Karaelmas Fen ve Müh. Derg. 13(2):329-342, 2023

Karaelmas Science and Engineering Journal
Journal home page: https://dergipark.org.tr/tr/pub/karaelmasfen
DOI: 10.7212/karaelmasfen.1363912

Research Article
Received / Geliş tarihi : 20.09.2023
Accepted / Kabul tarihi : 22.11.2023

Detection of COVID-19 Severity and Mortality from Blood Parameters by Ensemble Learning Methods

COVID-19 Şiddeti ve Mortalitesinin Kan Parametrelerinden Kolektif Öğrenme Yöntemleri ile Tespiti

Gizemnur Erol Doğan[1,*], Betül Uzbaş[2]

[1]Konya Technical University, Faculty of Engineering and Natural Sciences, Software Engineering Department, Konya, Türkiye
[2]Konya Technical University, Faculty of Engineering and Natural Sciences, Computer Engineering Department, Konya, Türkiye

Abstract

COVID-19 is a pandemic that causes a high rate of spread and Acute Respiratory Distress Syndrome (ARDS). Severe pneumonia in infected individuals has resulted in too many patients being admitted to the Intensive Care Unit (ICU). This has placed unprecedented pressure on health systems by exceeding capacities. It is essential to detect the prognosis of this disease so that the health systems can remain active and the conditions of the patients who need to be hospitalized in the ICU do not become critical. In this study, COVID-19 prognosis was detected by using ICU admission (COVID-19 SEVERITY) and COVID-19 related death (COVID-19 MORTALITY) datasets with Machine Learning (ML) methods. The missing data of the datasets were filled with K-Nearest Neighbor (KNN), and Min-Max normalization was performed. Datasets were divided three times into training and test sets, and the data were balanced with the Synthetic Minority Oversampling Technique (SMOTE). Then, classification was carried out using Ensemble Learning (EL) methods. For COVID-19 SEVERITY and COVID-19 MORTALITY, 89.54% and 97.25% accuracy were achieved with the Adaboost classifier, respectively. Successful and rapid COVID-19 prognosis detection with ML methods will help to use the ICU more efficiently and relieve the pressure on health systems.

Keywords: Classification models, hematological parameters, machine learning

Öz

COVID-19, yüksek yayılım hızına ve Akut Solunum Sıkıntısı Sendromuna (ARDS) neden bir pandemidir. Enfekte bireylerde gelişen şiddetli pnömoni, çok fazla hastanın Yoğun Bakım Ünitesine (ICU) kabul edilmesine neden olmuştur. Bu da, sağlık sistemlerinde kapasitelerin aşılarak benzeri görülmemiş bir baskı meydana getirmiştir. Sağlık sistemlerinin aktif kalabilmesi ve ICU'ya yatması gereken hastaların durumlarının kritikleşmemesi için bu hastalığın prognozunun belirlenmesi oldukça önemlidir. Bu çalışmada, ICU'ya kabul edilen (COVID-19 SEVERITY) ve COVID-19 nedeni ile ölen (COVID-19 MORTALITY) hastaların bilgilerini içeren veri setleri, Makine Öğrenmesi (ML) yöntemleri kullanılarak COVID-19 prognoz tespiti yapılmıştır. Veri setlerinde bulunan eksik veriler K-En Yakın Komşu (KNN) ile tamamlanmış ve Min-Max normalizasyonu yapılmıştır. Veri setleri, eğitim ve test setleri olarak bölünmüş ve veriler Sentetik Azınlık Aşırı Örnekleme Tekniği (SMOTE) ile dengelenmiştir. Ardından, Kolektif Öğrenme (EL) yöntemleri kullanılarak sınıflandırma gerçekleştirilmiştir. COVID-19 SEVERITY ve COVID-19 MORTALITY için Adaboost sınıflandırıcısı ile sırasıyla %89.54 ve %97.25 başarı elde edilmiştir. ML yöntemleri ile COVID-19 prognozunun başarılı ve hızlı bir şekilde tespit edilmesi, ICU'yu daha verimli kullanmaya ve sağlık sistemlerinin üzerindeki baskıyı hafifletmeye yardımcı olacaktır.

Anahtar Kelimeler: Hematolojik parametreler, makine öğrenmesi, sınıflandırma modelleri

---

*Corresponding author: gerol@ktun.edu.tr

Gizemnur Erol Doğan orcid.org/0000-0001-9347-9775
Betül Uzbaş orcid.org/0000-0002-0255-5988

This work is licensed by "Creative Commons Attribution-NonCommercial-4.0 International (CC)".

1. Introduction

COVID-19 is the new viral respiratory disease that causes ARDS. Since its first appearance in 2019, it has been detected in more than 200 countries in the world within 4 months and has taken the world under its influence (Brinati et al. 2020). Because of this global epidemic, this disease has been defined as a pandemic. COVID-19 has caused severe effects on people with chronic conditions such as diabetes, cancer, and heart disease, causing permanent illness and even death (WHO 2020, WHO 2022). Serious illness occurred in about 15% of infected patients, and about 6% of them reached critical levels (Wang et al. 2020). Too many people were admitted to the ICU due to severe pneumonia and ARDS (Pasquier et al. 2021). In this pandemic situation where emergency departments and ICUs are overcrowded, early detection of patients who need intensive care is very important for the proper functioning of health systems (Negant et al. 2020). In addition, one of the most important epidemiological criteria to be detected in infectious diseases is the mortality of the disease (Ghani et al. 2005). Detecting the mortality of COVID-19, which has such a high contagiousness and ICU admission rate, makes it easier to determine the prognosis of the patients and to follow the pandemic surveillance.

In addition to biasing individual health, COVID-19 has also affected countries in terms of economic, social, and health infrastructure systems (Dey et al. 2020, Fong et al. 2020). For example, in northern Italy, with 1,381 patients admitted to the ICU in the first wave of COVID-19, healthcare systems have experienced unprecedented pressure, exceeding 133% of capacities (Grasselli et al. 2020). For this reason, artificial intelligence-based studies have been carried out to support healthcare professionals in many tasks to slow down this epidemic and control infrastructure systems (Chen et al. 2020, Erol Doğan and Uzbaş 2023, Kumar et al. 2023, Yang et al. 2020). While many studies have been conducted for the diagnosis of COVID-19-infected individuals, the development of prognostic models for ICU admission and mortality estimation has lagged (Cabitza et al. 2021; Wynants et al. 2020). However, health systems in many countries have come under a great burden with COVID-19. In addition, critically ill patients admitted to the ICU have been found to have a higher mortality rate than normal COVID-19 patients (Moore and June 2020, Willyard 2020). Therefore, early diagnosis of the prognosis and lethality of this disease is very important both for the health systems to remain active and for the condition of infected individuals not to become critical.

