Journal of Intelligent Systems and Internet of Things
Vol. 18, No. 01, PP. 341-348, 2026
DOI: https://doi.org/10.54216/JISIoT.180126

Symptom-Based Detection of COVID-19 Cases Using Machine Learning Algorithms

Hussein Ibrahim Hussein[1,2,*], Lateef Abd Zaid Qudr[1], Weal Hasan Ali Almohammed[3]

[1]Department of computer engineering techniques, Alsafwa university college, Almamalie str Karbala, Iraq

[2]Department of Information Security, college of information technology, University of Babylon, Hillah, Iraq

[3]Department of Computer Science, College of Computer Science and Information Technology, University of Kerbala, Karbala, Iraq

Emails: Hussein.sarhan@alsafwa.edu.iq; latifkhder@alsafwa.edu.iq; wael.h@uokerbala.edu.iq

Abstract

Mammals are susceptible to the lethal disease called coronavirus. This virus often infects humans through the aerial precipitation of any fluid released from the bodily part of the affected entity. This viral variant is deadlier than other sudden viruses. Given the ongoing thread which COVID-19 on health systems in the worldwide, there is a rising interest in development a mechanism that effective in terms of cost and classification. A mechanism for categorizing and scrutinizing the estimations derived from this virus' symptoms is proposed in this paper. The precision of various machine-learning classifiers is calculated in this study in order to determine the optimal classifier for COVID-19 identification. Because the COVID-19 dataset has the greatest precision of 100%, it was classified using AdaBoost and Bagging. Additionally, precision, recall, and F-score measures together with the ROC were deployed for evaluating detection performance to ensure the approach is capable and successful.

Received: March 01, 2025 Revised: May 23, 2025 Accepted: July 02, 2025

Keywords: Covid-19; Detection; Bagging; AdaBoost; Machine Learning

1. Introduction

Living a healthy life requires maintaining and improving one's health. However, the COVID-19 outbreak has emerged as the greatest threat to people's ability to survive. The recently identified COVID-19 virus has been the cause of the deadly, widespread disease known as COVID-19. In the Chinese province of Wuhan, this sickness was first discovered at the end of 2019. The COVID-19 disease, that raised by the coronavirus SARS-CoV-2, has deep negative impacts on different life's aspects such as healthcare, social, commercial, and so on since its evolution in 2019 [1]. This new coronavirus, COVID-19, is formed by a brand-new member of the coronavirus family. The results demonstrate that COVID-19, which causes severe respiratory issues in those who are infected, is passed from individual to individual. Like past pandemics, COVID-19 is a present global threat that affects the health system and poses a significant risk to the world economy. The most typical signs of this virus, according to WHO, are fatigue, fever, and a dry cough [2]. People who have these minor symptoms can get healed without needing any extra care or medicine. Few of the patients reported the following additional symptoms: sore throat, runny nose, nasal congestion, pain, soreness, or diarrhoea.

Due to the wide and swift spread across borders, health systems in both developing and developed countries have been swamped because of the high request for diagnostic and treatment services. Hence, an accurate and timely diagnosis has become crucial for efficient patient' management, as well as preventing community transmission. However, there are traditional diagnostic mechanisms such as reverse transcription polymerase chain reaction (RT-PCR), CT scans, and X-Rays have been utilized in the battle against COVID-19. Although, there is high accuracy in the diagnosis of these traditional mechanisms, there are different challenges including but not limited to time-

consumingness, costliness, lack of early detection, and limited availability [3, 4]. To overcome such challenges, computational methods, in particular machine learning algorithms, are employed to develop scalable, effective, and timely solutions for COVID-19 classification. ML algorithms have shown the ability to promote pre-screening by analysing huge datasets that combine patient-reported symptoms, medical images, and clinical data. They are learning symptomatic patterns, predicting infection probabilities, supporting the triage of patients for further tests, and automating the final diagnostic decision-making. Eventually, the combination of machine learning algorithms in COVID-19 diagnosis promises to alleviate the strain on healthcare infrastructures and improve the efficiency of medical resource utilization [5].

