Research*, 46(1–2);131–142. DOI:10.1016/S0169-8095(97)00056-2

[14] Poots, G., & Skelton, P.L. (1995). Simulation of wet-snow accretion by axial growth on a transmission line conductor. *Applied Mathematical Modelling*, 19(9);514-518. DOI:10.1016/0307-904X(95)00012-9

[15] Ay, S. (2018). *Enerji iletim sistemleri Cilt 4 Hava hatlarının mekanik hesaplamaları.* İstanbul: Birsen Yayınevi.

[16] Ajder, A. (2022). Analysis of non-uniform accreted ice in overhead power lines using SAP2000. *IEEE Access*, 10;128951–128958. DOI:10.1109/ACCESS.2022.3227648

[17] TEİAŞ, (2021). *Yüksek gerilim enerji iletim hatları için direk tasarımı direk tasarımı teknik şartnamesi*, İletim hatları tesis daire başkanlığı, Türkiye Elektrik İletim A.Ş., Haziran, 12–52.

International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 9-No.1 (2023) pp. 36-41
http://dergipark.org.tr/en/pub/ijcesen

ISSN: 2149-9144

Research Article

Diabetes Prediction Using Colab Notebook Based Machine Learning Methods

Önder YAKUT*

Kocaeli University, Faculty of Technology, Department of Information Systems Engineering 41001, Kocaeli-Turkey
* **Corresponding Author**: **Email:** onder.yakut@kocaeli.edu.tr - **ORCID:** 0000-0003-0265-7252

Article Info:

DOI: 10.22399/ijcesen.1185474
Received : 07 October 2022
Accepted : 16 March 2023

Keywords :

Cloud Computing
Diabetes Prediction
Google Colaboratory
Machine Learning

Abstract:

Diabetes is getting more and more common around the world. People suffer from diabetes or live at risk associated with this disease. It is necessary to prevent health problems caused by diabetes, to reduce the risk of diabetes and to reduce the load of diabetes on the health system. Therefore, it is important to diagnose and treat diabetic patients early. In this study, Pima Indian Diabetes (PID) database was used to predict diabetes. The PID database was divided into 2/3 for the training dataset and 1/3 for the test dataset. Then, the test and training datasets were fed into the machine learning models using five-fold cross-validation. Random Forest Classifier, Extra Tree Classifier and Gaussian Process Classifier machine learning methods were used to predict whether individuals have diabetes or not. In this study, the proposed method with the highest prediction accuracy was determined as the Random Forest Classifier. The proposed method's accuracy was 81.71%, precision was 88.79%, recall was 84.83%, F-score was 86.76% and ROC AUC was 88.03%. The proposed method was developed to assist clinicians in predicting the diagnosis of diabetic patients. The machine learning methods developed in this study were applied using Colab Notebook a Google Cloud Computing service.

1. Introduction

Insulin is a hormone that regulates blood sugar in the human body. Diabetes is a chronic disease that occurs when the pancreas cannot produce the necessary insulin, or the body cannot use the insulin produced effectively. Over time, diabetes causes serious damage to the cardiovascular system, eyes, kidneys, and nerves [1]. Diabetes does not only affect the individual who is sick. It is also a disease that affects the family of the sick individual and the whole of society. Care and treatment costs due to diabetes and the complications it causes increase rapidly and puts a load on the health system. In addition, the patient's quality of life decreases and this situation negatively affects the patient's family. Diabetes has become a global problem. Approximately 422 million people have diabetes, according to the World Health Organization. Most of these people live in low and middle-income countries. 1.6 million people die each year due to diabetes [2].