When the studies in the literature are examined, Podder et al. (2021) have developed various ML models for patients admitted to the ICU or semi-ICU. RF and voting classifiers (RF, Logistic Regression (LR), Support Vector Machines (SVM)) achieved the best performance in classifying patients who should be admitted to ICU and semi-ICU. Voting classifiers achieved 98% success and were presented as the most successful result. Rodriguez-Nava et al. (2020) determined clinical AUC scores to determine ICU admission and mortality, and ultimately obtained AUC scores of 0.761 and 0.781 for ICU admission and mortality, respectively. Kong et al. (2020) determined the importance of biomarkers with RF by taking blood samples from infected individuals every 3-7 days to determine the severity of COVID-19 and evaluated the diagnostic accuracy of the most important proteins with the Recipient Operating Characteristic (ROC) curve. Famiglini et al. (2021) developed 3 prognostic ML models with blood parameters of individuals for the admission of COVID-19 patients to the ICU. These models take as input the dataset BIG-DATA-COVID19, which includes blood parameters of COVID-19 patients, and predict whether they need to be transferred to the intensive care unit within 5 days. They obtained an AUC value of 0.880 with the ensemble model obtained with the XGBoost (XGB), RF, and LR classifiers. Alabad et al. (2022) developed an ML model estimating the length of ICU stay of COVID-19 patients in eastern Saudi Arabia. In this model, they achieved 94.16% success by using the RF classifier. Moulaei et al. (2022) compared several ML models that predicted COVID-19 mortality using initial hospital admission data from 1500 COVID-19 patients. As a result, the RF model was chosen as the most successful model to predict COVID-19 mortality with 95.03% success and 0.990 AUC. Elshennawy et al. (2022) developed 3 Deep Learning (DL) models that predict the mortality risk of infected individuals. The first of these models is CV-CNN, which was developed with the Convolutional Neural Network (CNN). CV-CNN was developed using the clinical dataset of 12020 patients, and this clinical dataset was also used in the second model, CV-LSTM + CNN, which was developed by combining Long Short-Term Memory (LSTM) and CNN. The third model, IMG-CNN, is a CNN model and this model estimates mortality using the clinical dataset used in the first two models converted to image data. As a result, the IMG-CNN model performed better than the other two models with a success of 94.14% and an AUC of 93.70%.

In this study, a dataset containing blood parameters of COVID-19 patients and 2 classes of information was used. The first class is the Severity class, which represents COVID-19 patients admitted and not admitted to the ICU. The other is the Dead class, which represents patients who died and did not die due to COVID-19. The general diagram of this study is presented in Figure 1. First, certain data preprocessing was applied to the dataset: KNN was used to fill in missing data (Erol et al., 2022) and Min-Max normalization was performed. Two sub-datasets were obtained from the filled and normalized datasets. The first sub-dataset, COVID-19 SEVERITY, was obtained by dividing the blood parameter characteristics and Severity class in the dataset. The other sub-dataset, COVID-19 MORTALITY, was obtained by dividing the blood parameter characteristics and Dead class. Each dataset was then divided three times into an 80% training set and a 20% test set. The data in the training sets were balanced with SMOTE. ML models were developed with the Severity and Mortality datasets and the EL method, Bagging, RF, and Adaboost classifiers.

There are 2 main limitations to this study. These are:

1. Blood values are influenced by genetic factors (Mahaney et al., 2005). Therefore, COVID-19 studies with blood parameters will be active for certain breeds.
2. Missing and unbalanced data is a known problem in medical data (Rahman and Davis, 2013; Hu et al., 2017). While missing blood parameter data can cause bias in statistical analyses, unbalanced data distribution between classes can lead to both bias in statistical results and overfitting in classification.

Considering these 2 limitations, the dataset in the study should be well analyzed. The data preprocessing steps to be applied to the dataset in the study directly affect the classification accuracy. Therefore, the completion of missing data and the elimination of data imbalance are the main steps of the study.

As a result of this study, severity and mortality detection of infected individuals was performed with a single dataset containing the blood parameters of COVID-19 patients, whether they were admitted to the ICU, and whether they died as a result of the disease. In addition, by applying appropriate data preprocessing to a single dataset that has all of this information, it has been shown that EL methods produce very successful results in determining the prognosis of the disease and keeping health systems active.

2. Material and Methods

2.1. Dataset

The BIGDATA-COVID19 (Famiglini et al. 2021) dataset used in this study includes age, sex, and routine blood test results of 1218 patients admitted to the COVID-19 emergency department of San Raffaele Hospital in Milan (Italy) between February 19 and May 31, 2020. For those hospitalized for at least 24 hours, blood samples were collected every day of hospitalization, resulting in 4995 observations in total. There are 2 classes in the dataset, Severity and Dead. The Severity class contains information on whether patients were admitted to the ICU within 5 days of hospitalization. The Dead class contains information on whether individuals died due to COVID-19 during this period. The features and classes in the dataset are given in Table 1.

The Severity class in the BIGDATA-COVID19 (Famiglini et al. 2021) dataset represents COVID-19 patients who were and were not admitted to the ICU, and the Dead class represents patients who died and did not die due to COVID-19. After applying data preprocessing to the data-

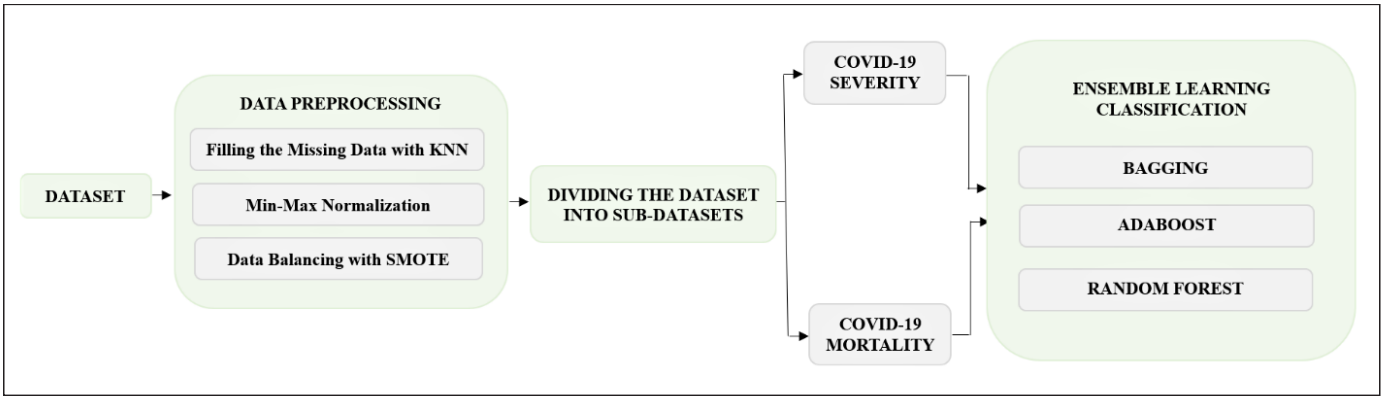


Figure 1. General diagram of the study.

Table 1. BIGDATA-COVID19 Dataset

Parameter		Acronym	Unit of Measure
Feature 1	Age	-	Years
Feature 2	Sex	-	Female-Male
Feature 3	Mean Corpuscular Volume	MCV	fL
Feature 4	Mean Platelet Volume	MPV	fL
Feature 5	Mean Corpuscular Hemoglobin	MCH	Pg/Cell
Feature 6	Mean Corpuscular Hemoglobin Concentration	MCHC	g Hb/dL
Feature 7	Neutrophils Count	NE - NET	% - 10^9/L
Feature 8	Lymphocytes Count	LY - LYT	% - 10^9/L
Feature 9	Monocytes Count	MO - MOT	% - 10^9/L
Feature 10	Eosinophils Count	EO - EOT	% - 10^9/L
Feature 11	Basophils Count	BA - BAT	% - 10^9/L
Feature 12	Hematocrit	HCT	%
Feature 13	Platelets	PLT	10^9/L
Feature 14	Red Blood Cell	RBC	10^{12}/L
Feature 15	White Blood Cells	WBC	10^9/L
Feature 16	Hemoglobin	HGB	g/dL
Feature 17	Erythrocyte Distribution Width	RDW	CV%
Class 1	Severity	-	(0,1)
Class 2	Dead	-	(0,1)

set, 2 sub-datasets were obtained. The first sub-dataset, COVID-19 SEVERITY, was obtained by dividing the blood parameter properties and Severity class in the dataset. The other sub-dataset, COVID-19 MORTALITY, was obtained by dividing the blood parameter characteristics and Dead class.

