Mathematical Biosciences and Engineering

MBE, 22(3): 554–584.
DOI: 10.3934/mbe.2025021
Received: 29 October 2024
Revised: 04 December 2024
Accepted: 11 February 2025
Published: 24 February 2025

https://www.aimspress.com/journal/MBE

Research article

AI-driven health analysis for emerging respiratory diseases: A case study of Yemen patients using COVID-19 data

Saleh I. Alzahrani[1,*], Wael M. S. Yafooz[2], Ibrahim A. Aljamaan[1], Ali Alwaleedi[3], Mohammed Al-Hariri[4] and Gameel Saleh[1]

[1] Biomedical Engineering Department, College of Engineering, Imam Abdulrahman Bin Faisal University, PO box 1982, Dammam 31451, Saudi Arabia
[2] Computer Science Department, Taibah University, Saudi Arabia
[3] Department of Epidemiology and Public Health, College of Medicine, Aden University, Aden, Yemen
[4] Department of Physiology, College of Medicine, Imam Abdulrahman Bin Faisal University, PO box 1982, Dammam 31451, Saudi Arabia

* **Correspondence:** Email: sialzahrani@iau.edu.sa; Tel: +966504905815.

Abstract: In low-income and resource-limited countries, distinguishing COVID-19 from other respiratory diseases is challenging due to similar symptoms and the prevalence of comorbidities. In Yemen, acute comorbidities further complicate the differentiation between COVID-19 and other infectious diseases. We explored the use of AI-powered predictive models and classifiers to enhance healthcare preparedness by forecasting respiratory disease trends using COVID-19 data. We developed mathematical models based on autoregressive (AR), moving average (MA), ARMA, and machine and deep learning algorithms to predict daily confirmed deaths. Statistical models were trained on 80% of the data and tested on the remaining 20%, with predicted results compared to actual values. The ARMA model demonstrated promising performance. Additionally, eight machine learning (ML) classifiers and deep learning (DL) models were utilized to identify COVID-19 severity indicators. Among the ML classifiers, the Decision Tree (DT) achieved the highest accuracy at 74.70%, followed closely by Random Forest (RF) at 74.66%. DL models showed comparable accuracy scores, around 70%. In terms of AUC-ROC, the kernel Support Vector Machine (SVM) outperformed others, achieving 71% accuracy, with precision, recall, F-measure, and area under the curve values of 0.7, 0.75, 0.59, and 0.72, respectively. These findings underscore the potential of AI-driven health analysis to optimize resource allocation and enhance forecasting for respiratory diseases.

Keywords: AI-powered health analysis; COVID-19 prediction models; respiratory disease forecasting; machine learning in healthcare; deep learning classifiers; ARMA time-series models

1. Introduction

Due to the conflict in Yemen since 2014, and according to the World Health Organization (WHO), only half of hospitals and health facilities are fully accessible to Yemenis [1−3]. Only 700 beds are available in the intensive care unit (ICU) for the entire population [4]. The acute shortage of healthcare services, staff, medicines, and equipment means that Yemeni people are at greater risk than people in many other countries [5]. These difficulties suggest that the Yemeni population may be moving gradually toward herd immunity against COVID-19 [6]. Based on the epidemiological situation, the United Nations (UN) Humanitarian Coordinator raised concerns in April 2020 that COVID-19 could affect up to 16 million (55%) Yemenis [7]. In January 2022, the rate of mortality was 19.4% (https://covid19.who.int/region/emro/country/ye). However, the exact impact of the disease is not easy to measure due to the lack of testing and the weakness of the health services due to the war in the country.

At the beginning of the pandemic, the number of confirmed cases in Yemen reached 1600. The mortality rate was 27%, which was five times the global average [8,9], with a testing rate of 162 tests per million people. Médecins Sans Frontières, which runs a COVID-19 center in Aden, reported that between April 30 and May 17, it had admitted 173 patients, of whom at least 68 had died [10]. The surging death rate in Aden at that time suggests that the virus and the number of confirmed cases spread much faster and further than anticipated [10].

The epidemiological situation was critical during the first two months before the official declaration of the first case on April 10, 2020, and two months after it was reported [11]. The data from 10 governorates, representing only one-third of the population, have been analyzed. Mortality was higher in patients younger than 60, and most were in critical condition when they reached the health facilities. The surveillance system captured mainly severe cases due to limited resources and the stigma associated with infection, which made it difficult to interpret mortality data [11]. Furthermore, the presence of other diseases among patients with COVID-19, such as swine flu (H1N1), dengue fever, malnutrition, and the effects of khat, could contribute to the high death rate in Yemen. In the context of these widespread acute comorbidities, COVID-19 is barely reported, and the lack of testing facilities makes it difficult to distinguish it from other infectious diseases.

The implementation of community surveillance (CBS) in the camps of internally displaced people (IDP) and urban settings in Yemen from mid-April to the end of September 2020 has been described in Baaees et al. [12]. It was found that CBS was useful for detecting suspected outbreaks in IDP camps; however, the system failed to detect suspected COVID-19 cases and other diseases despite the ongoing outbreaks reported through the Electronic Diseases Early Warning System (eDEWS).

The spread of COVID-19 and its impact have been estimated in many studies using forecasting models based on machine-learning and mathematical models [13]. These studies cover many countries, such as Japan, India, France, China, US, Italy, and Spain [14−21]. The simple mathematical models, among them, produced good results in predicting the spread of COVID-19 when compared with the artificial intelligence (AI) models that suffer from a lack of sufficient data as well as unreliable sources of data [22].

The mathematical models concerning epidemics are categorized as follows: Statistical methods

for epidemic surveillance, mechanistic state-space models, and empirical learning models [23]. Our work falls into the first category. Many prediction models belonging to this category have been introduced in the literature. A Gaussian Model (GM) is one such statistical method that detects the dynamic outbreak by monitoring the infected cases [24]. A Gaussian error function and a Monte Carlo simulation model were used to forecast the spread of COVID-19 in Italy and China and showed errors in the decelerating rate of positive cases and the substantial reduction in these cases that were recorded or reported [25,26]. The IHME COVID-19 team estimated the impact of the pandemic on hospital beds and the demand for ventilators in the USA [27]. Similar research was started in Germany and Europe to detect the exact date of the maximum number of cases during the first wave [28,29]. The Autoregressive Integrated Moving Average (ARIMA) model has been used to forecast the dynamics of COVID-19 in many countries and has shown good results [30]. Other researchers using ARIMA conducted studies in Italy, Spain, and Saudi Arabia, and the accuracy of the model for the prediction of the daily number of cases was acceptable [18,31].

The drawbacks of using mathematical models to forecast the spread of COVID-19 include their sensitivity to initial conditions, the number of parameters used, and the overfitting of the data [23]. The nature of a problem should be considered to identify the best-fit model. Some models are good at estimating the daily death and recovery rates, while others are suited to describing disease characteristics, and others have better descriptions of the effects of intervention measures; however, no single model meets all the requirements [23].

Other machine learning-based approaches were performed, such as in the work by Iwendi et al. [32], where Logistic Regressions, Decision Trees, and Random Forests algorithms were applied to the COVID-19 dataset of patients from Brazil and Mexico. In this study, some very promising results were provided. In [33], the clinical and laboratory data from 2566 COVID-19 patients were used to develop predictive models for hospitalization, ICU admission, and mechanical ventilation, achieving up to 88% accuracy. In [34], Logistic Regression and SVM models were used to predict COVID-19 mortality with up to 97% accuracy, identifying key predictors like age, lactate dehydrogenase, and C-reactive protein for early hospital resource allocation. The adaptive neuro fuzzy method is presented in reference [35]. Deep learning techniques were applied to the images, such as CT, X-ray, and MRI, of the COVID-19 patients [36]. Three examples from different locations around the globe were provided, including the challenges of the imaging techniques. Prinzi et al. [37] used clinical, laboratory, and radiomic features to train Support Vector Machine and Random Forest models for COVID-19 prognosis, achieving an AUC of 0.819 and accuracy of 0.733. Soda et al. [38] leveraged chest X-ray features, including handcrafted and CNN-extracted data integrated with clinical attributes, to predict severe COVID-19 outcomes, demonstrating promising performance on a dataset of 820 patients from six Italian hospitals. Wang et al. [39] developed radiomics, clinical, and combined models to predict COVID-19 outcomes, demonstrating good predictive performance using receiver operating characteristic curves, decision curves, and Delong's test.

Between 2020 and 2022, there were numerous researchers who used machine learning algorithms to study the respiratory failure in patients with COVID-19 [40−42], and other researchers compared techniques [43,44]. More research has been published in the last two years where ML is applied in medical applications [45−48].

Some researchers have reported SVM analysis for the classification of viral respiratory infections. The researchers in [44] found that the SVM classifier achieved a classification accuracy of 86.44% in the classification of Middle East Respiratory Syndrome Coronavirus (MERS-CoV). This result, based

on healthcare personnel class and the samples used, was around one-tenth of the total samples. Jiang et al. [45] used respiratory sounds and inspiratory cycles as input features to the SVM classifier to determine irregularities in respiratory diseases. They achieved a classification accuracy of 72%. In another study, Rohith Reddy [46] classified 30 chest X-ray images for COVID-19 and normal subjects using SVM. The classification accuracy was 57.1% using gray-level co-occurrence matrices (GLCMs) as features. In [48], 39 features were created from multimedia texts and used to detect fake news regarding COVID-19 using state-of-the-art deep learning models.

The reason for focusing on Yemen is primarily because, as reported, the high COVID-19 case-fatality ratio in Yemen is indicative of cases under ascertainment. Thus, COVID-19 may be unmitigated and undetected, and the real pandemic figures could be far higher than reported, as suggested elsewhere [49]. The secondary objective is to assist authorities in managing their limited resources effectively, taking into account the various risk factors and doing their best to evaluate the efficacy of therapeutic strategies and short-term resources. To our knowledge, this is one of the first studies to predict the spread of COVID-19 using mathematical models and to use classifications in Yemen. In this work, autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA) models have been used to forecast the number of new confirmed and death cases of COVID-19 in Yemen. In this study, we employed 8 supervised ML classifiers and 3 DL models to distinguish normal from COVID-19 patients based on 25 characteristics, including age and other features. These 8 ML classifiers and 3 DL models are Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes (NB), Gradient Boosting (GB), and AdaBoost (Ada), and the DL models being Long Short-Term Memory (LSTM), bidirectional being Long Short-Term Memory (BiLSTM), and Gated recurrent units (GRU). The selection of ML and DL models in this study was guided by their proven effectiveness in addressing similar biomedical and healthcare challenges. For example, Decision Trees and Random Forests were selected for their interpretability and robustness, especially when handling imbalanced datasets. Additionally, Logistic Regression and Support Vector Machine were chosen due to their well-established performance in binary classification tasks, while KNN was included for its effectiveness in smaller datasets. Among DL models, LSTM, BiLSTM, and GRU were selected for their ability to capture temporal patterns in sequential data, making them particularly suitable for COVID-19 symptom progression analysis. The objective is to specify an accurate classifier that can identify the most relevant factors that would serve as indicators for the severity of COVID-19. Our proposed method for forecasting the new COVID-19 cases can be used to help the government to obtain cost-effective decision-making policies to suppress the spread of the pandemic. In addition, the accurate prediction of COVID-19 cases is a cost-effective process since it will result in low-cost medical expenditure. Moreover, our proposed method can be useful in predicting the direct and indirect costs and facilities required in the future. The main contributions of this study are as follows:

1) Introducing a unique dataset of 35,265 suspected cases with numerous features that are being published for the first time.
2) Applying statistical methods to the proposed dataset to detect COVID-19 cases based on common symptoms.
3) Examining the performance of machine learning classifiers and deep learning models in detecting COVID-19 cases.
4) Investigating the important features that impact the detection of COVID-19 cases.