Predicting people with diabetes using machine learning methods will make the job of clinicians easier. Clinicians will ensure that people with diabetes are diagnosed and treated at an early stage. Thus, the load on the health system will be reduced and healthcare expenditures will be reduced. With the predetermination of diabetes, the disease will affect the lives of individuals less and increase the quality of life of individuals. Some studies in the literature that predict diabetes are given below. In these studies, the Pima Indians Diabetes database was used to predict diabetes. Febrian et al. used k-Nearest Neighbor and Naive Bayes algorithms comparatively to predict diabetes. In the proposed method, the Naive Bayes algorithm produced the best result for the PID data set [3]. Kibria et al. proposed a weighted voting classifier model to successfully predict diabetes risk. In the proposed method, ensemble learning was developed by combining Random Forest and XGBoost machine learning methods [4]. Chang et al. proposed an e-diagnosis system to detect and classify diabetes as an application of the Internet of Medical Things. They used Naive Bayes, Random Forest, and Decision Tree machine learning algorithms for classification in the proposed system [5]. Krishnamoorthi et al. Logistic Regression, k-Nearest Neighbor, Support Vector Machine and Random Forest machine learning algorithms were used for diabetes prediction. In the study, they proposed a smart

diabetes mellitus prediction framework [6]. Bhoi et al. proposed a model to predict diabetes in females in the PID dataset. In the proposed model, Classification Tree, Support Vector Machine, k-Nearest Neighbor, Naive Bayes, Random Forest, Neural Network, AdaBoost and Logistic Regression machine learning methods were used [7]. Guldogan et al. compared the prediction of the Multilayer Perceptron and Radial Based Function models to classify Type 2 Diabetes Mellitus [8]. Maulidah et al. proposed a method for diagnosing diabetes using the Naive Bayes algorithm and Particle Swarm Optimization technique utilization of the PID database [9]. Tigga et al. proposed the Random Forest Classifier method to predict Type 2 diabetes risk using the PID database [10]. Jakka et al. compared various machine learning methods to predict diabetic patients with high accuracy using the PID database [11]. Sisodia et al. used the Decision Tree, Support Vector Machine, and Naive Bayes machine learning classification algorithms to detect diabetes at an early stage [12]. Feng et al. proposed a variable-coded hierarchical fuzzy model for use in classification problems. The proposed method was used to diagnose diabetes. [13]. In this study, a machine learning method was proposed to be used in decision support systems for the early diagnosis of diabetes, which is becoming increasingly common. The proposed method has been thought to help patients start treatment early by identifying patients at risk for diabetes. The proposed method was developed to predict whether an individual had diabetes based on diagnostic criteria. For this purpose, training and test datasets were created using the patient data in the PID database. The PID database was divided into 2/3 for the training dataset and 1/3 for the test dataset. Thus, the developed machine learning models were fed with this training and test data set using five-fold cross-validation. The machine learning models developed in the study gained robust characteristics through rigorous training and testing. When the performance criteria obtained as a result of the experimental studies were evaluated, it is suggested that the proposed method has been used in computer-aided diagnosis systems to assist clinicians in decision-making.

The paper is structured as follows: material and methods are included in section 2. The experimental study is included in section 3. The experimental results are included in section 4. The discussion is included in section 5. The conclusions are included in section 6.

2. Material and Methods

In this section, the PID database used in the study is presented. Random Forest Classifier, Extra Tree Classifier and Gaussian Process Classifier machine learning methods used in the study are explained. In addition, it defined how the metrics that calculate the performance of the machine learning methods in the study are calculated.

2.1 Data Set

The data set used was obtained from the Pima Indians Diabetes (PID) Database of the National Institute of Diabetes and Digestive and Kidney Diseases. The PID database contains 768 records each with 9 attributes.

Table 1. *List of PID database's features [14].*

No	Name	Descriptions
1	Pregnancies	The number of pregnancies
2	Glucose	PGC 2 hours in an OGTT
3	Blood Pressure	Diastolic blood pressure
4	Skin Thickness	Triceps skinfold thickness
5	Insulin	2 Hours serum insulin
6	BMI	Body mass index
7	Diabetes Pedigree Function	Diabetes pedigree function
8	Age	Age
9	Outcome	Class label

The data set consists of two classes. The number of non-diabetic patients is 500 and it is labelled as 0. The number of patients with diabetes is 268 and is labelled 1. All records in the data set consist of women aged between 21 and 81 [14].

Table 1 shows the name and description of the PID database features in the data set. In this study, the PID database is divided into two a training and test data set. The training data set consists of 512 records (2/3). The training set was used to train the machine learning models. The test data set consists of 256 records. The test data set was used to test the machine learning models that have been trained.