2.2. Data Preprocessing

In this study, the BIGDATA-COVID19 (Famiglini et al. 2021) dataset was first subjected to certain data preprocessing: Missing data were filled with the KNN (Erol et al. 2022) and Min-Max normalization was performed in the range [0, 1]. Two sub-datasets were obtained from the filled and normalized datasets. The first sub-dataset, COVID-19 SEVERITY, was obtained by dividing the blood parameter characteristics and Severity class in the dataset. The other sub-dataset, COVID-19 MORTALITY, was obtained by dividing the blood parameter characteristics and Dead class. Each dataset was then divided three times into an 80% training set and a 20% test set. Each dataset has 3996 data in the training set and 999 data in the test set. In the COVID-19 SEVERITY training set, the number of ICU admissions and non-admissions are 1087 and 2909, respectively. In the COVID-19 MORTALITY training set, the number of deaths and non-deaths due to COVID-19 are 124 and 3872, respectively. This imbalance in the datasets was balanced with SMOTE and the minority classes in the COVID-19 SEVERITY and COVID-19 MORTALITY training sets were completed to 2909 and 3872, respectively.

2.2.1. Missing Data Filling with KNN Algorithm

Missing data from the dataset affects the performance of ML models and prevents them from producing accurate predictions. KNN fills the missing data in the dataset based on the average distance to its k nearest neighbors. Euclidean, Mahapolnis, and Manhattan distance metrics are used for distance calculation (Idri et al. 2016).

2.2.2. Min-Max Normalization

Min-Max normalization normalizes the data so that the smallest value is 0 and the largest value is 1 and spreads all data in the dataset over the range [0, 1]. This method ensures that all data is kept on a positive scale. Min-Max normalization is presented in Equation 1. According to Equation 1, X_{norm} represents the rescaled data and X_i represents the original data. With Min-Max normalization, each data X_i in the dataset is subtracted from the minimum value in the relevant feature column, and the X_{norm} value is obtained by dividing the maximum value of the relevant feature column by the difference of the minimum value.

$$X_{norm} = \frac{X_i - \min(X)}{\max(X) - \min(X)} \tag{1}$$

2.2.3. Data Balancing with SMOTE

When the number of data in classes in a dataset is not approximately equal, this is called imbalanced data. Unbalanced data produces success according to the majority class ignoring the minority class and giving misleading results for ML models. Therefore, many algorithms have been developed to balance the data in datasets. One of these algorithms, SMOTE, developed by Chawla et al. (2002), generates synthetic samples based on the k nearest neighbors of the samples examined in the minority class. The rationale of the method is to help the classifier generalize to the test data without disturbing the natural structure of the dataset by generating synthetic samples instead of copying the existing observations in the minority class (Douzas et al. 2018, Fernandez et al. 2018). This method overcomes the overfitting that occurs by oversampling through replication.

2.3. Machine Learning

ML learns meaningful relationships and patterns from the observations in the training set and predicts the information learned from the observations in the test set that it has never seen (Bishop 2006). In this way, it assimilates the datasets presented to it and aims to perform the classification task directly without instructions. There are many ML classifier models. In this study, EL models were used, which have been successful in many fields and outperformed single models (Anwar et al. 2014, Prusa et al. 2015, Shahzad and Lavesson 2013).

2.4. Ensemble Learning

EL is a classification methodology that refers to the use of multiple models together to create a stronger model rather than using a single classification model. Model diversity increases success and reduces the risk of overfitting (Mohammed and Kora 2023). In the literature, there are many studies on EL methods and different EL methods have been presented according to how the models are combined and trained (Dong et al. 2020, Sagi and Rokach 2018). The commonly used ones are as follows: Bagging, RF, and Adaboost.

2.4.1. Bagging Classification Algorithm

Bagging is a bootstrap algorithm developed by Breiman (1996). This algorithm randomly selects sub-samples from the dataset and presents each of these samples to a classifier model to form an ensemble of classifiers. Each of this ensemble of classifiers produces a prediction. The final class prediction is produced by combining all the predictions. In the combination of predictions, the average is taken for regression trees, while in classification trees the results are determined by multiple voting. Bagging generates random sub-samples of n samples from a dataset of n samples and each selected sample is put back into the dataset. The working principle of Bagging is presented in Figure 2.

2.4.2. Random Forest Classification Algorithm

RF is a very fast and robust classifier for overlearning (Breiman 2001). RF creates an ensemble forest with multiple trees and trains each tree in this forest on a different observation, generating various models, and then combines these models to obtain a final prediction, maximizing the classification success. The working principle of RF is presented in Figure 3.

2.4.3. Adaboost Classification Algorithm

Adaboost, developed by Freund and Schapire (1999), is a classification algorithm that works by boosting. Boosting is the process of combining several weak classifier models to create a strong classifier model. In the Adaboost algorithm, which is one of the boosting methods, the training set is first trained with a weak classifier, and then a training model is created again by giving more priority to the samples that are incorrectly predicted as a result of this training, that is, by increasing their weights. The model is continued by training the model so that the output of the weak classifier is the input of the other classifier, and finally, the results are combined to produce the final classifier model and class prediction. The working principle of Adaboost is presented in Figure 4.

2.5. Classification Performance Metrics

Classifier models that take the dataset as input provide information about their accuracy with performance measures in various criteria. These performance measures on various criteria are based on an error table called the Confusion Ma-

