Early prediction of COVID-19 infection using data mining and multi machine learning algorithms

Ahmed Jaddoa Enad[1], Mustafa Aksu[2]
[1]Institute of Natural and Applied Sciences, Department of Advanced Technologies, Kırşehir Ahi Evran University, Kırşehir,Turkey
[2]Faculty of Engineering and Architecture Department of Computer Engineering, Kırşehir Ahi Evran University, Kırşehir, Turkey

Article Info

Article history:

Received Jun 7, 2023
Revised Aug 4, 2023
Accepted Oct 12, 2023

Keywords:

Artificial intelligence
Coronavirus disease
Data mining
Machine learning
Prediction

ABSTRACT

The fields of artificial intelligence (AI) and machine learning (ML) have attracted significant interest and investment from a diverse range of industries, especially during the last several years. Despite the fact that AI methods have been used extensively and put through extensive testing in the healthcare industry, the recently discovered coronavirus disease (COVID-19) necessitates the use of these methods in order to prevent the emergence of the disease. The proposed system is based on six ML algorithms to predict COVID-19 infection as random forest (RF) algorithm, naive bayes (NB) algorithm, support vector machine (SVM) algorithm, decision tree (DT) algorithm, multi-layer perceptron (MLP), and k-nearest neighbor (KNN). It is based on two steps: first, we uploaded the dataset to train the model. Then, we test our model on those cases to work directly after making a trained classifier so it can directly discover with automatic COVID-19 prediction state of a patient suspected or not. The proposed system results showed the high accuracy of NB, DT, and SVM as 98.646%. Besides the better time to build the model and early predict the state of patients is 31 ms of the NB algorithm.

Corresponding Author:

Ahmed Jaddoa Enad
Institute of Natural and Applied Sciences, Department of Advanced Technologies
Kırşehir Ahi Evran University
40100, Kirşehir, Turkey
Email: aaggaa1981@gmail.com

1. INTRODUCTION

Since March 2020, a pandemic caused by the coronavirus disease (COVID-19) virus has been ongoing for about 2 years. In spite of the vaccination programs that are prevalent in a variety of nations, there is still an ongoing growth in the number of persons who are sick. Because of the many distinct forms and mutations that have occurred, the COVID-19 virus has become very contagious, fatal, and in some cases undetected [1]. This has led to an increase in the number of persons who have been infected with the virus [2].

The fields of artificial intelligence (AI) and machine learning (ML) have attracted significant interest and investment from a diverse range of industries, especially during the last several years [3]. Despite the fact that AI methods have been used extensively and put through extensive testing in the healthcare industry, the recently discovered COVID-19 necessitates the use of these methods in order to diagnose, forecast, and prevent the emergence of the disease [4]. It has been hypothesized that the use of AI methods would bring about a paradigm change in the field of healthcare [5], and it is possible that this would necessitate the application of these techniques to the ongoing COVID-19 pandemic. Improving the precision of COVID-19 diagnosis is imperative to expeditiously detect affirmative cases, thereby mitigating additional transmissions and guaranteeing prompt medical attention for patients [6].

ML is a subfield of AI that is concerned with the development of intelligent applications that can learn from data and enhance their accuracy without explicit programming [7]. Through the use of training algorithms, ML models can identify patterns and features in data, enabling them to make informed decisions and predictions based on new data [8]. The ultimate goal of ML is to achieve optimal performance in handling complex and dynamic real-world problems in healthcare researches side. ML algorithms typically operate through a structured sequence of stages, commencing with the identification and preparation of a training dataset [9].

The most related works in term of ML and deep learning algorithms used for COVID-19 infection have been discussed and overviewed as follow:

Alakus and Turkoglu [10] using a convolutional neural network (CNN), they demonstrated a technique for selecting and extracting features from images for further classification. CNNs may provide superior accuracy over other classifiers. The efficiency and precision are tested on both a regular CPU and a GPU. So, we draw the conclusion that CNNs are a great choice for picture categorization. Biometric features might be added to this system in the future.