The work presented in this article is organized as follows: In Section 2, we present the data collection and describe prediction algorithms and classifiers and the performance metrics. In Section 3, we present the simulation results. In Sections 4 and 5, we describe the limitation and conclusion, respectively.

2. Materials and methods

2.1. Data description and preprocessing

In this work, we use patients' data under IRB REC-124-2022. The dataset includes all COVID-19 cases in Yemen reported by the Yemeni Ministry of Health from the first measured day in April 2020 to the end of August 2022. Data are collected by the ministry's epidemiological surveillance team, which belongs to the electronic integrated Disease Early Warning System (eDEWS). Of the 35,265 suspected cases, 11,925 are confirmed. The dataset provides comprehensive information about each subject, including demographic (such as age, gender, and governorate of residence), clinical (such as the presence of chronic diseases and comorbidities such as diabetes, cardiovascular disease, kidney disease, and asthma), and symptomatic (such as headache, chest pain, muscle and joint pain, loss of smell, loss of taste, and shortness of breath) features.

Data preprocessing is a necessary step to prepare the dataset for predication and classification tasks and to ensure the integrity of the analyses. Irrelavent features, such as personal identifiers (e.g., patient names and contact information), are removed. Additionally, features with a high percentage of missing values ($>$ 90%), such as travel history, are also excluded. Missing numerical values, such as age, are addressed using median imputation, while missing categorical features, such as symptoms, are imputed using the mode to ensure consistency. Other categorical features, including gender, symptoms, and comorbidities, are one-hot encoded to represent each unique category as binary variables. The target variable, representing confirmed and suspected cases, is label-encoded to assign binary values (Yes = 1, No = 0). For time-series analysis, date and daily new cases are used. To evaluate the model's performance in forecasting the future trends, the dataset is split into training (80%) and testing (20%) subsets. For classification task, demographic features, clinical information, and symptomatic features are utilized. Numerical features, such as age, are standardized to ensure all features are on a similar scale.

The distribution of the governorate centers from which these data are collected is shown in Figure 1. The distribution of the major comorbidities associated with COVID-19 patients is shown in Figure 2. Of the 11,925 confirmed cases, 17.95% had a chronic disease, 4.67% had cardiovascular disease, 10.47% had diabetes, 11.47% had high blood pressure, 1.84% had kidney disease, 0.03% had human immunodeficiency virus (HIV), and 0.67% had liver disease. Furthermore, 2.23% had asthma, 86.55% had a fever, 62.98% had a sore throat, 79.82% had a cough, 38.15% had nose descent, 64.21% had difficulty breathing, 22.31% had a headache, and 22.69% had chest pain. Furthermore, 59.82% had muscle and joint pain, 4.41% had diarrhea, 18.64% lost their ability to smell and taste, 1.06% had visited an endemic area for the last 14 days prior to their infection, 5.16% had contacted a suspected case, 1.53% smoked, and 2.56% had been vaccinated.

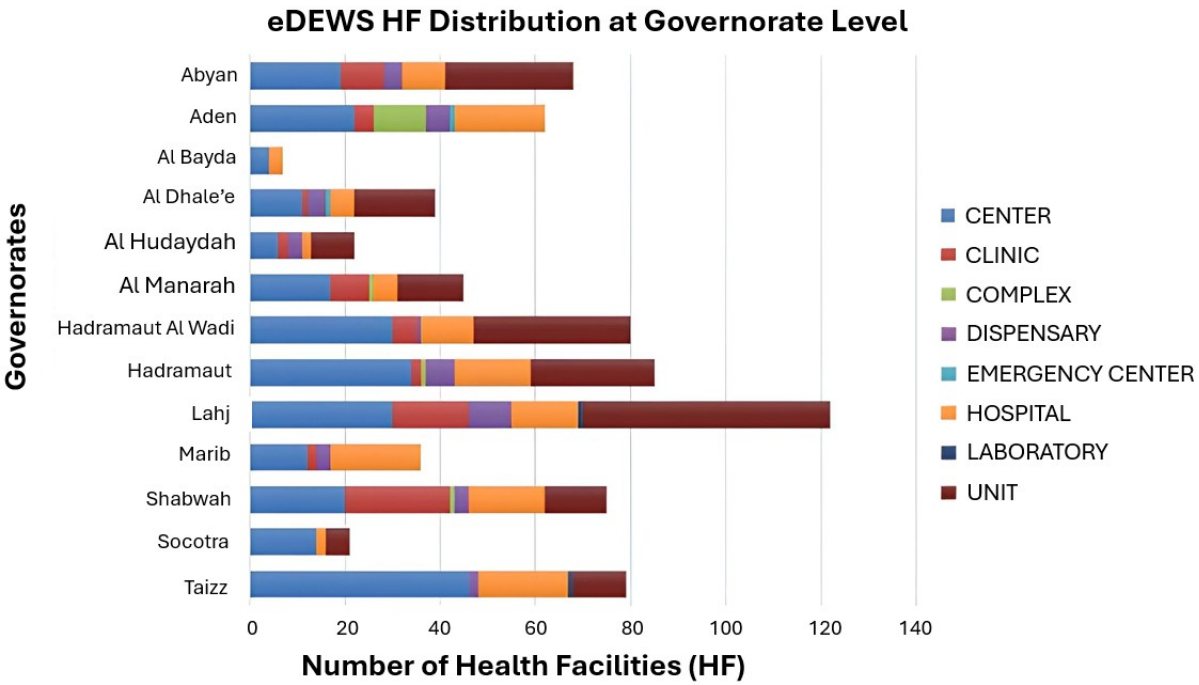


Figure 1. Distribution for providers of COVID-19 medical centers in Yemen.

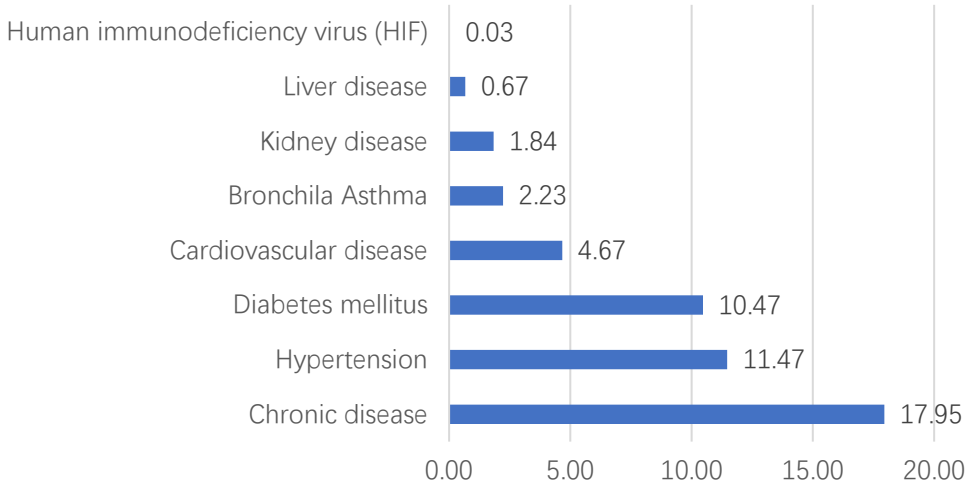


Figure 2. Comorbidities associated with COVID-19 patients.

2.2. Description of prediction models

We focus on using certain statistical models to forecast the spread of COVID-19 in Yemen. The autoregressive (AR) model of order p, known as the AR (p) model, is considered first. It is expressed as follows [50]:

$$y_t = \phi_0 + \phi_1 y_{(t-1)} + \phi_2 y_{(t-2)} + \phi_3 y_{(t-3)} + \cdots + \phi_p y_{(p-1)} + \varepsilon_t, \quad (1)$$

where ε_t is white noise and ϕ_p is an AR coefficient at lag p. In this model, the current value (y_t) depends only on its previous values $y_{(t-1)}, y_{(t-2)}, y_{(t-3)}$, etc.

The moving average (MA) model of order q, known as the MA (q) model, can be expressed as follows (Hyndman & Athanasopoulos, 2018):

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} + \theta_3 \varepsilon_{(t-3)} + \cdots + \theta_q, \tag{2}$$

where μ is the constant mean of the process, ε_t is the error term at time t, and $\varepsilon_{(t-1)}, \ldots, \varepsilon_{(t-q)}$ are the errors in the past. Unlike the AR model, the current value y_t in the MA model is expressed in terms of past random errors, which follow a white noise process.

An autoregressive moving-average process of order p and q, known as the ARMA (p, q) model, is expressed as a combination of (1) and (2) as follows [51]:

$$y_t = \phi_0 + \phi_1 y_{(t-1)} + \phi_2 y_{(t-2)} + \phi_3 y_{(t-3)} + \cdots + \phi_p y_{(p-1)} + \mu + \varepsilon_t + \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} + \theta_3 \varepsilon_{(t-3)} + \cdots + \theta_q \varepsilon_{(t-q)}, \tag{3}$$

where the current value here depends on both p and q past values of the white noise disturbances.

These three time-series models require stationarity. Data are stationary when their statistical properties, i.e., mean, variance, and covariance, are time-invariant. Non-stationary data transform to stationary by enabling differentiation of the data series. This is achieved by subtracting the previous value from the current value. The model that includes this step is known as the ARIMA model.

This model was first introduced by Box and Jenkins in 1976 [52]. It is one of the most frequently used time series based on statistical models for short-term predictions of future observations. It is a combination of three univariate time-series models: AR, MA, and ARMA models.