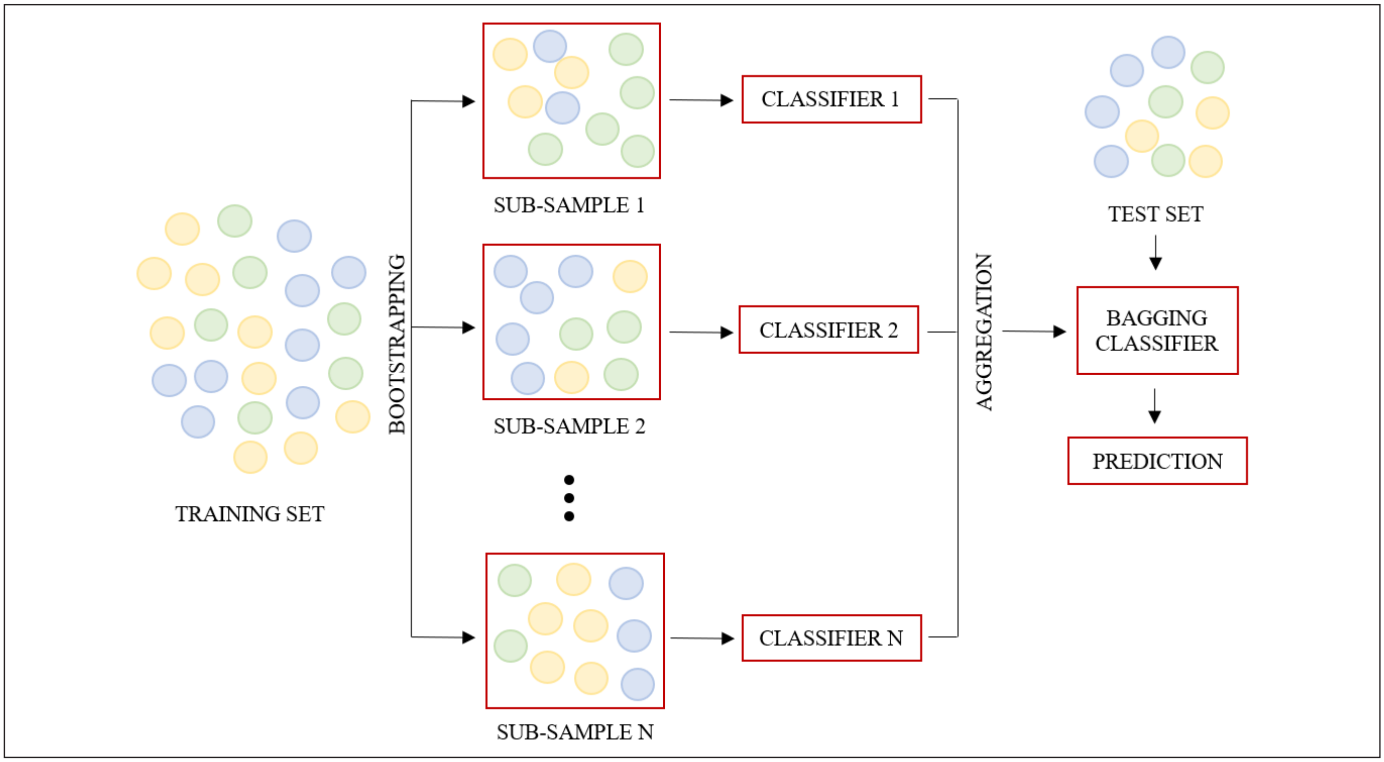


Figure 2. Working principle of bagging classification algorithm.

trix (CM) (Japkowicz and Shah 2011). The CM is a special two-dimensional table that presents the number of correct and incorrect data produced by the predicted results versus the actual results. The basic structure of the CM is presented in Figure 5, based on a binary classification model.

Each row of the matrix represents the instances in the predicted class and each column represents the actual class of those instances. As shown in Figure 5, for a two-class classifier model with positive and negative labels:

True- Positive (TP): This represents a dataset with a positive class label correctly predicted by the classifier.

True -Negative (TN): This represents the dataset with a negative class label correctly predicted by the classifier.

False-Positive (FP): This represents the dataset with a positive class label that is incorrectly predicted by the classifier.

False - Negative (FN): This represents a dataset with a negative class label that is incorrectly predicted by the classifier.

Some basic metrics have been established to measure the performance of a classifier using CM. These are Accuracy, Recall, Precision, F-Score, and Area under the ROC Curve (AUC).

2.5.1. Accuracy

Accuracy presents the rate at which the classifier model built using the training set correctly classifies the data in the test set. This metric is calculated as the ratio of correctly predicted areas in the model to the total dataset, as presented in Equation 2. This helps to evaluate the overall performance of the classifier by treating all class labels equally. The accuracy metric alone is not enough to evaluate a classifier model, especially for imbalanced datasets. Therefore, other performance metrics should also be used when evaluating the performance of a classifier model (Dai et al. 2022).

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

2.5.2. Recall and Precision

Precision and Recall are metrics used to evaluate information extraction by measuring the test performance of a classifier and are suitable for applications such as artificial neural networks that perform information extraction (Abeel et al. 2009, Huang and Bader 2009). The standard to aim for when building a classifier model is a high TP and low FP ratio. The ability of the classifier to detect positive classes is called recall. In other words, recall is a metric that shows

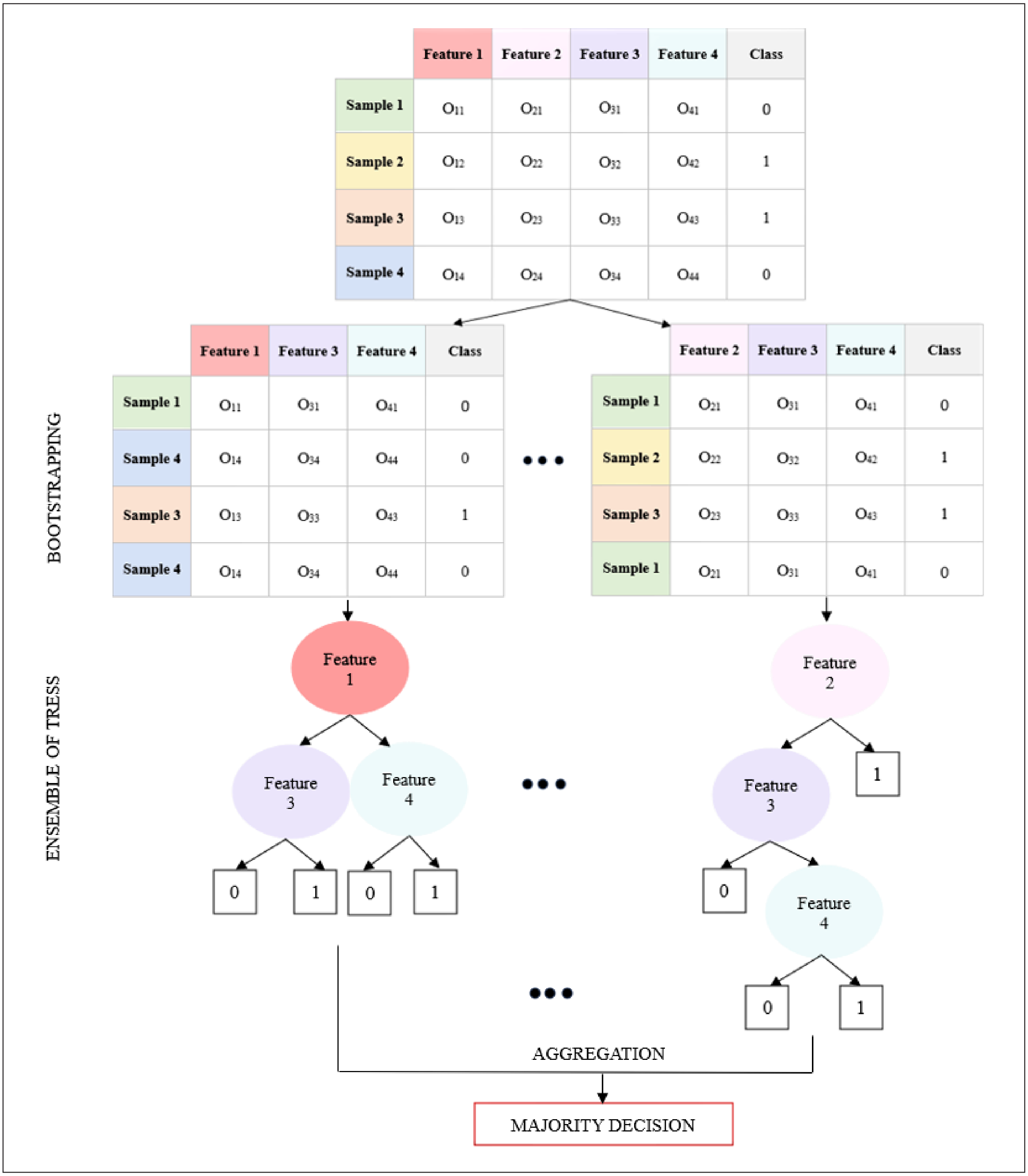


Figure 3. Working principle of random forest classification algorithm.