Arpaci *et al.* [11] developed of six distinct prediction models for COVID-19 diagnosis, utilising six distinct classifiers, including Bayes net, logistic, instance based learner (IBk), PART algorithm, and decision tree (DT). The classifiers were developed using a collection of 14 clinical features. According to the findings, the CR meta-classifier demonstrates a notable degree of precision, particularly 84.21%, in its ability to forecast affirmative and negative instances of COVID-19, particularly in situations where RT-PCR kits are insufficient in verifying the existence of the infection. Furthermore, these results could be beneficial to countries, particularly those with limited resources, that encounter difficulties in obtaining RT-PCR assays and specialised facilities.

ML techniques are provided in [12]. According to the findings of our experiments, the blood glucose level is the factor that has the most impact on one's ability to predict COVID-19 in this specific dataset. According to the findings, XGBoost has the greatest accuracy value for the case of cv, with a value of 92.67%, while LR has the second best accuracy value, with a value of 92.58%. On the other hand, the values for precision, recall, and F1 score for both XGBoost and LR are the same, at 93%. LR demonstrates the maximum level of testing accuracy, which is 94.06%, when the holdout technique is used with 20% of the testing data samples. As a result, XGBoost and LR are both viable options for predicting COVID-19.

Ong *et al.* [13] utilized a deep learning neural network and a random forest (RF) classifier. They employed a convenience sampling technique to gather information from a cohort of 800 respondents. The principal aim of the investigation was to assess a range of factors, encompassing knowledge about COVID-19, it was found that a substantial majority of 97.32% of the participants attributed the perceived usefulness of COVID-19 to their understanding of the disease. Furthermore, the results indicate that the RF classifier achieved a precision rate of 92%, accompanied by a standard deviation of 0.00. The findings of this investigation suggest that a favourable association exists between possessing knowledge regarding COVID-19 and the perception of vulnerability to it, as well as an augmented perception of the efficacy of measures implemented to hinder its transmission.

Moulaei *et al.* [14] aimed to assess multiple ML algorithms to predict the COVID-19 mortality rate based on patient data collected during their initial hospital admission. It was found that the three most significant predictors were dyspnea, hospitalisation in the intensive care unit, and treatment with oxygen. The analysis encompassed a total of 38 distinct characteristics. The study revealed that smoking, alanine aminotransferase, and platelet count exhibited the lowest precision in forecasting mortality due to COVID-19. The experimental findings indicate that the RF technique outperformed other ML algorithms in terms of accuracy, sensitivity, precision, and specificity, with scores of 95.03%, 90.70%, 94.23%, and 95.10%, respectively. Additionally, the receiver operating characteristic (ROC) score was 99.02%.

In this paper, comparing ML algorithms for predicting COVID-19 infection has been proposed. It is totally presented as follows: section 1 is introduction, section 2 is method, section 3 is results and discussion, and section 4 is conclusion.

2. METHOD

The system under consideration has been executed using the java eclipse programming environment. Java is utilized for the implementation of ML algorithms. The process comprised three primary phases, namely:

- The initial stage involves pre-processing of data mining on the complete COVID-19 dataset to convert the raw data into a format that is both effective and efficient.
- In the second stage, the pre-processed training dataset is utilized to generate value attributes.
- Phase three involves the utilization of ML algorithms to obtain outcomes, as the stages are shown in Figure 1.

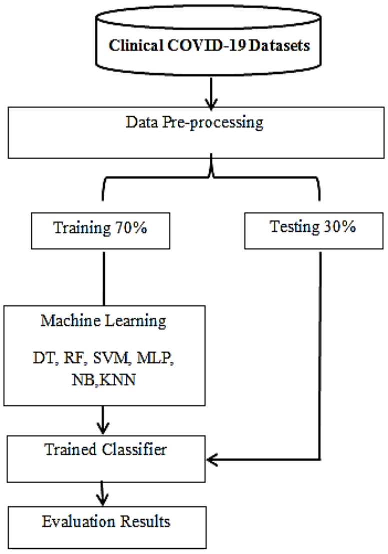

Figure 1. The proposed system model

It is considered one of the important processes in extracting unstructured data and the task of converting it into a meaningful and effective format on the other jobs, alongside, it based on the useful data from preprocessing to evaluated with ML classifiers as it showed in Figure 2 as the main steps of data preprocessing. So, the main contribution of this paper is how to create an early prediction of COVID-19 and related diseases. The still open problem in this field which we dealt with is the accuracy to predict COVID-19 and the required time to build the prediction system to get accurate results.