2.2 Machine Learning Methods

In this study, Random Forest Classifier, Extra Tree Classifier and Gaussian Process Classifier machine learning methods used to predict whether individuals have diabetes or not are explained below.

2.2.1 Random Forest Classifier

Random Forest is a method that aims to improve the classification result by using more than one decision tree. The number of decision trees used in the Random Forest is parametric. Decision trees formed within the scope of this parameter are formed from subsets chosen randomly from the data set. Training takes place on randomly selected subsets and prediction is made on each decision tree. As a result

of these estimates, the decision tree and estimate with the highest success rate are selected as the result [15, 16].

2.2.2. Extra Trees Classifier

Extra Trees is a tree-based community learning algorithm proposed by Geurts et al in 2006. Extra Trees is an algorithm that continues its process through this ensemble by combining predictions from multiple decision trees to derive the classification result. While Random Forest calculates the best variable selection during the best split Extra Trees chooses a random variable and a random cut point instead of the calculation. Thus, the diversity between trees increases and the number of splits decreases. Since the calculation cost for the split process decreases, the training time of the model is also shortened [16, 17].

2.2.3. Gaussian Process Classifier

The Gaussian Process defines various kernel functions on the input data and creates an output with the weighted sums of these functions. Radial-based functions are used as kernel functions. These functions produce a Gaussian output according to the distance of the input data from a point [18].

2.3 Performance Metrics

These are the metrics used to evaluate the prediction result of the developed machine learning model. These metrics are calculated using the parameters True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN). To calculate the prediction result of the developed model, the formulas for performance metrics are given below [19, 20].

$$Accuracy = \frac{(TP+TN)}{(TP+FN+FP+TN)} \quad (1)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (2)$$

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

$$F1-Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (4)$$

3. Experimental Study

In this study, a method is proposed to predict whether an individual is diabetic or not. The block diagram of the proposed method is shown in Fig. 1. The proposed method was developed using cloud computing-based Google Colab [21] and Google Drive service. Users must have a Google account to use these services. By using this account, the data set is uploaded to Google Drive and becomes accessible via cloud storage. Through Google

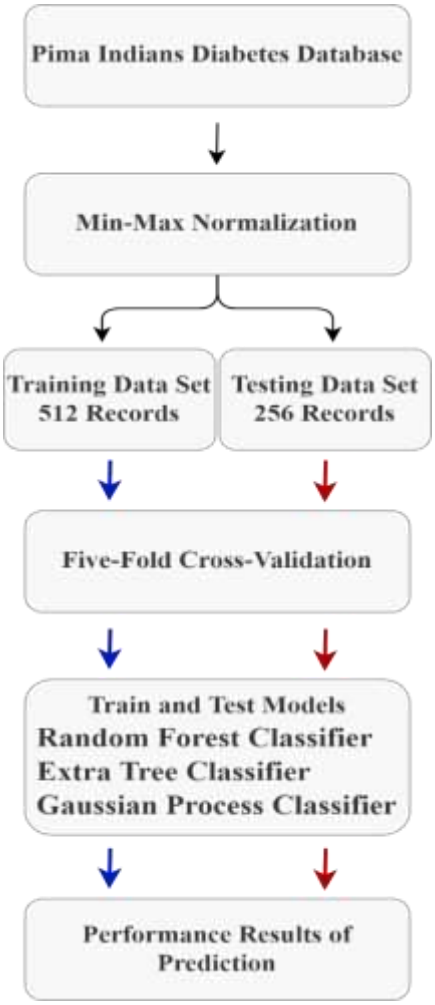


Figure 1. *Block diagram of the proposed method for diabetes prediction.*

Colab, pre-processing can be done on the data set in Python and the prediction results can be evaluated by developing machine learning models. Thus users can meet their software and hardware needs through the Google Cloud Computing Platform saving time and money.

In the proposed method, the PID database was loaded on Google Drive and made accessible via Google Colab. PID database features were normalized according to min-max normalization using Python programming language via Google Colab. Then, the PID database was divided into a training data set (512 records) and a test data set (256 records). With these data sets, machine learning models were fed by using the cross-validation technique. In this study, Random Forest Classifier, Extra Tree Classifier and Gaussian Process Classifier machine learning methods were used. Models were created using machine learning methods. Thus, the created models were trained, and training performance results were obtained. Then, the final prediction results were obtained by testing the trained models.