The method of diagnosis of COVID-19 based on machine learning algorithms is divided into two main categories based on the type of data that is used for detecting it, which are image-based detection and symptom-based detection. In the first type, image-based detection, the system utilizes images from CT scans or X-rays of lungs of patients in order to detect the COVID-19 infection. In particular, there are two key methods: features extraction and transfer learning using pre-trained deep learning algorithms. In the case of the features extraction method, the features have been selected manually or automatically by using machine-learning techniques. Besides that, the transfer learning method uses pre-trained deep learning algorithms to extract features and test on unseen data. However, image-based detection suffers, as previously discussed, from scalability and costliness issues. As well as, the clinical images are difficult to obtain in a short period. Hence, they are constrained in use, especially for limited resource systems [6,7]. On the other hand, symptom-based detection, as a practical method, uses the clinical assessment of symptoms to detect the presence of COVID-19. With this intention, prevalent symptoms, including but not limited to fatigue, loss of smell, cough, fever, and loss of taste, are employed in this method. These data are collected via questionnaire, audio data laboratory examination, and self-reported symptoms [8].

Considering that, symptom-based detection is a valuable method since it has several advantages in terms of cost-effectiveness, rapid deployment, scalability, and accessibility across various levels of the healthcare systems [9]. Furthermore, the drawbacks of image-based detection mentioned above, this study focuses on the symptom-based detection method for early COVID-19 detection. However, despite their practical benefits, symptom-based detection methods generally exhibit lower accuracy compared to image-based methods. Recently, several works have explored the use of machine learning to improve the effectiveness of symptom-based detection, yielding encouraging yet still suboptimal results. Nevertheless, improving accuracy is still an ongoing issue. Therefore, this study aims to develop and evaluate multiple supervised machine learning models for early COVID-19 detection using symptom data, with the goal of identifying approaches that significantly improve classification performance.

The rest of paper is organized as follows. Section 2 reviews some related existing studies. In Section 3, the proposed method has been detailed including data collection and description, pre-processing techniques, and experiments and evaluation metrics. The experimental results and comparisons with traditional descriptors are described in Section 4. Finally, Section 5 provides a comprehensive summary of main results and recommendations.

2. Related Works

A global COVID-19 spread has been reported due to the pandemic situation. Many scholars have made predictions on the kind of algorithm that will be used to battle this virus. Both supervised and unsupervised machine learning techniques were utilized for this purpose. Generally, the result of related works highlights the effect of machine learning algorithms in supporting rapid and accurate COVID-19 early detection. In [10], the researchers have been employed different supervised machine learning techniques such as SVM, Random Forest, Decision Tree, KNN and others. They used symptom-based dataset for training issue. Random Forest and Decision Tree classifiers that reached 98.52% have obtained best achievement. Likewise, the scholars in [11] have been proposed ensemble classifier to build a detection model using patient symptoms. This work detected the presence of COVID-19 cases with overall accuracy 97.88%. Additionally, another work [12] employed Bayesian belief network algorithm to trained and tested on patient symptoms and it indicated a solid detection capability where the accuracy rate yield as high as 98%.

In the same context, another work [13] proposed a hybrid model combining the Fuzzy C-Mean clustering algorithm and the Back Propagation classification algorithm, and it was trained and tested using a public dataset. The proposed model was deployed to a mobile application to easily detect COVID-19 and it achieved accuracy up to 89%. In addition, other researchers [15] conducted a study aimed at developing a prediction method by using Lasso-logistic regression, and they collected a real-world clinical dataset. The findings have shown that the proposed model attained 85.9% accuracy at its best performance. Similarly, a recent work presented an automatic machine learning model using support vector machine binary classifier (SVM), as well as other models, that diagnosis the cases and detects severity. It is analysing a vast volume of data representing the patients' symptoms. SVM has accomplished the best detection accuracy of 96%. Notably, these samples of previous works show a wide range of detection accuracies following different datasets, feature selection, and employed models as

highlighted in figure 1. This difference confirms the necessity for more investigation to find out models that offer both a high accuracy rate and generalizability across diverse settings. Therefore, in this study, the Adaboost and bagging classifiers are employed, and the finest classifier for this dataset is chosen based on a comparative analysis of the Adaboost and bagging classifiers.