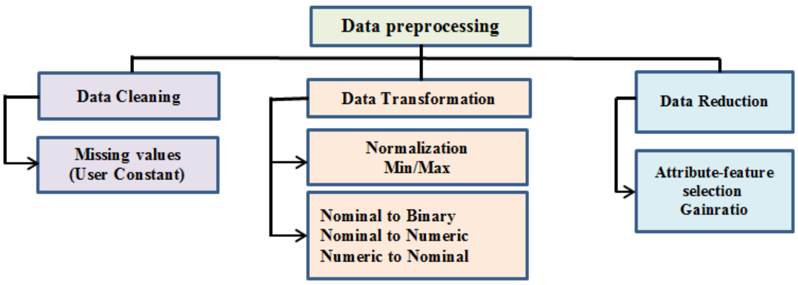

Figure 2. The used data mining pre-processing methods

a. Normalization: it is done in order to scale the data values in a specified range.
b. Attribute-feature-selection: it is very important to use attributes that are related and interrelated with each other and it is possible to get rid of other characteristics as it has become necessary to use a high level of importance of characteristics and ignore other characteristics of little importance is very important in the implementation process. The proposed method used a gain ratio to determine the splits and to select the most important features.
c. The missing-values: this value can be exchanged for days, the reason, the maximum value, or the average value. In some cases, the zero value can be used for the missing values. Also, fixed values can be adopted as an alternative to the missing action. Within the proposed system method used as user constant value to replace missed value attributes in dataset records.
d. The method employed by the proposed system for converting nominal attribute values into binary states was nominal to binary. The present methodology involves the utilization of a system that operates by converting nominal attributes (i.e. string values) within the COVID-19 dataset into binary data, represented as (0, 1). This approach is deemed optimal for use in the proposed ML algorithm, as it serves to enhance the accuracy of prediction.
e. Nominal to numeric: the nominal to binary method was utilized by the proposed system to transform nominal attribute values into binary states. The current approach entails the utilization of a system that functions by transforming nominal attributes, specifically string values, present in the COVID-19 dataset

into binary data, which is denoted as (0, 1). The aforementioned methodology is considered to be the most advantageous for implementation in the proposed ML algorithm, as it effectively improves the precision of forecasting. In this work, the method is used to convert string values in dataset attributes such as ever_married, smoking_status, and residence_type attributes.

f. Numeric to nominal: previously, it was utilized to transform numerical data into categorical data. In this work, the used method applied to class values to deal with them as nominal values through classify the class state as patient_test_status (class 1) with COVID-19 and non-COVID-19 (class 0).

Furthermore, when gathering the healthcare dataset pertaining to COVID-19, it is observed that the data comprises both categorical and numeric variables. As ML algorithms are designed to comprehend numeric data, it is recommended to transform the categorical data into numeric data through techniques such as label encoder or one hot encoding.

The label encoder technique, which falls under data mining transformation techniques, involves the conversion of categorical data into numeric data. The process involves the conversion of ascending numerical values into a numeric data range of 0 to n-1. Besides, Table 1 showed the used data preprocessing methods for each attribute in the COVID-19 dataset.