The general nonseasonal model is known as ARIMA (p, d, q), where p is the order of the autoregressive, d is the degree of differencing, and q is the order of the moving average. The values of the p and q parameters are chosen based on statistical tools such as the Autocorrelation Function (ACF) graph, the Partial Autocorrelation Function (PACF) graph, the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC). A common representation of an ARIMA model can be written as follows:

$$\phi_p(B)(1-B)^d y_t = \mu + \theta_q(B)\varepsilon_t, \tag{4}$$

where B indicates the backward linear operator, $(1-B)^d$ is the difference filter, and d is the degree of differencing needed to make the data stationary.

2.2.1 Estimation of parameters

Before choosing the best parameters for the model, the data are checked to see whether they are stationary or not. In order to do so, two methods are employed: The Augmented Dickey–Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests. The ADF test is a statistical significance test that is conducted with the assumption that the time-series data is non-stationary, and if this null hypothesis is rejected (i.e., the p-value is greater than the significance level), then the data is considered stationary. The KPSS is a type of unit root test that checks the stationarity of a time series around a deterministic trend. Its null hypothesis is the opposite to that of the ADF test, i.e., if the p-value is greater than the significance level, then the time series is stationary. Both methods are implemented

using the Python scripting language.

To choose the best parameters for each model, both AIC and BIC are calculated for different values of p and q as follows:

$$AIC = -2\log(l) + 2k, \tag{5}$$

$$BIC = k\ log(n) - 2l, \tag{6}$$

where l is the likelihood of the model and N is the total number of estimated parameters in the model. The parameters that give the minimum AIC and BIC values are chosen for the model for future prediction.

Additionally, the performance of each model is evaluated using three common statistical measures of forecast precision in univariate time series data: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). These measures are given as:

$$RMSE = \sqrt{\{1/N \sum_{i=1}^{N}(y_i - \tilde{y}_i)^2)}, \tag{7}$$

$$MAE = \sum_{i=1}^{N}|y_i - \tilde{y}_i| /N, \tag{8}$$

where y_i and $\tilde{y}_i$ are actual and predicted values, respectively.

R^2 is defined as follows:

$$R^2 = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(z_i - \tilde{z}_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(z_i - \bar{z}_i)^2}, \tag{9}$$

where $\bar{z}_i$ is defined as

$$\bar{z}_i = \frac{1}{N}\sum_{i=1}^{N} z_i, \tag{10}$$

2.3. Classification algorithms

Six different classification algorithms are applied in this work based on alternative concepts. The Logistic Regression (LR) method is a binary classification logarithmic-based technique that applies statistics to classify outputs out of two alternatives using the sigmoid function [53]. For an independent variable x and dependent variable y, the values of y between 0 and 1 will be predicted by finding the probability of x. Commonly, the maximum likelihood is applied to predict the best model that fits the sigmoid line using the probability of x as in (11):

$$ln\left(\frac{P(x)}{1-P(x)}\right) = e^{b_0 + b_1 x}, \tag{11}$$

where the predicted dependent variable is the probability of the independent variable, $\hat{y} = P(x)$.

Unlike the Logistic Regression approach, Linear Discriminant Analysis (LDA) is a statistical binary approach where the distribution of the data is estimated given a class. It creates new access that maximizes the separation of the two classes [54]. This supervised approach is commonly used to reduce dimensions by discovering new linear combination features. LDA commonly applies Eigen decomposition, which results in better efficiency, compression, and illustration.

Naive Bayes (NB) classifiers are another linear approach based on statistics that use the probability of independent features to classify between categories based on the Bayes rule [54].

Although the method is old, it is still common, especially because of the high dimensions p of the feature space. In this approach, the model features X_k are assumed independent and given a class $G = j$:

$$f_j(X) = \prod_{k=1}^{p} f_{jk}(X_k), \tag{12}$$

where f_{jk} is the individual class-conditional marginal densities.

The Decision Tree (DT) method is a mapping binary method that recursively splits the data set by applying a sequence of tests until we are left with one type of class [55]. It consists of four elements: Alternative branches, the two decisions or choices that stem from the main branch; Decision Node, the decision made out of the tree that is usually square shaped; Chance Nodes, which depict possible alternative outcomes in a circular shape; and End Nodes, which depict the final results as triangles. Usually, the Decision Tree approach can be performed in five steps as follows:

1) Start with the idea (decision node)
2) Create chance and decision nodes
3) Extend the tree until the end point
4) Calculate tree values using:

$$\begin{aligned} Expected\ value\ (EV) = {} & (1st\ possible\ outcome\ x\ Likelihood\ of\ outcome) + \\ & (2nd\ possible\ outcome\ x\ Likelihood\ of\ outcome) - Cost, \end{aligned} \tag{13}$$

5) Analyze the outcomes and make optimal decisions

The Random Forest (RF) method is an extended approach to the decision tree method where the data is combined into multiple trees called forests [55]. It may also be used for prediction and classification. The RF method can be implemented by dividing the dataset into subsets for an ensemble of Decision Trees, then training each of the decision tree subsets; after that, combining the decision tree outputs into one result; and finally, validating the approximated model.

The SVM is a regression-based method where the optimal hyperplane is approximated, and that separates data into classes [56]. Kernel SVM is used in this work, where a higher-dimensional plane is used to classify the data. The objective function in this approach is to find the optimum hyperplane that maximizes the margin separation. There are also SVM techniques for prediction and classification with linear and nonlinear algorithms. The hyperplane for one dimension is a point, a line for two-dimensional planes, a surface for three dimensions, and so on. In this method, the main goal is to find the best hyperplane that maximizes the margin, called the street, which separates the positive from the negative samples. Support vectors are the nearest data points to the hyperplane. Unlike most of the machine learning algorithms, such as neural networks, only the difficult points, or support vectors, have an impact on the final decision. In SVM, the output Y_i is predicted by finding the optimum weight W_j with input X_i that represent the feature and bias b as follows:

$$Y_i = W_j X_i + b, \tag{14}$$

2.3.1. Performance metrics

To assess the efficiency and performance of each classifier and measure the performance of our models, the accuracy, precision, recall, the F1 score, and the area under the curve (AUC) are calculated. Accuracy is defined as the ratio of the number of observations that are correctly classified (true positives and negatives) to the total number of observations. Mathematically, it is written as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \tag{15}$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Precision refers to the proportion of correctly classified positive observations among all positive observations. It is computed as follows:

$$Precision = \frac{TP}{TP+FP}, \tag{16}$$

Recall is the percentage of correctly classified positive observations out of the total number of positive observations. It is calculated as shown in (17):

$$Recall = \frac{TP}{TP+FN}, \tag{17}$$

The F1 score (also called the F measure) combines precision and recall into a single metric by calculating their harmonic means as follows:

$$F1\ score = \frac{2\times precision\times recall}{precision + recall}, \tag{18}$$

The AUC of receiver operator characteristics (ROC) is a curve that visualizes the ability of a classification model to distinguish between two or more classes. The curve is plotted with the true positive rate (TPR) against the false positive rate (FPR). The TPR and FPR are defined as:

$$TPR = \frac{TP}{TP+FN}, \tag{19}$$

$$FPR = \frac{FP}{FP+TN}, \tag{20}$$

3. Results and discussion

In this section, we describe the results of the experiments that are conducted to predict the confirmed or suspected cases of COVID-19 along with the experimental settings. The experiments are divided into two types: The first type of experimentation is done using statistical methods, ML classifiers; and the second type of experimentation is done using deep DL Models.

3.1. Experiment settings

All experiments were conducted using Google Colab with Python 3 utilizing GPU. The ML libraries used are SKlearn for splitting the dataset and the scaling process. The dataset is divided into three parts for training, validation, and testing the ML and DL models, at 70%, 10%, and 20%, respectively. Table 1 illustrates the ML hyperparameters and Table 2 shows the DL Experiment Hyperparameters.

Table 1. ML experiment hyperparameters.

ML Classifiers	Values
Random frest	n_estimators = 100, criterion = 'gini', max_depth = None, min_samples_split = 2, max_features = 'auto'
Decision tee	Criterion = 'gini', splitter = 'best', max_depth = None, min_samples_split = 2, max_features = None
Logistic Regression	Penalty = 'l2', C = 1.0, solver = 'lbfgs', max_iter = 100
Support vctor mchine	C = 1.0, kernel = 'rbf', degree = 3, gamma = 'scale', probability = True
K-Nearest Neighbors	n_neighbors = 5, weights = 'uniform', algorithm = 'auto', leaf_size = 30, p = 2, metric = 'minkowski'
Naive byes	var_smoothing = 1e-9
Gradient Boosting	n_estimators = 100, learning_rate = 0.1, max_depth = 3, subsample = 1.0, min_samples_split = 2, max_features = None
AdaBoost	n_estimators = 50, learning_rate = 1.0, algorithm = 'SAMME.R', base_estimator = DecisionTreeClassifier (max_depth = 1)

Table 2. DL experiment hyperparameters.

Hyperparameter	Value
Random sate	42
Encoding	LabelEncoder
Scaling	StandardScaler
Dense lyer 1 uits	128
Dense lyer 2 uits	64 (conditional on model type)
Dropout rte	0.5
Output lyer ativation	sigmoid
Optimizer	adam
Loss fnction	binary_crossentropy
Metrics	accuracy
Epochs	20
Batch sze	32
Model tpes	LSTM, BiLSTM, GRU

According to researchers, RF and GB are robust ensemble techniques that manage high-dimensional data and lessen overfitting [57]. While LR is frequently utilized for its effectiveness and explainability in feature-target connections [58], DT provides interpretability and acts as a baseline for tree-based models [59]. KNN efficiently uses local structures, especially in smaller datasets [60], while

SVM is excellent at establishing non-linear decision boundaries using kernel functions [61]. NB is well with categorical data and is computationally efficient [62]. It is known that ensemble techniques like GB and AdaBoost can improve weak learners step-by-step and increase prediction accuracy [63]. In line with best practices in current machine learning research, this wide collection of models guarantees a thorough assessment of our dataset under several paradigms.

Starting hyperparameters that are known to deliver dependable performance over a range of datasets are based on accepted techniques and literature. For example, SVM utilizes an RBF kernel because of its capacity to manage intricate and non-linear relationships, while RF and Gradient Boosting utilized 100 estimators to guarantee a balance between accuracy and processing cost. Values like var_smoothing = 1e-9 for NB and n_neighbors = 5 for KNN are also frequently suggested settings that work with a lot of data. Additionally, grid search and cross-validation approaches for hyperparameter tuning is used, which involves modifying parameters like learning rate, depth, and regularization, in order to guarantee optimal performance and produce the best results for ML classifiers. Initial domain knowledge is used to set the hyperparameters for deep learning models, and validation trials are used to fine-tune them.