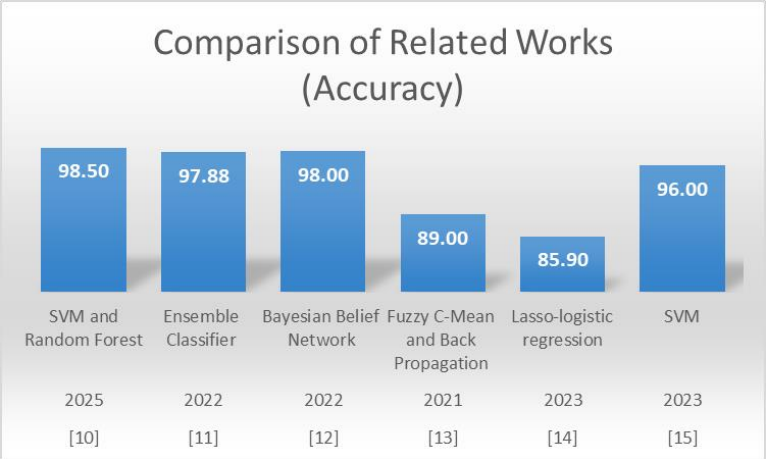

Figure 1. Comparison of Related works in term of Accuracy.

3. Method

Figure 2 depicts the technique of detecting COVID-19 based on machine learning classifiers. As depicted in figure 1, four steps need to be fulfilled to detect COVID-19: dataset gathering, pre-processing, training of machine learning classifier, and assessment of machine learning classifiers.

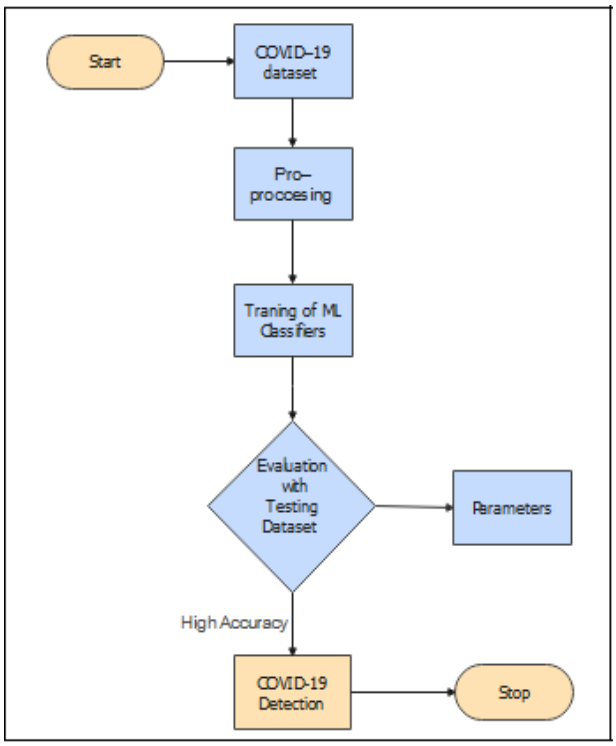

Figure 2. Model Framework for COVID-19 Detection.

3.1 Data Collection

We utilized the COVID-19 dataset published on Kaggle, which is freely accessible to all. This dataset comprises 20 Features and 5435 Instances. The two classes in this dataset are class one (Yes) with 4383 and class two (No) with 1051; there are no missing values in the data set. As can see in table1.

Table 1: Dataset Description

	Value	**Description**
Features	20	Breathing Problem, Fever, Dry Cough, Sore throat, Running Nose, Asthma, Chronic Lung Disease, Headache, Heart Disease, Diabetes, Hyper Tension, Fatigue, Gastrointestinal, Abroad travel, Contact with COVID Patient, Attended Large Gathering, Visited Public Exposed Places, Family working in Public Exposed Places, Wearing Masks, Sanitization from Market
Missing value	No	
Imbalance data	Yes	Imbalance class is No
Label	Yes or No	Binary classification
Class (Yes)	4383	
Class (No)	1051	
Instance	5435	

3.2 Pre-processing

The datasets pertaining to numerous real-world data science experiments comprise categorical variables. Text is the commonly used technique to store such properties. The significant aspect here is not the representation but the way the data is used for data analysis. Numerous machine-learning techniques can process categorical data without any need for modification. Therefore, Python tools comprising pandas and scikit-learn are used because they offer numerous processes that can be used to convert categorical data to a numeric representation. In this study used the label encoder to Encode target labels. These are transformers that are not intended to be used on features, only on targets.