Table 1. The present study aims to elucidate the data mining methodology employed for each attribute in the COVID-19 dataset

Column name	Type	Value	Data mining (pre-processing)
Gender	String	[Male female]	Transformation (nominal to binary)
Age	Double	[0.08] – [82]	Attribute-feature selection (gainratio)
Hypertension	integer	[0 1]	Normalization (min/max)
heart_disease	integer	[0 1]	Normalization (min/max)
ever_married	String	[Yes no]	Transformation (nominal to binary)
Patient_test_clinical	integer	[0 1]	Normalization (min/max)
Residence_type	String	[Urban rural]	Transformation (nominal to binary)
O2	String	[Yes, no]	Transformation (nominal to binary)
smoking_status	String	[formerly smoked, never smoked, smokes, unknown]	Transformation (nominal to numeric), cleaning replace missing value (user constant)
Patient_test_status	Integer	[0 1]	Transformation (numeric to nominal)

The study utilized a series of assessment scales that were founded on the confusion matrix framework [15]. Specifically, a set of equations with distinct nomenclature were employed, as exemplified in (1) through (6) [16].

- Precision: the metric being referred to is the ratio of true positives (TP) to the sum of true positives and false positives (TP+FP), commonly known as the TP rate or precision. The computation was performed using (1) [17].

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{1}$$

- Accuracy: the accuracy of a prediction model is determined by dividing the number of correct predictions by the total number of predictions. The calculation was performed using (2) [18]:

$$\text{Accuracy} = \frac{\text{TP + TN}}{\text{TP+TN+FP + FN}} \tag{2}$$

- Recall: the aforementioned expression denotes the ratio of true positives to the sum of true positives and false negatives. The computation of this metric can be derived from (3) as stated in [19].

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

- Detection rate (DR): it is the proportion of correctly identified positive (anomaly) instances [20], it is calculated by dividing the number of true positive instances by the total number of actual positive instances [21]. The computation of this metric is feasible by utilising (4) [22]:

$$\text{DR} = \frac{\text{TP}}{\text{TP+ FN}} \tag{4}$$

- False alert rate (FAR): the metric denotes the ratio of negative predictions that are erroneously classified as positive (anomalies) [23]. A lower value is considered to be more desirable. The computation of this metric can be derived from (5) [24]:

$$\text{FAR} = \frac{\text{FP}}{\text{FP+ TN}} \quad (5)$$

- Error rate (ERR): the operational definition of "accuracy" can be expressed as the proportion of incorrect predictions to the overall number of predictions conducted on a specific dataset, as depicted in (6) [25]:

$$\text{ERR} = \frac{\text{b+c}}{\text{a+ b+c+d}} \quad (6)$$

3. RESULTS AND DISCUSSION

The system under consideration is predicated on three distinct case studies that utilise ML, as illustrated in Figure 3. The utilized system was established within a framework that adhered to the specifications outlined in Table 2. The proposed system is based on the 1st case study on data mining (preprocessing) which is evaluated with ML classifiers with the maximum accuracy and minimum time required to build the system. The results show the decision tree (DT), support vector machine (SVM), and naive bayes (NB) algorithms are the best classifiers in the proposed system. Table 3 showed the used features of the dataset.

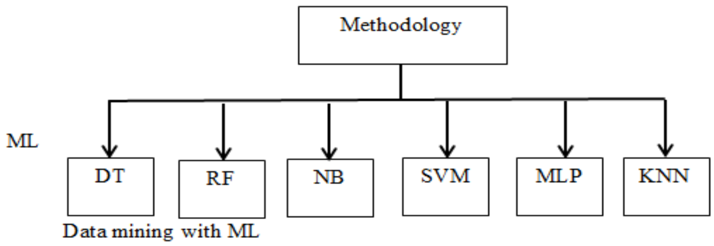

Figure 3. The used ML algorithms

Table 2. The environmental requirements for the system under consideration

Operating systems	Windows 7
CPU	Core (TM) I5-3630
RAM	4.00 GB
Implementation tools	Java, eclipse IDE for java EE developers luna SR2 v4.9

Table 3. The COVID-19 dataset contains a specific quantity of records and associated attributes

COVID-19 dataset features	
Number of data instances	2585
Number of data attributes	10

Besides, the main accuracy details of the proposed case study using data mining/ML on the COVID-19 dataset DT is the high accuracy of 98.646%, and the time to build model is 250 ms, while NB is the same accuracy as DT but the time take to build model is 31 ms. Besides, K-nearest neighbor (KNN) is a low result accuracy at 96.325%, and the time take to build the model is 5 ms. Table 4 presents the precision and temporal specifics, along with the confusion matrix assessed metrics such as FPR and FNR, for the COVID-19 data instance. Table 5 presents the results of correctly classified with incorrectly classified instances of the proposed data mining (preprocessing) with the highly accurate ML algorithms (DT, SVM, and NB) on the testing dataset used (COVID-19 dataset).