While LSTM, BiLSTM, and GRU models can handle temporal and sequential data, they can be quite useful. When it comes to COVID-19 diagnosis, prognosis, and prediction tasks, these models perform well at identifying patterns over time, such as shifts in a patient' symptoms, health, or reaction to treatment. Long-term dependencies in patient records can be captured by LSTM, which aids in forecasting the course of the illness by using historical data (like the evolution of symptoms over time) [64]. When predicting patient outcomes, BiLSTM can be especially helpful in instances where both past and future data are relevant, such as comprehending the possible evolution of symptoms and treatment effects from both directions [65]. GRU is helpful when working with big datasets or real-time tracking of COVID-19 patient statuses since it is computationally efficient and can handle sequential data well while using fewer resources [66]. These models are useful resources for comprehending the course of COVID-19, forecasting patient outcomes, and seeing trends that may aid medical practitioners in taking appropriate action.

3.2. Statistical methods

We use the time series of confirmed daily COVID-19 in Yemen. In this study, the AR, MA, and ARMA models are utilized to predict the number of new confirmed COVID-19 cases in Yemen based on recorded data from April 10, 2020, when the first confirmed case was reported, to August 30, 2022. The models are trained on 80% of the data and tested on the remaining 20%. Figure 3 shows the confirmed daily and cumulative COVID-19 cases reported in Yemen during that period. The total number of confirmed cases was less than 50 until May 10, 2020; later, it increased and exceeded 100 cases within five days. Four waves are observed: One in 2020, two in 2021, and one in 2022. The total number of confirmed cases until the end of August 2022 was 11,925.

Figure 4 illustrates the daily death rate among COVID-19 patients, which shows a direct relationship with the number of new confirmed cases. The number of deaths per day was less than 10 in the last few weeks, and it exceeded 50 within a few days in July 2020, with a maximum of approximately 60 deaths per day in February 2022.

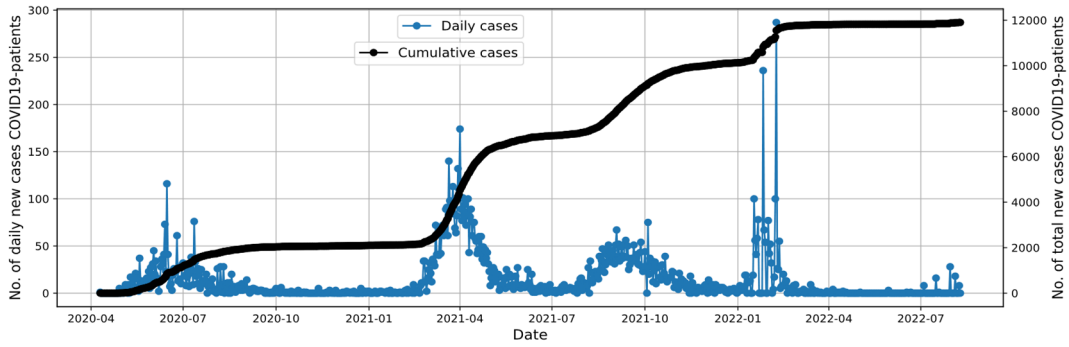


Figure 3. New (blue) and cumulative (black) confirmed COVID-19 cases daily.

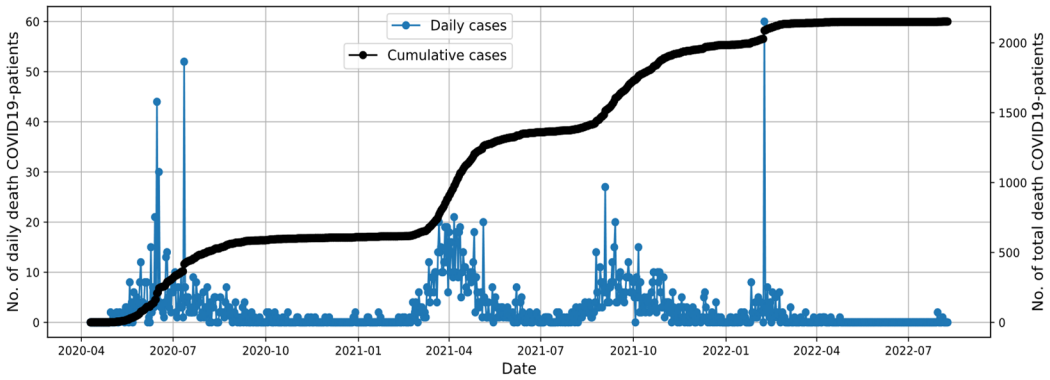


Figure 4. New (blue) and cumulative (black) death cases daily.

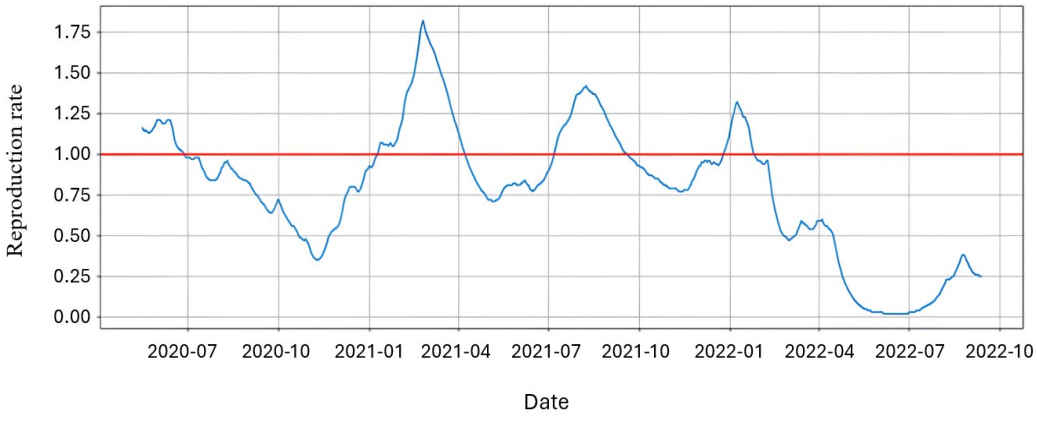


Figure 5. Estimates of the reproduction rate in Yemen from January 2020 to September 2022. The red line indicates $R = 1$.

The reasons for the higher percentage of death cases compared to confirmed cases are not only due to the lack of testing but also to the absence of precautionary measures, open borders with

neighboring countries, and low turnout for vaccination among people and healthcare workers [67]. A previous study in Aden, where more than a million people live, showed that 33% of the people were already infected in April 2021 (the second wave). Perhaps this explains what caused the reduction in confirmed cases during the next period that followed the first: Antibodies.

The reproduction rate, R, is an important factor that shows the speed of the spread of the disease. If the value is less than one, that means the epidemic has abated, whereas, if the value of R is greater than one, it means every COVID-19 patient infects more than one person. Figure 5 shows the reproduction rate in Yemen for the period from January 2020 to September 2022. The value of R fluctuates from zero at the minimum to more than 1.8 at the maximum. The estimated R was above 1 at the onset of the epidemic in Yemen, then fell below 1 rapidly from July to November 2020. However, from November 2020 to late March 2021, the estimated R drifted up toward 1 and reached its peak at 1.78. Since then, the estimated R has gone down substantially to around 0.735. Between April 2021 and February 2022, the pattern observed is similar to the previous months, where the estimated R exceeded 1 twice. After that, the estimated R dropped below 1, which indicates the effectiveness of the stringent health measures implemented in Yemen.

Based on the results of the ADF test (Augmented Dickey–Fuller) in Table 3, the p-value is less than 0.05, so the null hypothesis is rejected, which means the time-series data is stationary. Furthermore, the KPSS (Kwiatkowski–Phillips–Schmidt–Shin) test suggests that the data are stationary since p > 0.05. Hence, taking the difference is not necessary ($d = 0$).

Table 4 shows the AIC and BIC values for different p and q orders. The values of the p and q parameters range from 0 to 5. The best values for p and q that give the smallest AIC and BIC values are 5; based on this result, the AR (5), MA (5), and ARMA (5,5) models are considered the best-fit models for confirmed cases of COVID-19 in Yemen. In order to select the best model, a comparison is also conducted to investigate which model would best estimate the expected daily number of cases of COVID-19 in Yemen.

Table 3. Augmented Dickey–Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test results when the difference d = 0.

Test	ADF or KPSS Statistics	p-value	1% Critical vlue	5% Critical vlue	10% Critical vlue
ADF	-4.383	0.0003	-3.437	-2.865	-2.568
KPSS	0.255	0.1	0.739	0.463	0.347

After choosing the best parameters for the models, they are employed to forecast the number of confirmed cases. A total of 80% of trained data (blue), the actual (orange) of the last of 20% confirmed COVID-19 cases, and the predicted (green) are presented in Figure 6. Prediction accuracy metrics (RMSE, MAE, and R^2 values) for each model are illustrated in Figure 7. Overall, the most accurate and close estimate is achieved using an ARMA model, with an RMSE value of 21.51 and an MAE value of 6.06. The other two models have acceptable estimates and performances in the testing set, with an RMSE value of 23.74 and an MAE value of 7.01 for the AR model and an RMSE value of 24.06 and an MAE value of 9.79 for the MA model. Moreover, the results from the R^2 show that ARMA has a 77% score, which means that most of the dependent variables can be explained by the independent variables. The R^2 of the AR and MA are 55% and 46%, respectively. The prediction for the total number of cases for the next few months can be provided to the authorities

in Yemen to help them use their limited resources efficiently.

Table 4. AIC and BIC values for different p and q.

	0	1	2	3	4	5
0	(9063, 9073)	(8753, 8768)	(8649, 8669)	(8520, 8545)	(8480, 8509)	(8463, 8497)
1	(8508, 8523)	(8201, 8220)	(8202, 8227)	(8200, 8229)	(8202, 8236)	(8182, 8221)
2	(8373, 8393)	(8202, 8227)	(8191, 8220)	(8191, 8225)	(8192, 8231)	(8147, 8191)
3	(8282, 8306)	(8201, 8230)	(8191, 8225)	(8144, 8183)	(8144, 8188)	(8149, 8198)
4	(8267, 8297)	(8203, 8237)	(8192, 8231)	(8144, 8188)	(8161, 8209)	(8151, 8204)
5	(8241, 8275)	(8193, 8232)	(8151, 8195)	(8151, 8200)	(8141, 8195)	(8133, 8180)

To study if there is a statistically significant association between COVID-19 infection and 25 variables, a t-test and the Chi-square test are performed, and the results are shown in Figure 8 and Table 5. Since age is a numerical variable, the t-test is used to analyze the association between it and COVID-19 infection, while the Chi-square test is conducted on the other variables shown in Table 5 since they are categorical variables. The study population includes 23,075 cases (mean age: 46.4 ± 19.5 (SD) and 61.75% male) with 33.9% confirmed cases (35.8% > 50 years). Figure 8 shows that there is a strong association between age and COVID-19 infection ($p < 0.0001$). We conclude that older age increases susceptibility to infection.