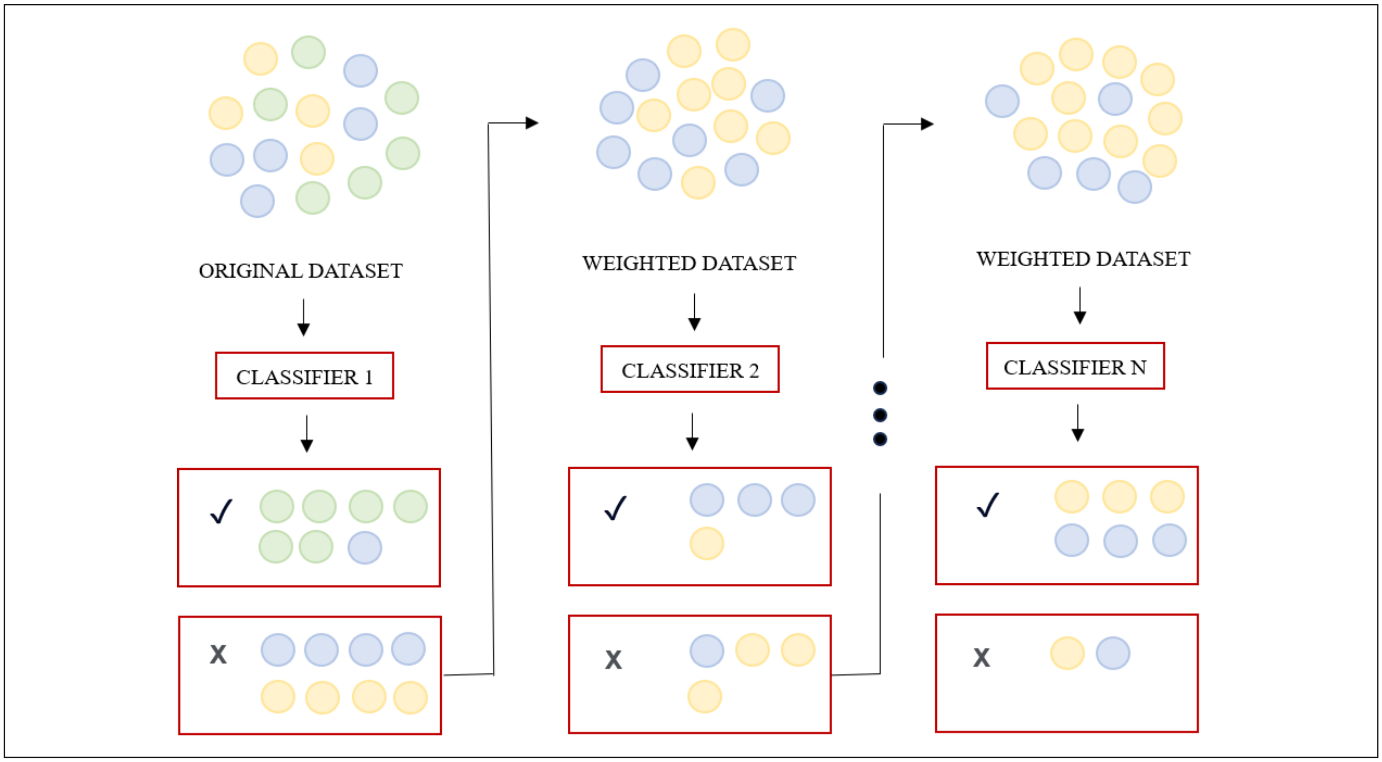
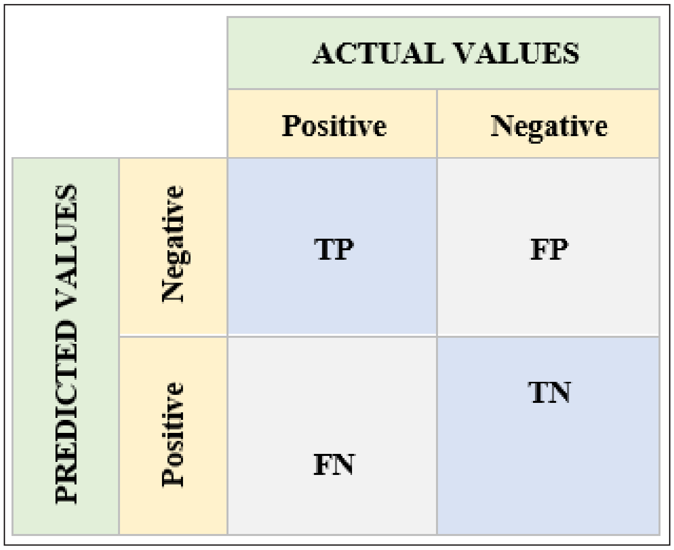


Figure 4. Working principle of adaboost classification algorithm.



		ACTUAL VALUES	
		Positive	**Negative**
PREDICTED VALUES	**Negative**	**TP**	**FP**
	Positive	**FN**	**TN**

Figure 5. Confusion matrix.

how much of the data that is known to belong to a positive class is labeled as a positive class after classification. The recall of a classifier is the number of TPs obtained from the CM divided by the number of instances in the true positive class. The TP ratio, the recall performance metric, is given in Equation 3.

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

Precision is a metric that indicates how many of the positively labeled values are correctly labeled positive. The precision metric is calculated as the ratio of the number of correctly labeled positive data to the number of data labeled as the predicted positive class and its formulation is given in Equation 4.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

2.5.3. F-Score

The Precision and Recall measures alone do not provide a meaningful comparison result for the classifier models. The F-Score, which evaluates the two measures together, is the harmonic mean of the precision and sensitivity values. The F-Score is commonly used as F1, F0.5, and F2 and takes values between 0 and 1. A classifier that makes accurate predictions is expected to have an F-Score value close to 1 (Japkowicz 2011). The formulation of the F-Score is given in Equation 5.

$$F_{\beta} - Score = \frac{(1 + \beta^2) \times (Precision \times Recall)}{\beta^2 + Precision \times Recall} \tag{5}$$

2.5.4. The Area Under the ROC Curve

The ROC curve is a graphical representation of the relationship between recall and specificity. On the x-axis of the ROC graph is the FP ratio, the specificity performance metric, while on the y-axis is the TP ratio, the sensitivity performance metric. Recall and specificity values are calculated for different threshold values and each point on the ROC curve represents the recall-specificity pair corresponding to a given decision threshold. AUC is used to test the performance of classifiers between 0 and 1. A value of AUC approaching 1 indicates that the classifier is making accurate predictions, a value of 0.5 indicates that the classifier is making random predictions, and values below 0.5 indicate that the classifier is not working correctly (Hulley et al. 2001).

3. Results and Discussion

In this study, we used the BIGDATA-COVID-19 (Famiglini et al. 2021) dataset, which contains blood parameters of COVID-19 patients and 2 classes of information. The classes in the dataset are as follows: Severity class representing COVID-19 patients who were and were not admitted to the ICU and Dead class representing patients who died and did not die due to COVID-19. First, certain data preprocessing was applied to the dataset: Missing data were filled with the KNN (Erol et al. 2022) and Min-Max normalization was performed in the range [0, 1]. Two sub-datasets were obtained from the filled and normalized datasets. The first sub-dataset, COVID-19 SEVERITY, was obtained by dividing the blood parameter characteristics and Severity class in the dataset. The other sub-dataset, COVID-19 MORTALITY, was obtained by dividing the blood parameter characteristics and Dead class. Each dataset was then divided three times into an 80% training set and a 20% test set. The data in the training sets were balanced with SMOTE. Training models were developed by applying 10-fold cross-validation to balanced training sets. Bagging, RF, and Adaboost classifiers were used. Kernel classifier RF was chosen for Bagging and Adaboost. The numFeatures parameter, which sets the number of randomly selected features for RF and kernel RF, was chosen as 4. Testing procedures were performed by applying test sets to the developed training models. The study was coded in Python 3.9.7 and classification processes were carried out with Weka 3.8.4.

Classification performances for COVID-19 SEVERITY and COVID-19 MORTALITY are presented in Table 2 with average Accuracy, Precision, Recall, F-Score, and AUC measurements of 3 sub-datasets obtained from each dataset.