The proposed method was developed using classification-based machine learning methods that predict diabetes. The prediction results of the models belonging to these methods were obtained by using the PID database. The prediction results of the models were compared with each other, and the most successful model was proposed for diabetes prediction.

4. Experimental Results

In this study, the machine learning methods used the PID database to predict whether individuals have diabetes or not. The prediction results of the developed models are shown in Table 2. The models in Table 2 are listed from high performance to low performance. When Fig. 2 is examined, it is seen that the training results of the machine learning models in the study

Table 2. *Prediction results of developed models trained and tested using PID database.*

Methods		Accuracy (%)	F1-Score (%)	Precision (%)	Recall (%)
Random Forest Classifier	Training	82.08	84.64	77.09	93.81
	Testing	81.17	86.76	88.79	84.82
Extra Trees Classifier	Training	81.43	84.17	77.09	92.66
	Testing	78.57	84.93	86.97	83.04
Gaussian Process Classifier	Training	80.82	85.55	88.92	82.43
	Testing	77.49	84.15	87.89	80.70

In Fig. 2, the ROC AUC (Receiver Operating Characteristic Area Under Curve) values obtained from the training and test results of each model are shown. The diabetes prediction performance of the developed machine learning models in Fig. 2 has been compared.

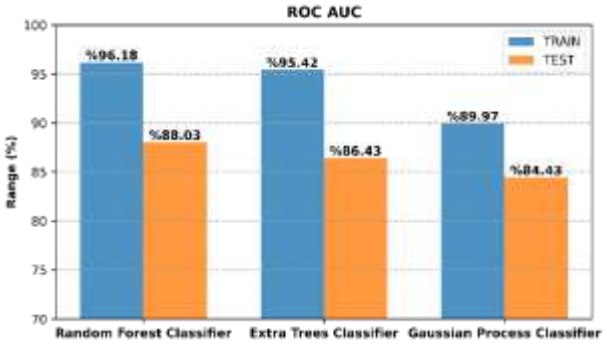


Figure 2. *ROC AUC of the developed machine learning models.*

are higher than the test results. When the test results of the machine learning models in the study were analyzed, the Random Forest Classifier performance result was 88.18%. The Extra Trees Classifier achievement result was 86.43%. The Gaussian Process Classifier performance result was 84.43%. When the performance values of the machine learning models in the study are investigated, the classifier with the highest value is Random Forest Classifier. Extra Trees Classifier and Gaussian Process Classifier are classifiers with second and third performance values respectively. The results of the second and third classifiers are comparatively lower than Random Forest Classifier.

5. Discussions

In Table 2, diabetes prediction made with the PID database was ranked from high performance to low performance. Random Forest Classifier method results were obtained Accuracy of 81.17%, F1-Score of 86.76%, Precision of 88.79%, and Recall of 84.82%. The performance results of the Extra Tree Classifier method were found Accuracy of 78.57%, F1-Score 84.93%, Precision of 86.97%, and Recall of 83.04%. The results of the Gaussian Process Classifier method achieved an Accuracy of 77.49%, F1- Score of 84.15%, Precision of 87.89%, and Recall of 80.70%. When the performances of the developed models were compared it was seen that the results were in the same range and close to each other. The machine learning methods in this study were compared with each other. So, it was concluded that the method with the highest performance was the Random Forest Classifier method.

In this study, ROC AUC results were obtained to show how well the machine learning methods predicted. The larger the area covered by the ROC AUC, the more effective the model was in distinguishing classes within the data set [22]. The ROC AUC values of the machine learning models in this study are shown in Fig. 2. When the ROC AUC values of the models were analyzed performances of models were listed as Random Forest Classifier 88.03%, Extra Tree Classifier 86.43% and Gaussian Process Classifier 84.43%. These results were evaluated according to the gradation of the AUC table [23]. Thus, it was concluded that the results obtained were quite close to each other and at the same time satisfactorily good.