The results of the Chi-square test shown in Table 5 reveal that there is no statistically significant difference between HIV, asthma, visiting an endemic area, smoking, and COVID-19 infection ($p > 0.05$). Since the Chi-square statistic is sensitive to sample size [67], in light of our results, we anticipate that a possible explanation for the non-significant difference might be attributed to the low reported cases (with HIV, asthma, visiting an endemic area, and smoking) due to the current Yemeni crisis and need to be confirmed by larger study samples. Additionally, due to cultural reasons, most patients hide such information. However, according to published reports, the association between asthma and COVID-19 severity and/or outcomes is controversial and continues to emerge [68]. Moreover, there is a statistically significant correlation between the other study variables and COVID-19 infection ($p < 0.05$) in Yemen.

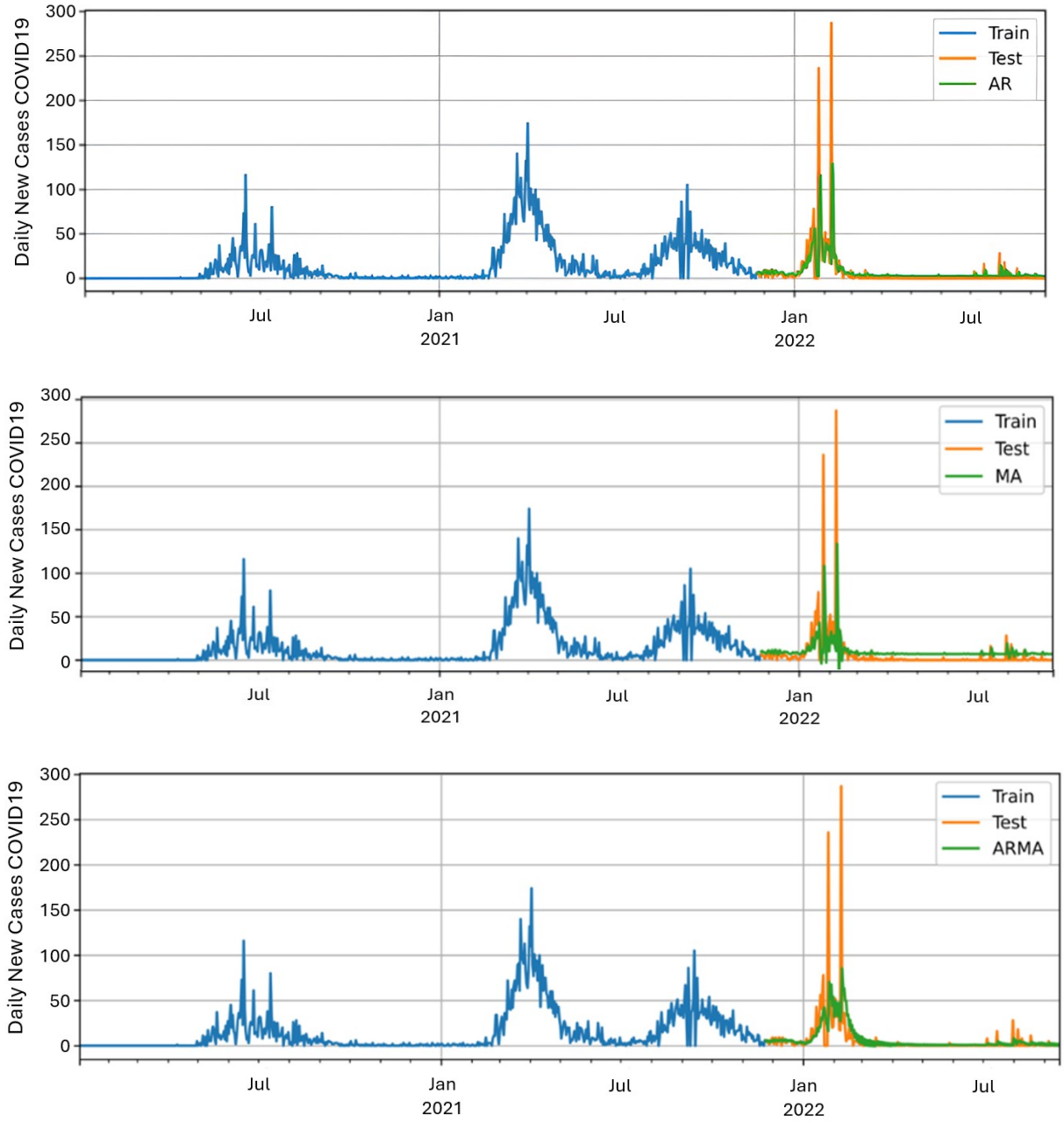
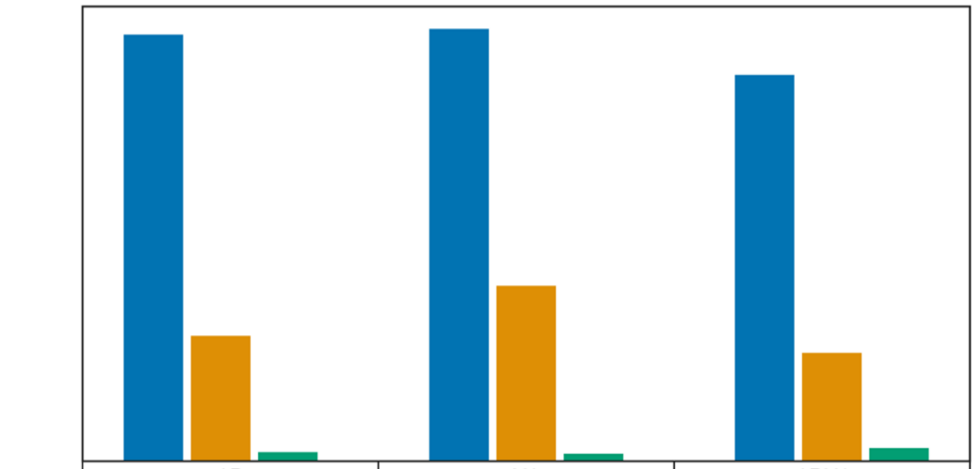


Figure 6. Prediction results from models. AR model (top), MA model (middle), and ARMA model (bottom).



	AR	MA	ARMA
RMSE	23.74	24.06	21.51
MAE	7.01	9.79	6.06
R^2	0.55	0.46	0.77

Figure 7. RMSE, MAE, and R^2 values for different time series models.

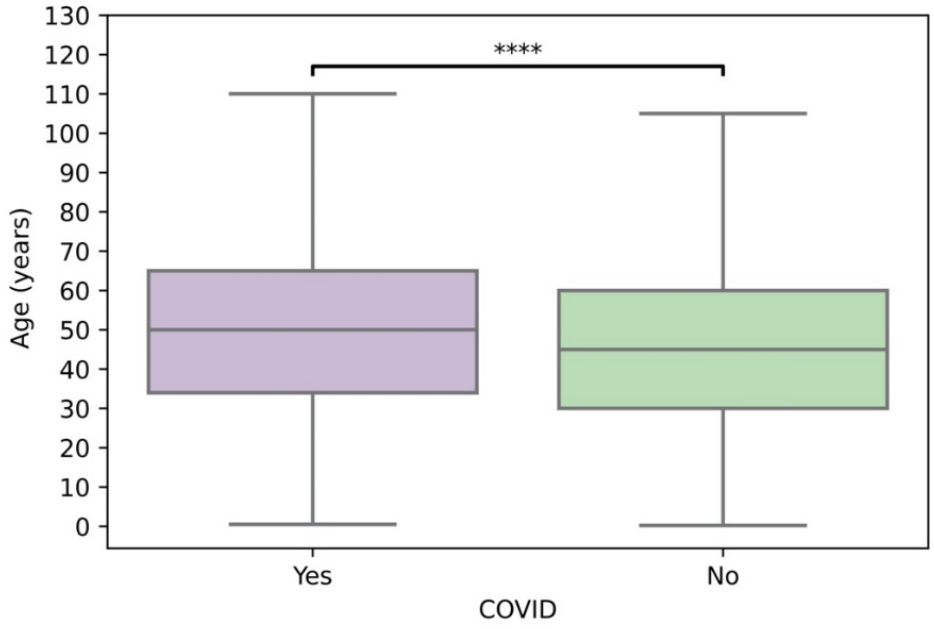


Figure 8. Box-plot illustrating the association between age and COVID-19. $p < 0.0001$.

Table 5. Chi-square test results between COVID-19 and various variables.

Variable	χ^2-statistics	p-value
Gender	77.99	< 0.0001
Chronic disease	6.27	0.01
Cardiovascular disease	62.15	< 0.0001
Diabetes	7.36	0.01
Blood pressure	5.62	0.02
Kidney disease	13.78	0.0002
HIV	0.035	0.85
Liver disease	5.58	0.02
Asthma	1.30	0.25
Visiting an endemic area	0.05	0.82
Fever	408.32	< 0.0001
Sore throat	598.75	< 0.0001
Cough	274.16	< 0.0001
Descent from nose	290.72	< 0.0001
Difficulty breathing	21.73	< 0.0001
Headache	991.46	< 0.0001
Chest pain	187.77	< 0.0001
Muscle and joint pain	1192.29	< 0.0001
Diarrhea	199.82	< 0.0001
Loss of smell	604.60	< 0.0001
Loss of taste	608.06	< 0.0001
Contact a suspected	19.76	< 0.0001
Smoking	0.67	0.41
Vaccination	170.42	< 0.0001

3.3. Experiment of ML classifiers

In this experiment, 8 ML classifiers are chosen, which include RF, DT, LR, SVM, KNN, NB, GB, and Ada. These ML classifiers are then put under experimentation to measure how well they can predict each patient getting COVID-19 based on the symptoms each person reported to have. This is measured under four distinct factors; Precision, Recall, F1-Score, and Accuracy [69], which can be seen from Table 6. Features used in the ML classifiers are selected based on their statistical significance, as determined by t-tests and Chi-square analyses. Variables with p-values < 0.05 are included in the feature set, ensuring that the models are built on predictors with proven associations to COVID-19 infection.