3.3 Training of Machine Learning Classifiers

Bagging (Breiman, 1996) [16], i.e. bootstrap aggregating, is a meta-algorithm to enhance classification and regression models about stability and classification precision. Even though bagging is typically used for decision tree classifiers, it can be utilised with any kind of model. The concept of bagging is straightforward and interesting, and is based on random sampling with replacement. The random selection with replacement of N objects X^b=X_1^b,X_2^b,..,X_n^b from the set of N objects X=X_1,X_2,..,X_nis known as a bootstrap replicate. Thus, few of the objects might be signified in a fresh set once, twice or even more times and few of the objects might not be signified at all. Through a bootstrap replicate of the training sample set, one can evade or attain fewer 'outliers' in the bootstrap training set. Therefore, the bootstrap projection of the data distribution parameters is strong and more precise compared to the plug-in estimates typically utilised in classification rules [17].

Freund and Schapire [18] recommended AdaBoost in 1996. Because of sound generalisation ability, swift performance and low implementation intricacy, boosting has emerged as one of the most renowned and effectual tools of classification in computer vision [16] as well as pattern recognition. Let X signify the input data space and Y be the set of possible class labels. We take into consideration the case of two classes Y={-1,1 and presume that X=R^n. Our objective is to construct a mapping function F:X→Y which, given the feature vector x∈ X, computes the (correct) class label y. Furthermore, we consider the scenario where a set of labelled data for training is available:

$$(x\_1, y\_1), \ldots, (x\_n, y\_n); x\_i \in X; y\_i \in Y \quad (1)$$

Table 2: Parameter used for algorithms for AdaBoosting

Parameter	Range	Description
Base estimator	Class	Depend on decision tree classifier
N estimator	50	The maximum number of estimators at which boosting is terminated
Learning rate	1	Weight applied to each classifier at each boosting iteration
Algorithm	SAMME.R	
Random stat	None	Given at each base estimator at each boosting iteration

Table 3: Parameter used for algorithms for Bagging

Parameter	Range	Description
Base estimator	Class	Depend on decision tree classifier
N estimator	10	The number of best estimators in ensemble
Random stat	None	Given at each base estimator at each boosting iteration
Max sample	1	The number of samples to training each base estimator
Max feature	1	The number of features from train to each estimator

In the training stage, a learning algorithm utilises the training data to form a classification model (classifier). In the testing stage, the learned classifier is assessed by utilising the testing dataset to obtain the appropriate classification precision. If the right classification precision for the testing dataset is adequate, the trained classifier can be utilised. Else, few further processes can be conducted for enhancing the classification precision; for instance, parameters tuning or further processing of the data. For ensuring that the approach is effectual and competent, precision, recall, and F-score metrics coupled with the ROC are employed for evaluating classification performance.

$$Precision = \frac{True\ Postive}{True\ Postives+False\ Postives} \quad (2)$$

$$Recall = \frac{True\ Postive}{True\ Postives+False\ Negative} \quad (3)$$

4. Results Analysis and Discussion

In this section, the proposed was implemented in Python to assess system performance. The data was split into two sets having 70% and 30% data, respectively. The testing set (30%) was used to determine the accuracy pertaining to every class. The assessment metrics include overall accuracy, recall, F-measure, precision, and receiver operating characteristic curve (ROC). The results indicate that Bagging and AdaBoost classification was effectiveness.

Table 4: Number of training and testing belong to each class

Class	Training	Testing	Total
Yes	3080	1303	4383
No	723	328	1051
Total	3803	1631	5434

Table 5: The model accuracy for each algorithm

Algorithm	Precision	Recall	F-score	Accuracy
AdaBoost	100	100	100	100
	100	100	100	
Bagging	100	100	100	100
	100	100	100	

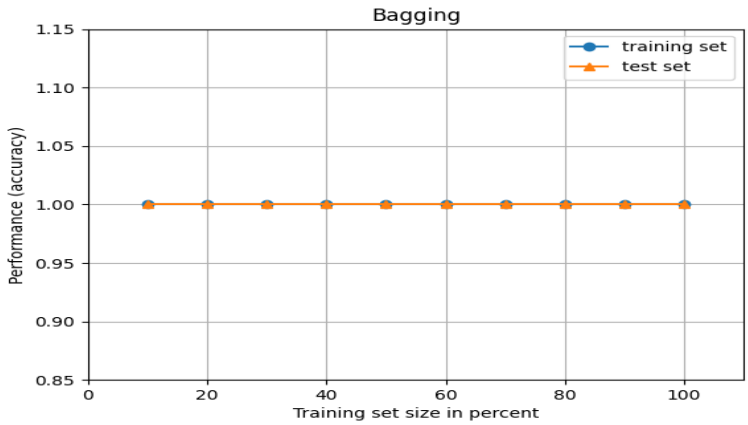

Figure 3. Learning carve for Bagging learning algorithm.