Table 4. The results of ML for COVID-19 data analysis

Item	Method name	Accuracy (%)	Confusion matrix		Time (ms)
			False positive rate	False negative rate	
1	DT	98.646	7	0	250
2	SVM	98.646	7	0	499
3	RF	98.4526	6	2	655
4	NB	98.646	7	0	31
5	Multi-layer perceptron (MLP) neural	98.4526	7	1	9033
6	KNN	96.325	6	13	5

Table 5. Correctly/incorrectly classified testing instances of the data preprocessing

ML algorithm	Correctly classified	Incorrectly classified
DT	510=98.646%	7=1.354%
SVM	510=98.646%	7=1.354%
RF	509=98.4526%	8=1.5474%
NB	510=98.646%	7=1.354%
MLP neural	509=98.4526%	8=1.5474%
KNN	498=96.325%	19=3.675%

Furthermore, the evaluation criteria used in the proposed system as mean absolute error (MAE), root mean squared error (RMSE), and ERR shown in Table 6 and Figure 4 show the prediction of the evaluation criteria of the proposed algorithms, the SVM as 0.0135 almost lower the MAE value, so it is the better compared with others. DT results of the RMSE statistic are lower as the better 0.1157 compared with other algorithms. DT, SVM, and NB results of ERR were better for these algorithms.

Table 6. MAE and RMSE for the COVID-19 ML

Evaluation criteria	Predication					
	DT	SVM	RF	NB	MLP	KNN
MAE	0.0309	0.0135	0.0317	0.0415	0.02827	0.0372
RMSE	0.1157	0.1164	0.1269	0.1215	0.09142	0.1916
ERR	0.01353	0.01353	0.01547	0.01353	0.01547	0.03675

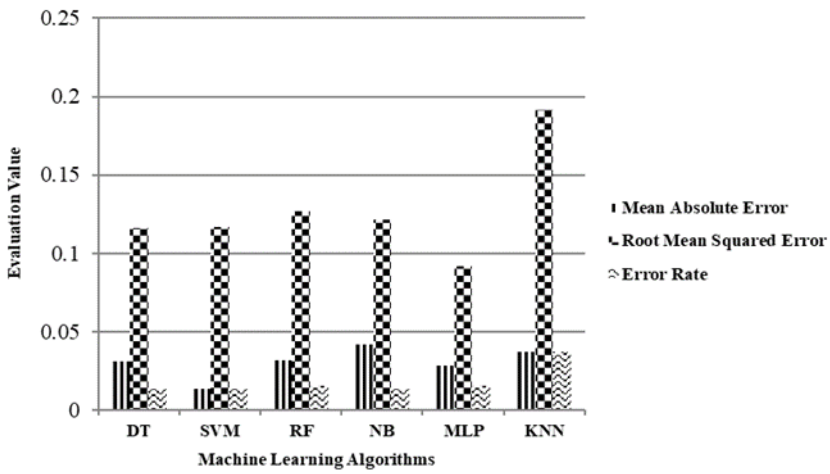

Figure 4. The main evaluation parameters MAE, RMSE for the COVID-19 dataset analysis

In addition, the proposed system includes various evaluation classifiers, as outlined in Table 7, which are based on six distinct ML classifiers. These classifiers have been implemented for both normal cases without COVID-19 (class 0) and cases with COVID-19 (class 1), as illustrated in Figure 4 of the COVID-19 case based on confusion matrix values. DT, SVM, and NB precision as 0.98646 can be seen as a measure of high quality to return more relevant results than irrelevant ones and recall as a measure of quantity. In addition, Figure 5 is showed precision, recall, F-measure, DR, and FAR of the proposed machine-learning algorithms. Furthermore, the case studies demonstrate that the proposed system yields superior accuracy, particularly in the context of NB, DT, and SVM, where the accuracy rate reaches 98.646%, as it showed in Table 8.