Generally, the ML classifier that is the best at predicting which patients would end up having COVID-19 based on their symptoms in terms of accuracy is DT. DT has an accuracy score of 74.70%, which is the highest in comparison to the other 8 ML classifiers. The RF classifier attains a score of 74.66%, which means it falls closely behind the DT classifier with a minor difference in accuracy of only 00.04%, which can be perceived from Table 6. This further proves that these two classifiers (DT and RF) can predict which patient is infected with COVID-19 based on their symptoms, understanding the importance of each symptom, and how this can lead to COVID-19. This demonstrates that these classifiers effectively predict COVID-19 based on symptoms. However, a classifier that struggles in

predicting COVID-19 for patients is NB, attaining an accuracy score of only 59.22%. Similarly, another classifier that did not perform too well in this experiment is LR, which has an accuracy higher than the NB classifier, at 67.01%. These results demonstrate that the DT and RF classifiers are well-equipped in predicting certain illnesses a person may end up having based on their symptoms; in this case, COVID-19. This can be further shown by relating to the confusion matrix and AUC-ROC metrics [70] for the ML classifiers in Figures 9 and 10. On the other hand, some other ML classifiers such as NB and LR may struggle significantly in predictions and in understanding the correlations between the features (symptoms), which can be viewed in Figure 11. Additionally, Figure 12 illustrates the relative importance of the features used in classification. From Figure 12, it can be clearly seen that symptoms such as muscle and joint pain, headache, and sore throat are the most influential predictors of COVID-19 severity. Furthermore, the accuracy ratings, as shown in Figure 10, are enhanced by the AUC values, which offer further information about the classifiers' capacity to discriminate positive and negative situations. Strong AUC ratings, for example, demonstrate the DT and RF classifiers' robustness in differentiating COVID-19 positive and negative patients, despite their greatest accuracy at 74.70% and 74.66%, respectively. Classifiers like NB and LR, on the other hand, demonstrate weaker AUC performance and lower accuracy ratings at 59.22% and 67.01%, respectively, indicating that they have difficulty in efficiently differentiating the two classes. The addition of AUC analysis and feature importance will give a better picture of the classifiers' prediction skills and aid to clarify how well they perform across thresholds.

Table 6. Comparison of ML classifiers using Precision, Recall, F1-Score, and Accuracy for the first experiment of all features (with 95% confidence intervals).

Classifier	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
RF	70.96 ± 1.57	56.35 ± 2.23	62.82 ± 1.23	**74.66** ± 1.06
DT	71.13 ± 1.78	56.20 ± 2.09	62.79 ± 2.01	**74.70** ± 1.13
LR	59.34 ± 1.93	41.77 ± 1.99	49.03 ± 1.22	67.01 ± 1.56
SVM	66.53 ± 2.55	52.27 ± 2.17	58.54 ± 1.33	71.88 ± 1.63
KNN	59.98 ± 1.78	58.56 ± 2.34	59.26 ± 1.17	69.42 ± 1.89
NB	47.63 ± 1.33	73.86 ± 2.89	57.92 ± 1.29	**59.22** ± 1.03
GB	62.07 ± 2.01	47.08 ± 1.89	53.54 ± 1.54	68.97 ± 1.69
Ada	58.86 ± 1.67	42.68 ± 2.22	49.48 ± 1.76	66.90 ± 1.77

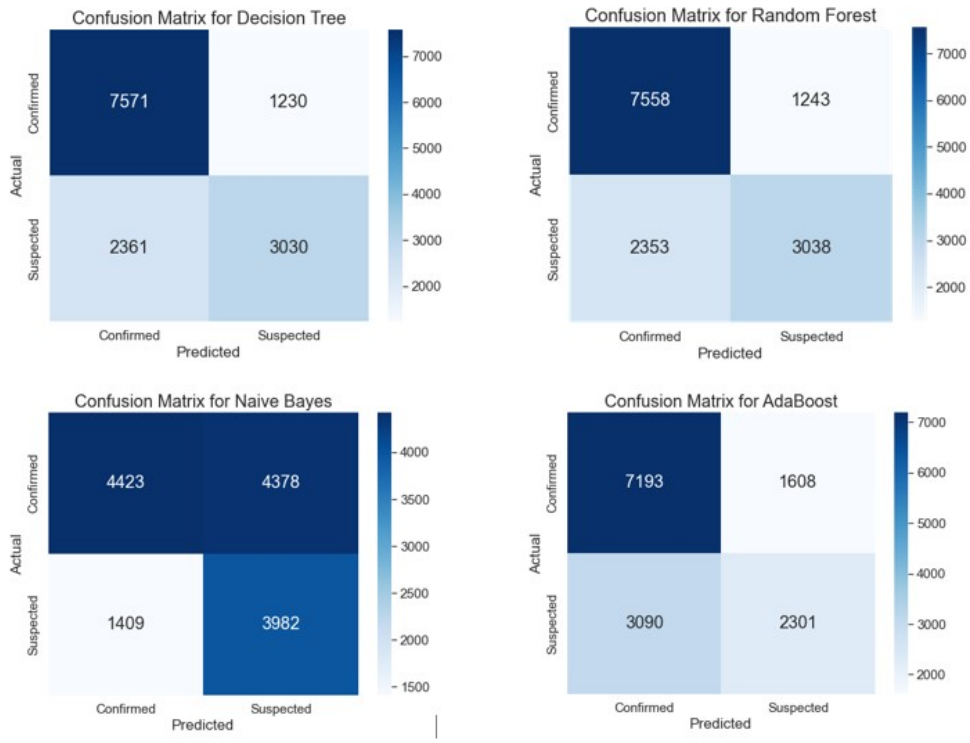


Figure 9. Confusion matrix for the ML classifiers.

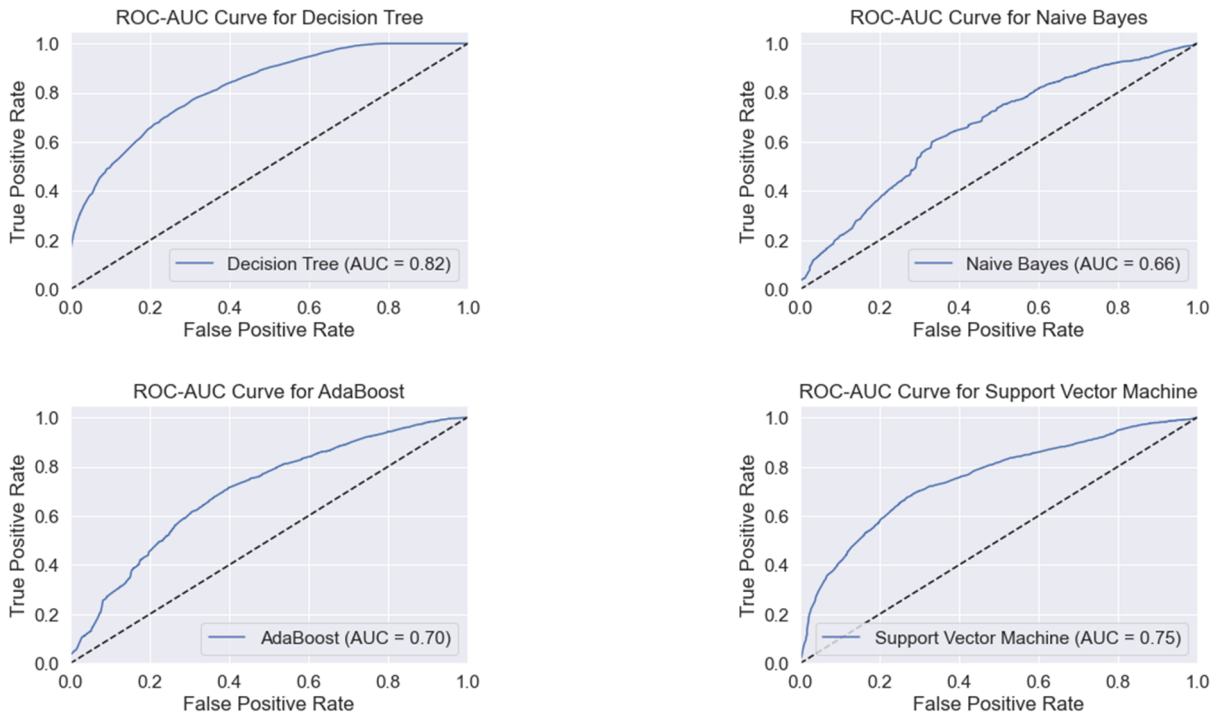


Figure 10. AUC-ROC of ML classifiers.

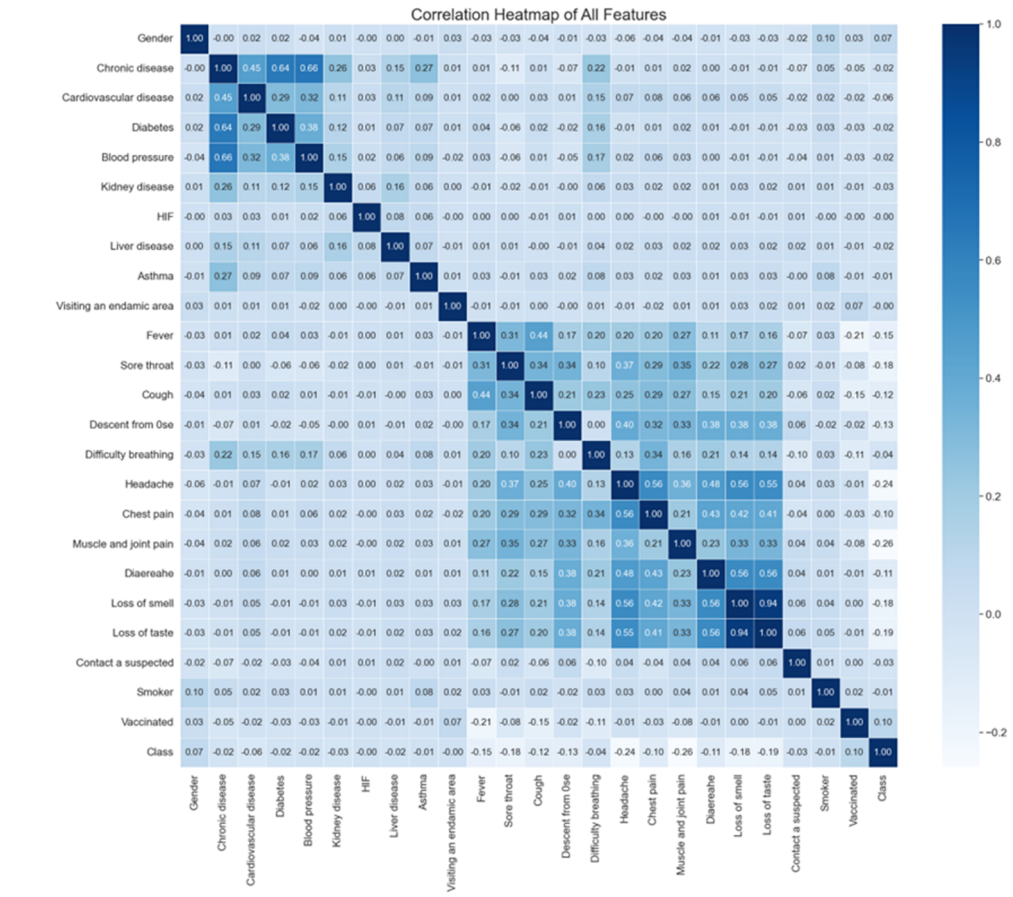


Figure 11. Correlation between the features (symptoms).