When the results of the models in Table 2 and Figure 2 are examined, the method with the highest performance was the Random Forest Classifier. Within the scope of this study, the Random Forest Classifier method was proposed to estimate whether individuals had diabetes or not. The results of the Random Forest Classifier model are compared with

Table 3. *Comparison of the proposed method with studies in the literature.*

Model	Accuracy (%)
Proposed Method (Random Forest Classifier)	81.71
Naive Bayes [3]	78.57
Voting Classifier (XGB + RF) [4]	90.00
Random Forest [5]	79.57
Logistic Regression [6]	86.00
Logistic Regression [7]	76.80
Multilayer Perceptron [8]	78.10
Naive Bayes + Particle Swarm Optimization [9]	77.34
Random Forest [10]	75.00
Logistic Regression [11]	77.60
Naive Bayes [12]	76.30
Hierarchical Fuzzy Rule-based Evolutionary [13]	79.17

other studies in the literature in Table 3. In Table 3, the prediction accuracy of studies that predict diabetes using the PID database is compared with the prediction accuracy of the proposed method. As a result of the comparison, it is observed that the prediction degree of the proposed method is in the same range and close to each other with the models in the literature. High-performance models in the literature have produced better results by using the ensemble approach and editing the dataset. The data set used in our study is divided into 66% (2/3) for training and 34% (1/3) for testing. In such a case, the trained models are trained with less data. It is also being tested with more data. As a result, the model produces more durable, valid and reliable results. Also, the proposed method makes predictions using a single machine learning model. This situation affects the performance result of the proposed method. The Logistic Regression [6] model used the data set as 80% for training and 10% for testing. In this case, the Logistic Regression [6] model was trained with more data and tested with less data. Thus, the performance of the model is increased. The Voting Classifier (XGB + RF) [3] model used the data set as 70% for training and 30% for testing. They reduced the possibility of making mistakes in machine learning methods by reducing the number of features by making feature selections. In addition, they used two classifiers based on the ensemble learning approach in the prediction. As a result, they achieved a high level of accuracy.

When the ROC AUC result of the Random Forest Classifier shown in Figure 2 is examined, it is determined that the performance of this method is satisfactorily high. Also, the accuracy of the Random Forest Classifier is given in Table 2 as 81.71%. Therefore, the Random Forest Classifier method was proposed in this study, considering it useful and sufficient to predict diabetes. It is concluded that the proposed method will help clinicians' decision-making processes in computer-aided diagnosis systems.

6. Conclusions

In this study, a method that predicts whether individuals have diabetes is proposed by using Random Forest Classifier, Extra Tree Classifier and Gaussian Process Classifier methods. When the prediction results are analyzed, the Random Forest Classifier method, which has a satisfactory degree of success has been proposed within the scope of this study. The proposed method has been developed to assist clinicians in predicting the diagnosis of diabetic patients. It is thought that the proposed method can be useful in the decision support process by using computer-aided diagnosis systems. In future studies, a diabetes data set can be created for larger audiences and includes more records. The new data set can be analyzed using more advanced machine learning methods.

Author Statements:

- The conducted research is not related to either human or animal use.
- The authors declare that they have equal right on this paper.
- The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
- The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- The authors declare that they have nobody or no company to acknowledge.

References

[1]Diabetes Overview, (2023). *https://www.who.int/news-room/fact-sheets/detail/diabetes*

[2]Diabetes, (2023). *https://www.who.int/health-topics/diabetes#tab=tab_1*

[3]Febrian, M. E., Ferdinan, F. X., Sendani, G. P., Suryanigrum, K. M., & Yunanda, R. (2023). Diabetes prediction using supervised machine learning. *Procedia Computer Science*, 216; 21-30. https://doi.org/10.1016/j.procs.2022.12.107

[4]Kibria, H. B., Nahiduzzaman, M., Goni, M. O. F., Ahsan, M., & Haider, J. (2022). An ensemble approach for the prediction of diabetes mellitus using a soft voting classifier with an explainable AI. *Sensors*, 22(19); 7268. https://doi.org/10.3390/s22197268