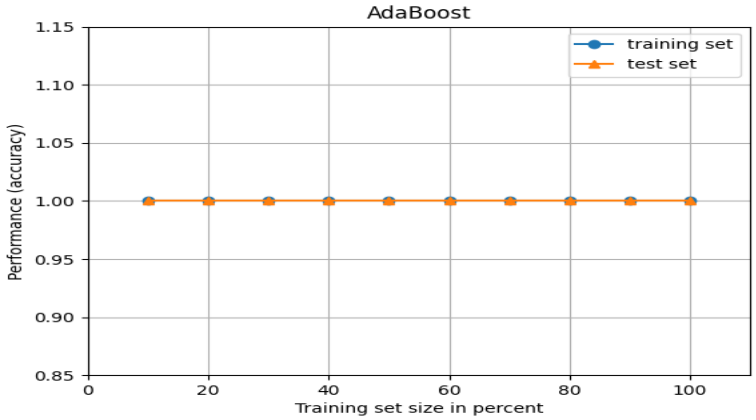

Figure 4. Learning carve for AdaBoost algorithm

The proposed was found to provide better as well as faster results. Furthermore, an increase in performance was noted when datasets were processed. As can see in figure 3 and figure 4, the performance of the training dataset was excellent.

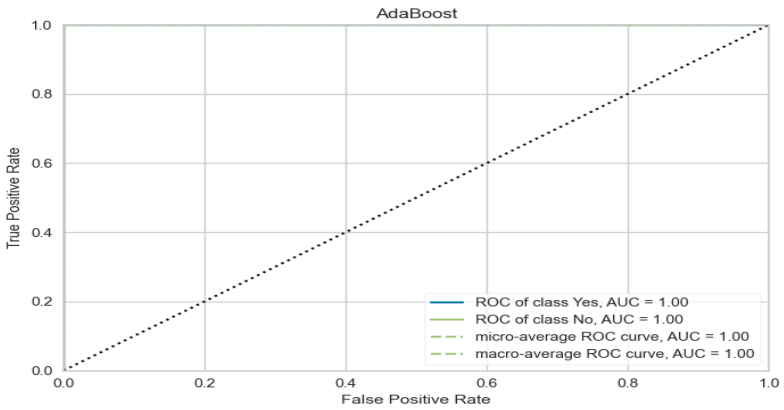

Figure 5. AdaBoost algorithm with ROC for each class.

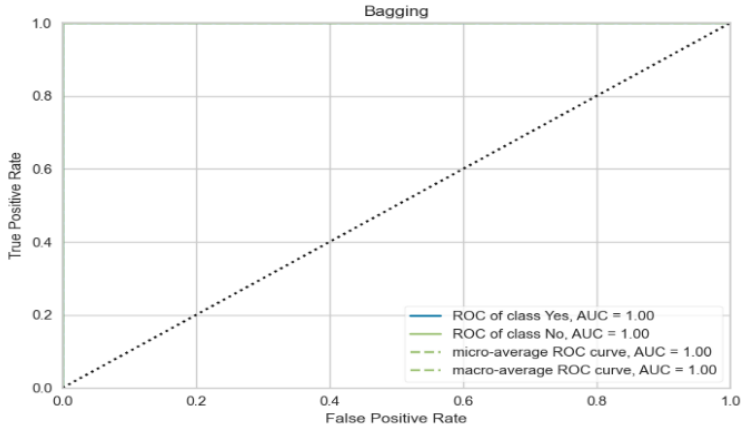

Figure 6. ROC and AUC for Bagging algorithm for each class.

The significant efficiency and efficacy of the proposed approach is depicted in figure 5 and figure 6, where the AUC of the 1.00 for AdaBoost and Bagging. The ideal scenario for the model's predictions, because When AUC=1, the classifier is able to correctly distinguish between the entire positive and the negative class points.

5. Conclusion

COVID-19 is a global health hazard and virus, which could infect a human through respiratory droplets formed from the infected individual's body. In this study, various machine learning classification algorithms like those that Bagging and AdaBoost are tested on COVID data and then comparatively scrutinised based on their training data performance measures. The COVID-19 dataset is classified using Bagging and AdaBoost as it attained the highest precision of 100%. Furthermore, we deployed precision, recall, and F-score metrics coupled with the ROC for evaluating classification performance.