Table 7. Using ML for massive data analysis: a COVID-19 presence evaluation

Evaluation parameters	ML algorithms					
	DT	SVM	RF	NB	MLP	KNN
Precision	0.98646	0.98646	0.98452	0.98646	0.98452	0.96325
DR	1.0	1.0	0.99607	1.0	0.99803	0.97450
FAR	1.0	1.0	0.85714	1.0	1	0.85714
TP rate	510	510	508	510	509	497
TN rate	0	0	1	0	0	1

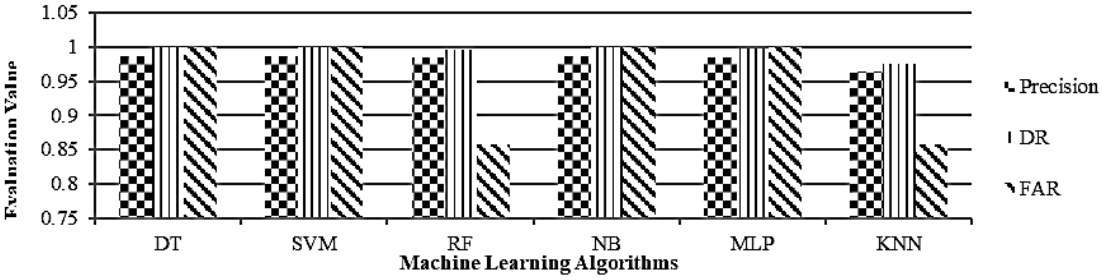

Figure 5. Precision, recall, F-measure, DR and FAR of the proposed ML algorithms

Table 8. The results of COVID-19 dataset analysis with the compared systems

Ref	Year	AI technique	Accuracy (%)
[10]	2020	CNN-LSTM	92.30
[11]	2021	Logistic regression with CR classifier	84.21
[12]	2021	Logistic regression and XGBoost	94 and 92
[13]	2022	Neural network and RF	97.32 and 92
[14]	2022	RF	95.03
Proposed-system		NB, DT, and SVM	98.646

4. CONCLUSION

The COVID-19 hospitalized patients are always at risk of death. ML algorithms can be used as a potential solution for predicting mortality in COVID-19 hospitalized patients. It used ML to create a fast detection system like a real-time detection warning system to identify those who are suspected. The pre-trained algorithms to classify the dataset contents. We cleaned the dataset with preprocessing methods by removing duplicates and normalized the attributes to increase the model's accuracy; in addition to preprocessing our data, the data set was loaded from the folder as input to the system model. The data has been tested on COVID-19 patients utilizing 10 independent variables. The classification scheme utilized in the present study involves two distinct categories: class 0, which denotes the absence of COVID-19 in patients, and class 1, which signifies the presence of COVID-19 in patients. Next, we were able to improve the model's accuracy by preprocessing the data set with the used data mining pre-processing methods. The future research direction is to detect COVID-19 from chest x-ray images through the application of transfer learning techniques using ResNet50, ResNet101, DenseNet121, DenseNet169, and InceptionV3 models.

REFERENCES

[1] A. Morand *et al.*, “COVID-19 virus and children: What do we know?,” *Archives de Pediatrie*, vol. 27, no. 3, pp. 117–118, Apr. 2020, doi: 10.1016/j.arcped.2020.03.001.

[2] K. Usher, J. Durkin, and N. Bhullar, “The COVID-19 pandemic and mental health impacts,” *International Journal of Mental Health Nursing*, vol. 29, no. 3, pp. 315–318, Jun. 2020, doi: 10.1111/inm.12726.