As can be seen in Table 2, the highest success with the COVID-19 SEVERITY dataset was 89.54% with Adaboost. Bagging and RF achieved 89.11% and 89.41%, respectively. For the COVID-19 MORTALITY dataset, the highest success was 97.25% with Adaboost. Bagging and Adaboost achieved 96.69% and 96.89%, respectively. As a result, it was revealed that Adaboost, one of the EL methods, produced the most successful results in detecting both COVID-19 patients who were admitted to the ICU and those who were not, and individuals who died and did not die due to COVID-19.

The AUC performance criterion provides the optimal threshold for the separation of positive and negative classes and plays a central role in discriminating between populations (Hajian-Tilaki 2013, Metz 1978). Adaboost was selected as the most successful model for COVID-19 SEVERITY and COVID-19 MORTALITY. Adaboost, which was selected as the most successful model for COVID-19 SEVERITY and COVID-19 MORTALITY, produced 0.949 and 0.953 AUC for these two datasets, respectively. In contrast, the most successful AUC value for COVID-19 MORTALITY is 0.959 with Bagging.

Table 2. Classification results.

Classification Results							
Classifier	**Dataset**	**Accuracy**	**Precision**	**Recall**	**F-Score**	**AUC**	**TP Rate**
Bagging	**COVID-19 Severity**	89.11%	89.30%	89.13%	89.20%	0.948	89.13%
	COVID-19 Mortality	96.69%	97.26%	96.70%	96.93%	0.959	96.70%
Random Forest	**COVID-19 Severity**	89.41%	89.46%	89.43%	89.43%	0.948	89.40%
	COVID-19 Mortality	96.89%	97.13%	96.90%	97.0%	0.953	96.90%
Adaboost	**COVID-19 Severity**	**89.54%**	89.60%	89.56%	89.56%	**0.949**	89.56%
	COVID-19 Mortality	**97.25%**	97.33%	97.26%	97.30%	**0.953**	97.26%

In this study, the imbalance in the dataset is eliminated in the training set to ensure unbiased training and the accuracy of the models is checked with TP Rate in the test model. The TP Rate performance criterion provides the average number of true positives predicted for the true positive class. As seen in Table 2, the most successful TP Rate for COVID-19 SEVERITY and COVID-19 MORTALITY is 89.56% and 97.26% with Adaboost, respectively. Despite the unbalanced data distribution in the test set, thanks to the data balancing applied in the training model, the models perform their detection without bias. As a result, the blood parameters of COVID-19 patients can successfully determine whether the patients will be admitted to the ICU and whether they will die due to the disease.

The comparison of this study with other COVID-19 severity and mortality detection studies conducted in the literature is presented in Table 3.

When Table 3 is examined, Rodríguez-Nava et al. (2020) examined the blood parameters of 313 COVID-19 patients in the Northwestern United States and determined the patients' clinical AUC scores for determining ICU admission and deaths. As a result, the AUC for ICU admission and mortality in COVID-19 was determined to be 0.761 and 0.781, respectively. In this study, AUCs of 0.949 and 0.953 were obtained for COVID-19 SEVERITY and COVID-19 MORTALITY, respectively. This shows that the ML detection models developed in this study achieved higher AUC values determined by Rodriguez-Nava et al. (2020) for COVID-19 severity and mortality. In the study conducted by Podder et al. (2021), a dataset containing 111 blood samples of 5644 COVID-19 patients was used to predict whether individuals would be admitted to ICU or semi-ICU. Missing data in the dataset were completed with median values and then feature selection was made with Extra Trees (ET). They developed various ML models for patients admitted to ICU or semi-ICU with 10 selected

Table 3. Comparison of this study with the literature

Study	Dataset	Purpose	Result
Rodriguez-Nava et al. (2020)	Blood samples of 313 COVID-19 patients in the Northwestern United States	Determining clinical AUC scores for determining intensive care unit admissions and deaths.	AUC of 0.761 and 0.781 for ICU admission and mortality, respectively
Podder et al. (2021)	111 blood samples from 5644 patients at Hospital Israelita Albert Einstein, Brazil.	Classifying ICU and semi-ICU admission by selecting features from the blood parameters of COVID-19 patients.	98% accuracies with voting classifiers
Famiglini et al. (2021)	4995 blood samples taken from 1218 COVID-19 patients (BIGDATA-COVID19)	Predicting whether patients need to be transferred to the ICU within 5 days by blood parameters.	0.880 AUC with the ensemble model obtained with XGB, RF and LR classifiers
Moulaei et al. (2022)	First hospital admission data of 1500 (1386 survivors and 144 deaths) COVID-19 patients	Comparing various ML models predicting COVID-19 mortality	95.03% accuracy and 0.990 AUC with RF
This Study	4995 blood samples taken from 1218 COVID-19 patients (BIGDATA-COVID19)	Determining the disease severity and mortality with a single blood dataset containing information about whether COVID-19 patients were admitted to the ICU and whether they died	89.54% accuracy and 0.949 AUC and 97.25% accuracy and 0.953 AUC with Adaboost for COVID-19 SEVERITY and COVID-19 MORTALITY, respectively.

features. Voting classifiers (RF, Logistic Regression (LR), Support Vector Machines (SVM)) showed the best performance, achieving 98% success in classifying patients who should be admitted to intensive care and semi-intensive care. Famiglini et al. (2021) developed an ML model that predicts whether infected individuals should be transferred to the ICU within 5 days using the BIG DATA-COVID19 dataset used in this study. As a result, they achieved an AUC value of 0.880 with the ensemble classifier model consisting of XGBoost (XGB) RF and LR. In this study, an AUC of 0.949 was obtained with COVID-19 SEVERITY representing transfer to ICU using the same dataset. Although data imbalance was resolved with SMOTE in both studies, differences in pre-processing directly affected success. The success of the data preprocessing combination presented in the study was demonstrated by obtaining a higher AUC value with the same dataset. In this study, in addition to the prediction model of transfer to ICU, a prediction model of whether patients will die due to disease was also developed. In the study conducted by Moulaei et al. (2022), they developed ML models that predict COVID-19 mortality by applying certain data preprocessing and feature selection to the data of 1500 COVID-19 patients, 1386 of whom were alive and 144 of whom were dead. Among the models, RF was chosen as the most successful model with 95.03% accuracy and 0.990 AUC value. In this study, higher accuracy (97.25%) and AUC (0.953) for COVID-19 MORTALITY were achieved with Adaboost.

4. Conclusion and Suggestions

COVID-19 is a pandemic respiratory disease that targets the lungs and can cause lung damage and death in severe cases. The lethality of this disease varies between countries (Asch et al. 2021, Strålin et al. 2021, Strålin et al. 2022), and workload in hospitals affects the lethality of the disease (Fegin et al. 2022, Strålin et al. 2022). About 15% of infected patients developed severe illnesses and about 6% of them reached critical status (Wang et al. 2020). Although many people were admitted to the ICU due to severe pneumonia and ARDS, 2.9% of those with severe illness had no abnormalities on initial admission (Guan et al. 2020, Pasquier et al. 2021). This has led to serious confusion in healthcare systems as COVID-19 patients' conditions worsened and the level of ICU need could not be determined.