References

[1] P. Shi, Y. Wang, S. Abbasi, and A. Wong, "COVID-Net Assistant: A deep learning-driven virtual assistant for COVID-19 symptom prediction and recommendation," arXiv preprint arXiv: 2211.11944, 2022.

[2] D. A. Kass, P. Duggal, and O. Cingolani, "Obesity could shift severe COVID-19 disease to younger ages," *Lancet*, vol. 395, no. 10236, pp. 1544–1545, 2020.

[3] Y. Raddad, A. Hasasneh, O. Abdallah, C. Rishmawi, and N. Qutob, “Integrating Statistical Methods and Machine Learning Techniques to Analyze and Classify COVID-19 Symptom Severity,” *Big Data and Cognitive Computing*, vol. 8, no. 12, p. 192, 2024.

[4] M. J. Binnicker, “Challenges and controversies to testing for COVID-19,” *Journal of Clinical Microbiology*, vol. 58, no. 11, Oct. 2020, doi: 10.1128/jcm.01695-20.

[5] M. A. Callejon-Leblic, “Loss of Smell and Taste Can Accurately Predict COVID-19 Infection: A Machine-Learning Approach,” *Journal of Clinical Medicine*, vol. 10, no. 4, p. 570, 2021, doi: 10.3390/jcm10040570.

[6] H. Mohammad-Rahimi, M. Nadimi, A. Ghalyanchi-Langeroudi, M. Taheri, and S. Ghafouri-Fard, "Application of machine learning in diagnosis of COVID-19 through X-ray and CT images: A scoping review," *Frontiers in Cardiovascular Medicine*, vol. 8, p. 638011, 2021, doi: 10.3389/fcvm.2021.638011.

[7] L. Wang and A. Wong, "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Scientific Reports*, vol. 10, no. 1, p. 1, 2020, doi: 10.1038/s41598-020-76550-z.

[8] E. Dritsas and M. Trigka, "Supervised machine learning models to identify early-stage symptoms of SARS-CoV-2," *Sensors*, vol. 23, no. 1, p. 40, 2022, doi: 10.3390/s23010040.

[9] M. Malik, "Determination of COVID-19 patients using machine learning algorithms," *Intelligent Automation & Soft Computing*, vol. 31, no. 1, pp. 207–222, 2022, doi: 10.32604/iasc.2022.018753.

[10] M. Pant and A. Vidyarthi, "COVID-19 prediction classification models comparison by patient's symptoms and medical history," in *Challenges in Information, Communication and Computing Technology*, CRC Press, 2025, pp. 381–386, doi: 10.1201/9781003559085-67.

[11] A. Kumari and A. K. Mehta, "Effective prediction of COVID-19 using supervised machine learning with ensemble modeling," in *Algorithms for Intelligent Systems*, pp. 537–547, 2022, doi: 10.1007/978-981-16-5747-4_45.

[12] A. Ekong, "Supervised machine learning model for effective classification of patients with COVID-19 symptoms based on Bayesian belief network," *Researchers Journal of Science and Technology*, vol. 2, no. 1, pp. 27–33, 2022.

[13] A. F. Al-Zubidi, N. F. Al-Bakri, R. K. Hasoun, S. H. Hashim, and H. T. Alrikabi, "Mobile application to detect COVID-19 pandemic by using classification techniques: Proposed system," *International Journal of Interactive Mobile Technologies*, vol. 15, no. 16, 2021, doi: 10.3991/ijim.v15i16.24195.

[14] Z. H. Arif and K. Cengiz, "Severity classification for COVID-19 infections based on LASSO-logistic regression model," *International Journal of Mathematics, Statistics, and Computer Science*, vol. 1, pp. 25–32, 2023, doi: 10.59543/ijmscs.v1i.7715.

[15] L. S. Suma, H. S. Anand, and S. S. Vinod Chandra, "Nature inspired optimization model for classification and severity prediction in COVID-19 clinical dataset," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 3, pp. 1699–1711, 2023, doi: 10.1007/s12652-021-03389-1.

[16] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123–140, 1996.

[17] G. Liang, X. Zhu, and C. Zhang, "An empirical study of bagging predictors for different learning algorithms," in *Proc. AAAI Conf. Artif. Intell*, vol. 25, no. 1, pp. 1802–1803, Aug. 2011.

[18] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. 13th Int. Conf. Machine Learning (ICML)*, vol. 96, pp. 148–156, July 1996.