[3] D. Gondauri and M. Batiashvili, “The Study of the Effects of Mobility Trends on the Statistical Models of the COVID-19 Virus Spreading,” *Electronic Journal of General Medicine*, vol. 17, no. 6, p. em243, Apr. 2020, doi: 10.29333/ejgm/8212.

[4] A. Sepehrinezhad, A. Shahbazi, and S. S. Negah, “COVID-19 virus may have neuroinvasive potential and cause neurological complications: a perspective review,” *Journal of NeuroVirology*, vol. 26, no. 3, pp. 324–329, Jun. 2020, doi: 10.1007/s13365-020-00851-2.

[5] C. M. Toquero, “Challenges and Opportunities for Higher Education amid the COVID-19 Pandemic: The Philippine Context,” *Pedagogical Research*, vol. 5, no. 4, p. em0063, Apr. 2020, doi: 10.29333/pr/7947.

[6] T. Chuenyindee, A. K. S. Ong, Y. T. Prasetyo, S. F. Persada, R. Nadlifatin, and T. Sittiwatethanasiri, “Factors Affecting the Perceived Usability of the COVID-19 Contact-Tracing Application ‘Thai Chana’ during the Early COVID-19 Omicron Period,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 7, p. 4383, Apr. 2022, doi: 10.3390/ijerph19074383.

[7] L. Bai *et al.*, “Chinese experts’ consensus on the Internet of Things-aided diagnosis and treatment of coronavirus disease 2019 (COVID-19),” *Clinical eHealth*, vol. 3, pp. 7–15, 2020, doi: 10.1016/j.ceh.2020.03.001.

[8] D. N. Vinod and S. R. S. Prabaharan, “COVID-19-The Role of Artificial Intelligence, Machine Learning, and Deep Learning: A Newfangled,” *Archives of Computational Methods in Engineering*, vol. 30, no. 4, pp. 2667–2682, May 2023, doi: 10.1007/s11831-023-09882-4.

[9] H. Swapnarekha, H. S. Behera, J. Nayak, and B. Naik, “Role of intelligent computing in COVID-19 prognosis: A state-of-the-art review,” *Chaos, Solitons and Fractals*, vol. 138, p. 109947, Sep. 2020, doi: 10.1016/j.chaos.2020.109947.

[10] T. B. Alakus and I. Turkoglu, “Comparison of deep learning approaches to predict COVID-19 infection,” *Chaos, Solitons and Fractals*, vol. 140, p. 110120, Nov. 2020, doi: 10.1016/j.chaos.2020.110120.

[11] I. Arpaci, S. Huang, M. Al-Emran, M. N. Al-Kabi, and M. Peng, “Predicting the COVID-19 infection with fourteen clinical features using machine learning classification algorithms,” *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 11943–11957, Mar. 2021, doi: 10.1007/s11042-020-10340-7.

[12] P. Podder, S. Bharati, M. R. H. Mondal, and U. Kose, "Application of machine learning for the diagnosis of COVID-19," in *Data Science for COVID-19 Volume 1: Computational Perspectives*, Elsevier, 2021, pp. 175–194. doi: 10.1016/B978-0-12-824536-1.00008-3.

[13] A. K. S. Ong *et al.*, "Utilization of Random Forest and Deep Learning Neural Network for Predicting Factors Affecting Perceived Usability of a COVID-19 Contact Tracing Mobile Application in Thailand 'ThaiChana,'" *International Journal of Environmental Research and Public Health*, vol. 19, no. 10, p. 6111, May 2022, doi: 10.3390/ijerph19106111.

[14] K. Moulaei, M. Shanbehzadeh, Z. Mohammadi-Taghiabad, and H. Kazemi-Arpanahi, "Comparing machine learning algorithms for predicting COVID-19 mortality," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, p. 2, Jan. 2022, doi: 10.1186/s12911-021-01742-0.

[15] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Mining*, vol. 16, no. 1, p. 4, Feb. 2023, doi: 10.1186/s13040-023-00322-4.