[5]Chang, V., Bailey, J., Xu, Q. A., & Sun, Z. (2022). Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. *Neural Computing and Applications*, 1-17. https://doi.org/10.1007/s00521-022-07049-z

[6]Krishnamoorthi, R., Joshi, S., Almarzouki, H. Z., Shukla, P. K., Rizwan, A., Kalpana, C., & Tiwari, B. (2022). A novel diabetes healthcare disease prediction framework using machine learning techniques. *Journal of Healthcare Engineering*, https://doi.org/10.1155/2022/1684017

[7]Bhoi, S. K. (2021). Prediction of diabetes in females of pima Indian heritage: a complete supervised learning approach. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(10); 3074-3084. https://doi.org/10.17762/turcomat.v12i10.4958

[8]Güldoğan, E., Zeynep, T. U. N. Ç., Ayça, A. C. E. T., & ÇOLAK, C. (2020). Performance evaluation of different artificial neural network models in the classification of type 2 diabetes mellitus. *The Journal of Cognitive Systems*, 5(1); 23-32.

[9]Maulidah, N., Abdilah, A., Nurlelah, E., Gata, W., & Hasan, F. N. (2020). Seleksi Fitur Klasifikasi Penyakit Diabetes Menggunakan Particle Swarm Optimization (PSO) Pada Algoritma Naive Bayes. *Elkom: Jurnal Elektronika dan Komputer*, 13(2); 40-48.

[10]Tigga, N. P., & Garg, S. (2020). Prediction of type 2 diabetes using machine learning classification methods. *Procedia Computer Science*, 167, 706-716. https://doi.org/10.1016/j.procs.2020.03.336

[11]Jakka, A., & Vakula Rani, J. (2019). Performance evaluation of machine learning models for diabetes prediction. *Int. J. Innov. Technol. Explor. Eng.(IJITEE)*, 8(11). 10.35940/ijitee.K2155.0981119

[12]Sisodia, D., & Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. *Procedia computer science*, 132;1578-1585. https://doi.org/10.1016/j.procs.2018.05.122

[13]Feng, T. C., Li, T. H. S., & Kuo, P. H. (2015). Variable coded hierarchical fuzzy classification model using DNA coding and evolutionary programming. *Applied Mathematical Modelling*, 39(23-24);7401-7419. https://doi.org/10.1016/j.apm.2015.03.004

[14]Pima Indians Diabetes Database, (2022). https://data.world/data-society/pima-indians-diabetes-database

[15]Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

[16]Yakut Ö., Bolat E. D. (2020). Arrhythmia Diagnosis from ECG Signal Using Tree-based Machine Learning Methods. *International Journal of Mathematic Engineering and Natural Sciences*, 4(16);954-964. https://doi.org/10.38063/ejons.361

[17]Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63(1); 3-42. https://doi.org/10.1007/s10994-006-6226-1

[18]MacKay, D. J. (1998). Introduction to Gaussian processes. *NATO ASI series F computer and systems sciences*, 168; 133-166.

[19]Yakut, Ö. (2020, November 19-21). *Cloud Computing Based Voting Classifier Method Used For Survival Prediction of Heart Failure Patients*. International Conference on Engineering Technologies, Konya-Turkey. https://icente.selcuk.edu.tr/

[20]Yakut, Ö. (2020, November 28-29). *Comparison of Clustering Methods For Early Stage Diabetes Risk Prediction Using Cloud Computing*. International Black Sea Coastline Countries Symposium 5, Zonguldak-Turkey.

[21]Google Colab Notebbook, (2022). https://colab.research. google.com

[22]Yakut, Ö., Bolat, E. D. (2020). An Efficient Arrhythmic Heartbeat Classification Method Using ECG Morphology Based Features. *Euroasia Journal of Mathematics, Engineering, Natural & Medical Sciences*, 7(13);200-212. https://doi.org/10.38065/euroasiaorg.403

[23]Yakut, Ö., Timuş, O., Bolat, E. D. (2016). HRV Analysis Based Arrhythmic Beat Detection Us-ing kNN Classifier. *International Journal of Biomedical and Biological Engineering*, 10(2); 60-63. doi.org/10.5281/zenodo.1339067