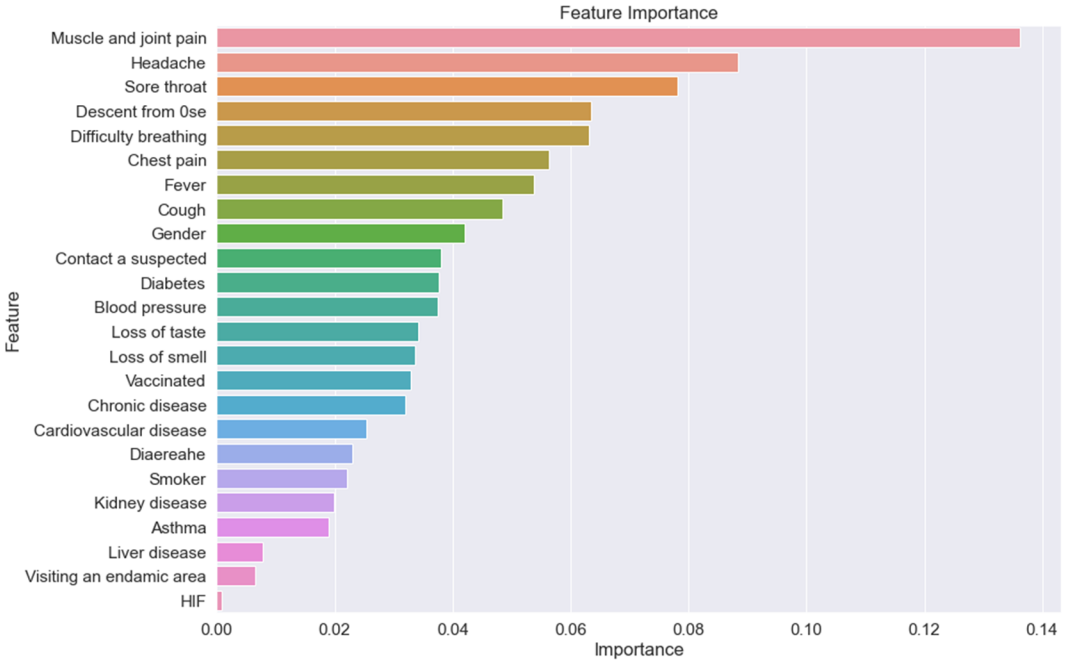


Figure 12. The importance of various features in predicting the severity of COVID-19 cases.

3.4. Experiments of deep learning models

This is the second type of experiment conducted to know the model's performance in predicting the affected cases based on the features (symptoms). In the DL experiments, three models are utilized, which are LSTM, BiLSTM, and GRU. The comparisons between precision, recall, F1, and accuracy are shown in Table 7.

Table 7. Comparison between precision, recall, F1, and accuracy using DL experiments (with 95% confidence intervals).

DL Models	Precision	Recall	F1	Accuracy
LSTM	67.57 ± 2.14%	46.84 ± 2.89%	55.32 ± 1.97%	71.27 ± 1.75%
BiLSTM	71.57 ± 2.09%	40.99 ± 2.77%	52.13 ± 2.15%	71.40 ± 1.83%
GRU	69.74 ± 2.33%	42.27 ± 3.01%	52.64 ± 2.21%	71.10 ± 1.80%

In this experiment, three DL models are used, which are LSTM, BiLSTM, and GRU. These models are picked to be experimented on to show how well they can predict COVID-19 for each person based on the symptoms each person is reported to have. This is measured based on precision, recall, F1, and accuracy. These three models all have similar performance in terms of precision, recall, and F1, but most notably in accuracy, since they all attain scores of around 70% in accuracy, for predicting cases of COVID-19 among people. We note that although the performance of the models is very similar, the best performing model is BiLSTM, attaining an accuracy score of 71.40%, followed closely behind LSTM, which attained a 71.27% accuracy score, and last by GRU, which reached an accuracy rate of 71.10% for predicting COVID-19 among patients. The confusion matrix, along with the accuracy and validation for the three aforementioned models, can be viewed in Figures 13 and 14.

Table 8 presents a comparative analysis of studies conducted in settings with characteristics similar to Yemen, including resource-constrained environments, limited healthcare infrastructure, and vulnerable populations. By examining studies from Ethiopia, Brazil, and India, this table highlights key aspects such as study objectives, datasets, models used, features analyzed, performance metrics, and challenges, situating our findings within the context of existing literature to underscore their significance.

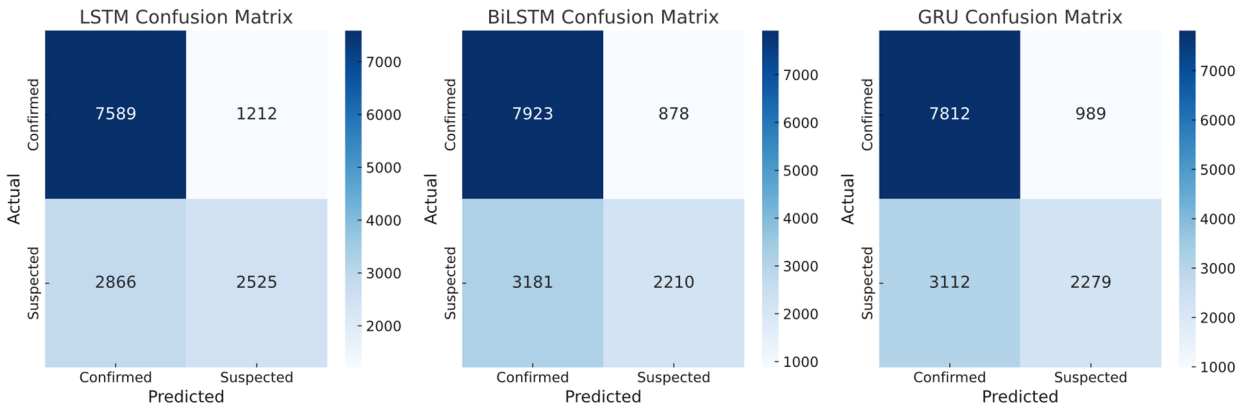


Figure 13. Confusion Matrix of the LSTM, BiLSM, and GRU.

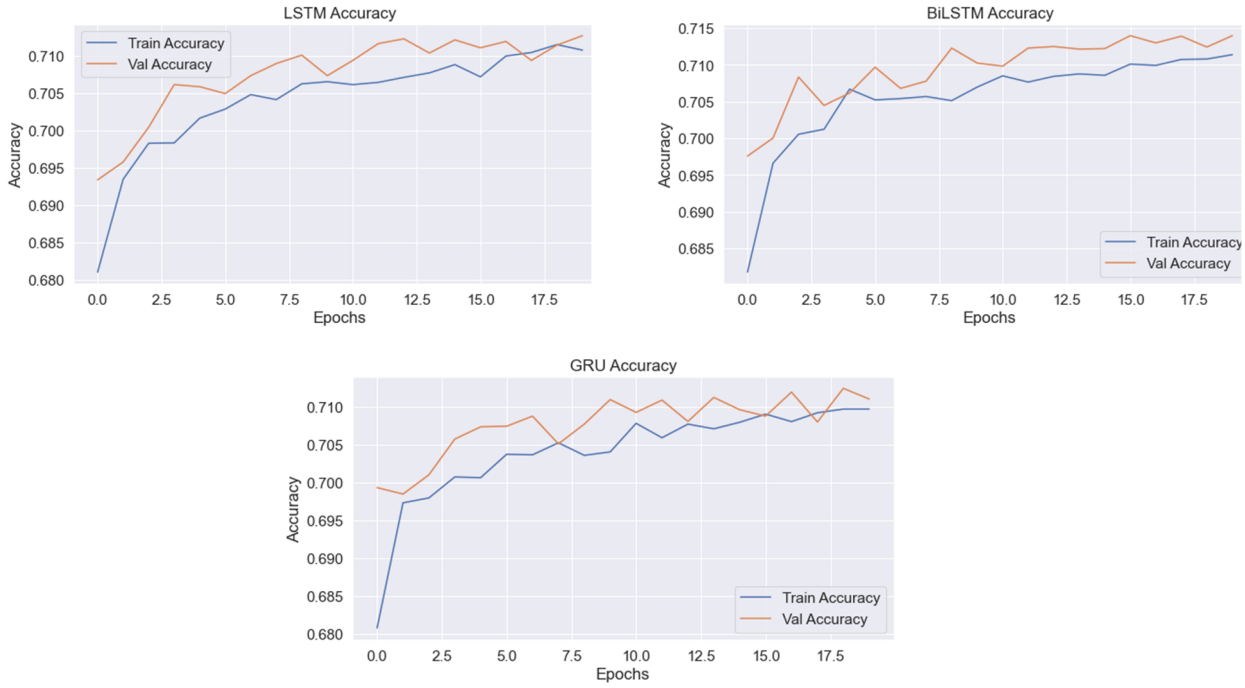


Figure 14. Accuracy of training and validation for LSTM, BiLSTM, and GRU.

Table 8. Comparison between this work and similar studies from literature.

Reference	Dataset characteristics	Models used	Key findings
[71]	696 hospitalized patients	J48 Decision Tree, RF, KNN, MLP, XGBoost, Logistic Regression, Naïve Bayes	KNN outperformed other models; top predictors were gender, ICU admission, and alcohol consumption
[72]	8443 patients	Logistic Regression (LR), LDA, KNN, Decision Trees (DT), XGBoost (XGB), SVM, Naïve Bayes (NB)	High recall in identifying severe cases but lower precision due to false positives; LDA and SVM were most consistent
[73]	5434 rows, 21 variables from Kaggle dataset: demographics, symptoms, and epidemiological factors	Random Forest, Logistic Regression, KNN	Random Forest demonstrated the highest accuracy and interpretability; validated model suitable for decision support
This study	35,265 suspected cases, 11,925 confirmed	AR, MA, ARMA, Decision Tree, Random Forest, SVM, LSTM	ARMA and DT models performed best; significant predictors include age and symptoms

4. Limitations

We acknowledge that this study has some limitations. The main limitation lies in the underrepresentation of certain subgroups within the dataset, such as individuals with rare comorbidities (e.g., HIV, liver disease). The small sample sizes for these subgroups may limit the ability of statistical tests to detect meaningful associations, potentially affecting the generalizability of our findings to these populations. Some of the features used to train the models are not measured for some patients, and hence the dataset is imbalanced, which may limit the performance of the model. To address this, we apply the Synthetic Minority Over-sampling Technique (SMOTE) to balance the dataset, which improves the reliability of the results. In future work, a large volume of datasets with more accurate and measurable features is recommended to be used to train the proposed models. Another limitation of this study is that we do not consider the impact of interventions imposed by administrations, healthcare system policies, and economic and sociodemographic situations, which may affect the predictive performance of our model. Though the data were reported and collected by the ministry's epidemiological surveillance team from thirteen governorates, the data from the other governorate were not available due to the conflict.