[16] S. Kaur and M. Singh, "Hybrid intrusion detection and signature generation using Deep Recurrent Neural Networks," *Neural Computing and Applications*, vol. 32, no. 12, pp. 7859–7877, Jun. 2020, doi: 10.1007/s00521-019-04187-9.

[17] R. N. Patil, S. Rawandale, N. Rawandale, U. Rawandale, and S. Patil, "An efficient stacking based NSGA-II approach for predicting type 2 diabetes," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 1, pp. 1015–1023, Feb. 2023, doi: 10.11591/ijece.v13i1.pp1015-1023.

[18] D. S. David, "Enhanced glaucoma detection using ensemble based CNN and spatially based ellipse fitting curve model," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 4, pp. 3303–3314, Apr. 2023, doi: 10.1007/s12652-021-03467-4.

[19] G. M. Amer, E. Abd El Hay, I. Abdel-Baset, and M. Abd El A. Mohamed, "Development Machine Learning Techniques to Enhance Cyber Security Algorithms. (Dept. E)," *MEJ. Mansoura Engineering Journal*, vol. 46, no. 4, pp. 36–46, Nov. 2021, doi: 10.21608/bfemu.2021.206401.

[20] P. Theerthagiri, I. Jeena Jacob, A. Usha Ruby, and V. Yendapalli, "Prediction of covid-19 possibilities using knearest neighbour classification algorithm," *International Journal of Current Research and Review*, vol. 13, no. 6 special Issue, p. S-156-S-164, 2021, doi: 10.31782/IJCRR.2021.SP173.

[21] D. Ngabo, W. Dong, E. Ibeke, C. Iwendi, and E. Masabo, "Tackling pandemics in smart cities using machine learning architecture," *Mathematical Biosciences and Engineering*, vol. 18, no. 6, pp. 8444–8461, 2021, doi: 10.3934/mbe.2021418.

[22] E. Mushtaq, A. Zameer, M. Umer, and A. A. Abbasi, "A two-stage intrusion detection system with auto-encoder and LSTMs," *Applied Soft Computing*, vol. 121, p. 108768, May 2022, doi: 10.1016/j.asoc.2022.108768.

[23] A. K. Balyan *et al.*, "A Hybrid Intrusion Detection Model Using EGA-PSO and Improved Random Forest Method," *Sensors*, vol. 22, no. 16, p. 5986, Aug. 2022, doi: 10.3390/s22165986.

[24] M. Abdullah, M. Al-Ayyoub, F. Shatnawi, S. Rawashdeh, and R. Abbott, "Predicting students' academic performance using e-learning logs," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 2, pp. 831–839, Jun. 2023, doi: 10.11591/ijai.v12.i2.pp831-839.

[25] M. Atif, F. Anwer, F. Talib, R. Alam, and F. Masood, "Analysis of machine learning classifiers for predicting diabetes mellitus in the preliminary stage," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 3, pp. 1302–1311, Sep. 2023, doi: 10.11591/ijai.v12.i3.pp1302-1311.

BIOGRAPHIES OF AUTHORS

Ahmed Jaddoa Enad is currently pursuing her master of Computer Engineering at Kırşehir Ahi Evran Universitesy. He will graduate in 2023. His Bachelor of Computer Science is from University of Babylon in Iraq-College of Computer Science. His main area of research is resource allocation in machine learning. He is currently focusing on the machine learning, artificial intelligence, prediction hospital mortality, and deep learning approaches. He can be contacted at email: aaggaa1981@gmail.com.

Mustafa Aksu is received his Bachelor's degree in Computer Science Engineering from Kocaeli University, Kocaeli, in 1998. He received His Master's degrees in Electrical and Electronics Engineering from the KSU University, Kahramanmaraş, Turkey, in 2004. He is currently a Ph.D. candidate in the Department of Computer Science, Inonu University, Malatya, Turkey. His research interests include data structures and algorithms, image processing, data mining, mobile systems, and applications. He can be contacted at email: mustafa.aksu@ahievran.edu.tr.