In this study, a dataset containing blood parameters of COVID-19 patients and 2 classes of information was used first, certain data preprocessing was applied to the dataset: KNN was used to fill in missing data (Erol et al. 2022) and Min-Max normalization was performed. Two sub-datasets were obtained from the filled and normalized datasets. The first sub-dataset, COVID-19 SEVERITY, was obtained by dividing the blood parameter characteristics and Severity class in the dataset. The other sub-dataset, COVID-19 MORTALITY, was obtained by dividing the blood parameter characteristics and Dead class. Each dataset was then divided three times into an 80% training set and a 20% test set. The data in the training sets were balanced with SMOTE. Training models were developed by applying 10-fold cross-validation to balanced training sets. Bagging, RF, and Adaboost classifiers were used. Testing procedures were performed by applying test sets to the developed training models. As a result of the classification, the most successful model was determined to be the Adaboost classifier. With Adaboost, the highest success was obtained with 89.54% for COVID-19 SEVERITY and for COVID-19 MORTALITY with 97.25%. In addition, the AUC values produced by this classifier for the most successful result are 0.949 for COVID-19 SEVERITY and 0.953 for COVID-19 MORTALITY.

As a result of this study, it has been demonstrated that severity and mortality detection of infected individuals can be performed with high success with a single dataset that includes blood parameters of COVID-19 patients, information on whether they were admitted to the ICU and whether they died as a result of the disease. Göreke et al. (2021) revealed that genetic differences are directly related to the ML models developed for COVID-19. In future studies, severity and mortality models can be developed based on the genetic factor with the knowledge of the blood parameters of many COVID-19 patients of different races, whether they were admitted to the ICU, and whether they died due to COVID-19. In addition, this study can be applied to many infectious diseases.

Acknowledgment: We want to thank the referees who carefully read our article and evaluated it in the light of valuable information and the attentive editor.

Author contribution: Author Gizemnur Erol Doğan: conceived the original idea, carried out software applications, and wrote the main manuscript text. Author Betül Uzbaş: developed the theory, analyzed the results, and edited the final version of the manuscript. All authors discussed the results and contributed to the final manuscript.

Ethics committee approval: This study does not require ethics committee permission or any special permission.

5. References

Abeel,T., Van de Peer, Y., Saeys Y. 2009. Toward a gold standard for promoter prediction evaluation. Bioinformatics, 25(12): 313–320. Doi: 10.1093/bioinformatics/btp191

Alabbad, D.A., Almuhaideb, A.M., Alsunaidi, S.J., Alqudaihi, K.S. 2022. Machine learning model for predicting the length of stay in the intensive care unit for Covid-19 patients in the eastern province of Saudi Arabia. Informatics in Medicine Unlocked, 30:100937. Doi: 10.1016/j.imu.2022.100937

Anwar, H., Qamar, U., Muzaffar Qureshi, A.W. 2014. Global optimization ensemble model for classification methods. Sci. World J., 2014:313164. Doi: 10.1155/2014/313164

Asch, D.A., Sheils, N.E., Islam, M.N., Chen, Y., Werner, R.M., Buresh, J., Doshi, J.A. 2021. Variation in US Hospital Mortality Rates for Patients Admitted with COVID-19 during the First 6 Months of the Pandemic. JAMA Intern. Med., 181(4):471–478. Doi: 10.1001/jamainternmed.2020.8193

Bishop, C. M. 2006. Pattern recognition and machine learning (Information science and statistics). Springer-Verlag New York, Inc.

Breiman L. 1996. Bagging predictors. Machine Learning. 24:123-140. Doi: ttps://doi.org/10.1007/BF00058655

Breiman L. 2001. Random forests. Machine Learning. 45:5-32. Doi: 10.1023/A:1010933404324

Brinati, D., Campagner, A., Ferrari, D., Locatelli, M., Banfi, G., Cabitza, F. 2020. Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study. Journal of Medical Systems, 44(135). Doi: 10.1007/s10916-020-01597-4

Cabitza, F., Campagner, A., Ferrari, D., Di Resta, C., Ceriotti, D., Sabetta, E., Colombini, A., … Carobene, A. 2021. Development evaluation and validation of machine learning models for covid-19 detection based on routine blood tests. Clinical Chemistry and Laboratory Medicine (CCLM), 59(2):421-431. Doi: 10.1515/cclm-2020-1294

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. 2002. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16(1):321-357. Doi: 10.1613/jair.953

Chen, S.G., Chen, J.Y., Yang, Y.P., Chien, C.S., Wang, M.L., Lina, L.T. 2020. Use of radiographic features in covid-19 diagnosis: Challenges and perspectives. Journal of the Chinese Medical Association, 83(7):644-647. Doi: 10.1097/JCMA.0000000000000336

Dai, W., Li, D., Tang, D., Wang, H., Peng, Y. 2022. Deep learning approach for defective spot welds classification using small and class-imbalanced datasets. Neurocomputing, 477(8):46-60. Doi: 10.1016/j.neucom.2022.01.004

Dey, N., Mishra R., Fong S.J., Santosh K.C., Tan S., Crespo R.G. 2020. COVID-19: Psychological and psychosocial impact, fear, and passion. Digit. Gov.: Res. Pract., 1, 1–4. Doi: 10.1145/3428088

Dong, X., Yu, Z., Cao, W., Shi, Y., Ma, Q. 2020. A survey on ensemble learning. Front. Comput. Sci., 14(2):241-258. Doi: 10.1007/s11704-019-8208-z

Douzas, G., Bação, F., Last, F. 2018. Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. Inform Sci., 465:1-20. Doi: 10.1016/j.ins.2018.06.056

Elshennawy, N.M., Ibrahim, D.M., Sarhan, A.M., Arafa, M. 2022. Deep-Risk: Deep Learning-Based Mortality Risk Predictive Models for COVID-19. Diagnostics (Basel), 12(8):1847. Doi: 10.3390/diagnostics12081847

Erol, G., Uzbaş, B., Yücelbaş, C., Yücelbaş, Ş. 2022. Analyzıng The Effect Of Data Pre-Processing Techniques Using Machine Learning Algorithms On The Diagnosis Of COVID-19. Concurrency and Computation-Practice & Experience, 34(28). Doi: 10.1002/cpe.7393

Erol Doğan, G., Uzbaşa, B. 2023. Diagnosis of COVID-19 from blood parameters using convolutional neural network. Soft Computing, 27(15):10555–10570. Doi: 10.1007/s00500-023-08508-y

Famiglini, L., Bini, G., Carobene, A., Campagner, A., Cabitza, F. 2021. Prediction of ICU admission for COVID-19 patients: a Machine Learning approach based on Complete Blood Count data. 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS). Doi: 10.1109/CBMS52027.2021.00065

Feigin, E., Levinson, T., Wasserman, A., Shenhar-Tsarfaty, S. 2022. Age-Dependent Biomarkers for Prediction of In-Hospital Mortality in COVID-19 Patients. J. Clin. Med., 11(10):2682. Doi: 10.3390/jcm11102682

Fernandez, A., Garcia, S., Herrera, F., Chawla, N. 2018. SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. Journal of Artificial Intelligence Research, 61(1):863-905. Doi: 10.1613/jair.1.11192

Fong, S., Li G., Dey N., Crespo R.G., Herrera-Viedma E. 2020. Finding an accurate early forecasting model from small dataset: A case of 2019-nCoV novel coronavirus outbreak. Int. J. Interact. Multimedia Artif. Intell., 6(1):1–10. Doi: 10.9781/ijimai.2020.02.002

Freud, Y. & Schapire, R.E. 1999. A Short Introduction to Boosting. Journal of Japanese Society for Artificial Intelligence, 14(5):771-780.

Ghani, A.C., Donnelly, C.A., Cox, D.R., Griffin, J.T., Fraser, C., Lam, T.H., Ho, L.M., … Leung, G.M. 2005. Methods for estimating the case fatality ratio for a novel, emerging infectious disease. Am J Epidemiol, 162(5):479–486. Doi: 10.1093/aje/kwi230

Göreke, V., Sarı V., Kockanat S. 2021. A novel classifier architecture based on deep neural network for COVID-19 detection using laboratory findings. Applied Soft Computing, 106(1):107329. Doi: 10.1016/j.asoc.2021.107329

Grasselli, G., Zangrillo, A., Zanella, A., Antonelli M., Cabrini, L., Castelli, A., Cereda, D., … COVID-19 Lombardy ICU Network, 2020. Baseline characteristics and outcomes of 1591 patients infected with sars-cov-2 admitted to icus of the lombardy region, italy. JAMA, 323(16):1574–1581. Doi: 10.1001/jama.2020.5394

Hajian-Tilaki, K. 2013. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Caspian Journal of Internal Medicinen, 4(2): 627–635.

Hu, Z., Melton, G.B., Arsoniadis, E.G., Wang, Y., Kwaan, M.R., Simon, G.J. 2017. Strategies for handling missing clinical data for automated surgical site infection detection from the electronic health record, 68: 112–120. Doi: 10.1016/j.jbi.2017.03.009

Huang, H., Bader, J.S. 2009. Precision and recallestimatesfortwo hybrid screens. Bioinformatics, 25(3):372–378. Doi: 10.1093/bioinformatics/btn640

Hulley, S.B., Cummings, S.R., Browner, W.S., Grady, D., Hearst, N., Newman, T.B. 2001. Studies of the Accuracy of Tests. Designing Clinical Research An Epidemiologic Approach. Second Edition. Lippincott Williams & Wilkins, 181-2.

Idri, A., Abnane, I., Abran, A. 2016. Missing data techniques in analogy-based software development effort estimation. The Journal of Systems and Software, 117, 595–611. Doi: 10.1016/j.jss.2016.04.058

Japkowicz, N., Shah, M. 2011. Evaluating Learning Algorithms, Cambridge University Press. Doi: 10.1017/CBO9780511921803

Kong, Y., Han, J., Wu, X., Zeng, H., Liu, J., Zhang, H. 2020. VEGF-D: a novel biomarker for detection of COVID-19 progression. Crit Care, 24(1):373. Doi: 10.1186/s13054-020-03079-y

Kumar, Y., Koul, A., Singla, R., Ijaz, M.F. 2023. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. Springer Nature, 14(7):8459–8486. Doi: 10.1007/s12652-021-03612-z

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Peng Z. 2020. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA, 323(11):1061–9. Doi: 10.1001/jama.2020.1585

World Health Organization (WHO) 2020. Health topics, coronavirus. https://www.who.int/health-topics/coronavirus#tab=tab _ 3

World Health Organization (WHO) 2022. WHO Coronavirus (COVID-19) Dashboard: https://covid19.who.int/

Willyard, C. 2020. Coronavirus blood-clot mystery intensifies. Nature, 581(7808):250. Doi: 10.1038/d41586-020-01403-8

Wynants, L., Van Calster, B., Collins, G.S., Riley, R.D., Heinze, G., Schuit, E., Bonten, M.M.J., … van Smeden, M. 2020. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ, 7(369):m1328. Doi: 10.1136/bmj.m1328

Mahaney, M.C., Brugnara, C., Lease, L.R., Platt, O.S. 2005. Genetic influences on peripheral blood cell counts: a study in baboons, 106(4): 1210–1214. Doi: 10.1182/blood-2004-12-4863

Metz, C.E. 1978. Basic principles of ROC analysis. Semin Nucl Med., 8(4):283-98. Doi: 10.1016/s0001-2998(78)80014-2

Mohammed, A., Kora, R. 2023. A Comprehensive Review on Ensemble Deep Learning: Opportunities and Challenges. Journal of King Saud University - Computer and Information Sciences, 35(2):757-774. Doi: 10.1016/j.jksuci.2023.01.014

Moulaei, K., Shanbehzadeh, M., Mohammadi-Taghiabad, Z., Kazemi-Arpanahi, H. 2022. Comparing machine learning algorithms for predicting COVID-19 mortality. BMC Medical Informatics and Decision Making, 22(1). Doi: 10.1186/s12911-021-01742-0

Moore, J.B., June, C.H. 2020. Cytokine release syndrome in severe COVID-19. Science, 368(6490):473–474. Doi: 10.1126/science.abb8925

Nagant, C., Ponthieuxa, F:, Smet, J., Dauby, N. 2020. A score combining early detection of cytokines accurately predicts COVID-19 severity and intensive care unit transfer. International Journal of Infectious Diseases, 101:342-345. Doi: 10.1016/j.ijid.2020.10.003

Pasquier, G., Bounhiol, A., Gangneux, F.R., Zahar, J.R., Gangneux, J.P., Novara, A., Bougnoux, M.E., Dannaoui, E. 2021. A review of significance of Aspergillus detection in airways of ICU COVID-19 patients. Mycoses. 64(9):980-988. Doi: 10.1111/myc.13341.

Podder, P., Khamparia, A., Mondal, M.R.H., Rahman, M.A. (2021). Forecasting the Spread of COVID-19 and ICU Requirements, International Journal of Online and Biomedical Engineering (iJOE) 17(05):81. Doi: 10.3991/ijoe.v17i05.20009

Prusa, J., Khoshgoftaar, T.M., Dittman, D.J. 2015. Using ensemble learners to improve classifier performance on tweet sentiment data. 2015 IEEE International Conference on Information Reuse and Integration, 252-257. Doi: 10.1109/IRI.2015.49

Rahman, M. and Davis, D.N. 2013. Addressing the Class Imbalance Problem in Medical Datasets, 3(1): 224. Doi: 10.7763/IJMLC.2013.V3.307

Rodriguez-Nava, G., Yanez-Bello, M.A., Trelles-Garcia, D.P., Chung, C.W., Friedman, H.J., Hines, D.W. 2020. Performance of the quick covid-19 severity index and the brescia-covid respiratory severity scale in hospitalized patients with covid-19 in a community hospital setting. International Journal of Infectious Diseases, 102:571-576. https://doi.org/10.1016/j.ijid.2020.11.003

Sagi, O., Rokach, L. 2018. Ensemble learning: A survey Wiley Interdiscip. Rev.: Data Min. Knowledge Discov., 8(4):e1249. Doi: 10.1002/widm.1249

Shahzad, R.K., Lavesson, N. 2013. Comparative analysis of voting schemes for ensemble-based malware detection. J. Wireless Mobile Netw., Ubiquitous Comput. Dependable Appl., 4 (1):98-117.

Strålin, K., Wahlström, E., Walther, S., Bennet-Bark, A.M., Heurgren, M., Lindén, T., Holm, J., Hanberger, H. 2021. Mortality Trends among Hospitalised COVID-19 Patients in Sweden: A Nationwide Observational Cohort Study. Lancet Reg. Health Eur., 4, 100054. Doi: 10.1016/j.lanepe.2021.100054

Strålin, K., Wahlström, E., Walther, S., Bennet-Bark, A.M., Heurgren, M., Lindén, T., Holm, J., Hanberger, H. 2022. Mortality in Hospitalized COVID-19 Patients Was Associated with the COVID-19 Admission Rate during the First Year of the Pandemic in Sweden. Infect. Dis., 54, 145–151. Doi: 10.1080/23744235.2021.1983643

Yang, H.S., Hou, Y., Vasovic, L.V., Steel, P.A., Chadburn, A., Racine-Brzostek, A.E., Velu, P., … Wang, F. 2020. Routine laboratory blood tests predict sars-cov-2 infection using machine learning. Clinical chemistry, 66(11):1396-1404. Doi: 10.1093/clinchem/hvaa200