5. Conclusions

In this work, mathematical models based on time series approaches are applied to classify and predict the number of new cases and deaths in Yemen. Three predictive models are investigated to get the best mathematical models for forecasting new cases and deaths in Yemen. This helps the health authorities take the necessary precautions and manage their limited resources. The ARMA model shows more accurate results compared to the AR and MA models, which still showed acceptable results. Validation tests are carried out, including testing multiple assessment tools such as the RMSE, MAE, and the coefficient of determination (R^2). The ARMA model reveals an RMSE value of 21.51, an MAE value of 6.06, and a coefficient of determination R^2 score of 77%. Eight ML classifiers and three DL models are employed to identify indicators of COVID-19 severity. The DT classifier achieves the highest accuracy of 74.70%, followed by RF with 74.66%. The DL models show similar accuracy scores, approximately 70%. The kernel Support Vector Machine (SVM) outperform others in terms of AUC-ROC, with a classification accuracy of 71% and precision, recall, F measure, and area under the curve values of 0.7, 0.75, 0.59, and 0.72, respectively. The results suggest that the implementation of ML algorithms and time series analysis shows their significance in forecasting and diagnosing COVID-19 cases. Perhaps it can assist decision-makers in making efficient decisions for the early detection of viral respiratory tract infections, thereby significantly enhancing management and control. We provide an advanced level of analysis, which may be helpful in controlling the pandemic. In future work, we can apply more advanced techniques as well as algorithms for developing the model, which will improve and forecast more precisely.

In future work, we plan to integrate explainable AI (XAI) mechanisms to enhance model transparency and trust by aligning findings with clinical literature, addressing ethical considerations, and fostering acceptance among healthcare professionals and patients, thereby ensuring the clinical applicability of our models.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was funded by the Deanship of Scientific Research (DSR), Imam Abdulrahman Bin Faisal University, with Grant No. Covid19-2020-046-Eng.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. World Health Organization, Survey reveals extent of damage to Yemen's health system: Regional Office for the Eastern Mediterranean, 2016. Available from: https://www.emro.who.int/media/news/survey-reveals-extent-of-damage-to-yemens-health-system.html.
2. S. M. Mousavi, M. Anjomshoa, COVID-19 in Yemen: A crisis within crises, *Int. J. Equity Health*, **19** (2020), 1–3. https://doi.org/10.1186/s12939-020-01231-2
3. J. Cole, M. Alsabri, L. M. Alsakkaf, A. Alhadheri, M. Amin, B. Nightingale, Conflict, collapse and Covid-19: Lessons from Yemen on elevated disease risk in regions under stress, *RUSI J.*, **166** (2021), 10–19. https://doi.org/10.1080/03071847.2021.1952106
4. Save the Children, First case of COVID-19 reported by authorities in Yemen, 2020. Available from: https://www.savethechildren.net/news/first-case-covid-19-reportedauthorities-yemen.
5. ACAPS, COVID-19: Impact on Yemen: Risk report–Update, 2020. Available from: https://reliefweb.int/sites/reliefweb.int/files/resources/20200409_acaps_risk_report_covid-19_impact_on_yemen_update.pdf.
6. M. Noushad, I. S. Al-Saqqaf, COVID-19: Is herd immunity the only option for fragile Yemen, *Int. J. Infect. Dis.*, **106** (2021), 79–82. https://doi.org/10.1016/j.ijid.2021.03.030
7. I. Abuelaish, The effect of war, violence, and hatred on children's development, *Dev. Med. Child Neurol.*, **63** (2021), 1360. https://doi.org/10.1111/dmcn.15012
8. World Health Organization, Yemen, 2022. Available from: https://covid19.who.int/region/emro/country/ye.
9. M. K. Looi, COVID-19: Deaths in Yemen are five times global average as healthcare collapses, *BMJ*, **370** (2020), m2997. https://doi.org/10.1136/bmj.m2997
10. S. Devi, Fears of "highly catastrophic" COVID-19 spread in Yemen, *Lancet*, **395** (2020), 1683. https://doi.org/10.1016/S0140-6736(20)31235-6
11. A. A. Al-Waleedi, J. D. Naiene, A. A. Thabet, A. Dandarawe, H. Salem, N. Mohammed, et al., The first 2 months of the SARS-CoV-2 epidemic in Yemen: Analysis of the surveillance data, *PLoS One*, **15** (2020), e0241260. https://doi.org/10.1371/journal.pone.0241260

12. M. S. O. Baaees, J. D. Naiene, A. A. Al-Waleedi, N. S. Bin-Azoon, M. F. Khan, N. Mahmoud, et al., Community-based surveillance in internally displaced people's camps and urban settings during a complex emergency in Yemen in 2020, *Confl. Health*, **15** (2021), 1–15. https://doi.org/10.1186/s13031-021-00394-1
13. G. R. Shinde, A. B. Kalamkar, P. N. Mahalle, N. Dey, J. Chaki, A. E. Hassanien, Forecasting models for coronavirus disease (COVID-19): A survey of the state-of-the-art, *SN Comput. Sci.*, **1** (2020), 1–15. https://doi.org/10.36227/techrxiv.12101547.v1
14. J. Kurita, T. Sugawara, Y. Ohkusa, Forecast of the COVID-19 outbreak, collapse of medical facilities, and lockdown effects in Tokyo, Japan, *medRxiv*, 2020. https://doi.org/10.1101/2020.04.02.20051490
15. R. Gupta, S. K. Pal, Trend analysis and forecasting of COVID-19 outbreak in India, *medRxiv*, 2020. https://doi.org/10.1101/2020.03.26.20044511
16. D. Fanelli, F. Piazza, Analysis and forecast of COVID-19 spreading in China, Italy, and France, *Chaos Solitons Fractals*, **134** (2020), 109761. https://doi.org/10.1016/j.chaos.2020.109761
17. M. A. Al-Qaness, A. A. Ewees, H. Fan, M. A. El Aziz, Optimization method for forecasting confirmed cases of COVID-19 in China, *J. Clin. Med.*, **9** (2020), 674. https://doi.org/10.3390/jcm9030674
18. Z. Liu, W. Guo, Government responses matter: Predicting COVID-19 cases in the US under an empirical Bayesian time series framework, *medRxiv*, 2020. https://doi.org/10.1101/2020.03.28.20044578
19. G. Perone, An ARIMA model to forecast the spread and the final size of the COVID-19 epidemic in Italy, in *HEDG-Health Econometrics and Data Group Working Paper Series, University of York*, 2020. https://doi.org/10.2139/ssrn.3564865
20. P. Monllor, Z. Su, L. Gabrielli, P. Taltavull de La Paz, COVID-19 infection process in Italy and Spain: Are data talking? Evidence from ARMA and vector autoregression models, *Front. Public Health*, **8** (2020), 550602. https://doi.org/10.3389/fpubh.2020.550602
21. F. Petropoulos, S. Makridakis, Forecasting the novel coronavirus COVID-19, *PLoS One*, **15** (2020), e0231236. https://doi.org/10.1371/journal.pone.0231236
22. J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, M. Luengo-Oroz, Mapping the landscape of artificial intelligence applications against COVID-19, *J. Artif. Intell. Res.*, **69** (2020), 807–845.
23. V. Vytla, S. K. Ramakuri, A. Peddi, K. K. Srinivas, N. N. Ragav, Mathematical models for predicting COVID-19 pandemic: A review, *J. Phys. Conf. Ser.*, **1797** (2021), 012009. https://doi.org/10.1088/1742-6596/1797/1/012009
24. S. G. Kwak, J. H. Kim, Central limit theorem: The cornerstone of modern statistics, *Korean J. Anesthesiol.*, **70** (2017), 144–156. https://doi.org/10.4097/kjae.2017.70.2.144
25. I. Ciufolini, A. Paolozzi, An improved mathematical prediction of the time evolution of the COVID-19 pandemic in Italy, with a Monte Carlo simulation and error analyses, *Eur. Phys. J. Plus*, **135** (2020), 1–13. https://doi.org/10.1140/epjp/s13360-020-00488-4
26. C. J. L. Murray, Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days, and deaths by US state in the next 4 months, *medRxiv*, 2020. https://doi.org/10.1101/2020.03.27.20043752
27. R. Schlickeiser, F. Schlickeiser, A Gaussian model for the time development of the SARS-CoV-2 corona pandemic disease, *Physics*, **2** (2020), 164–170. https://doi.org/10.3390/physics2020010

28. J. Schüttler, R. Schlickeiser, F. Schlickeiser, M. Kröger, COVID-19 predictions using a Gauss model, based on data from April 2, *Physics*, **2** (2020), 197–212. https://doi.org/10.3390/physics2020013
29. P. Kumar, H. Kalita, S. Patairiya, Y. D. Sharma, C. Nanda, M. Rani, et al., Forecasting the dynamics of COVID-19 pandemic in top 15 countries in April 2020: ARIMA model with machine learning approach, *medRxiv*, 2020. http://dx.doi.org/10.1101/2020.03.30.20046227
30. S. I. Alzahrani, I. A. Aljamaan, E. A. Al-Fakih, Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions, *J. Infect. Public Health*, **13** (2020), 914–919. https://doi.org/10.1016/j.jiph.2020.06.001
31. C. Iwendi, C. Huescas, C. Chakraborty, S. Mohan, COVID-19 health analysis and prediction using machine learning algorithms for Mexico and Brazil patients, *J. Exp. Theor. Artif. Intell.*, **34** (2022), 621–642. https://doi.org/10.1080/0952813X.2022.2058097
32. C. Iwendi, K. Mahboob, Z. Khalid, A. R. Javed, M. Rizwan, U. Ghosh, Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system, *Multimed. Syst.*, **28** (2022), 1223–1237. https://doi.org/10.1007/s00530-021-00774-w
33. B. Hao, S. Sotudian, T. Wang, T. Xu, Y. Hu, A. Gaitanidis, et al., Early prediction of level-of-care requirements in patients with COVID-19, *eLife*, **9** (2020), e60519. https://doi.org/10.7554/eLife.60519
34. T. Wang, A. Paschalidis, Q. Liu, Y. Liu, Y. Yuan, I. C. Paschalidis, Predictive models of mortality for hospitalized patients with COVID-19: Retrospective cohort study, *JMIR Med. Inform.*, **8** (2020), e21788. https://doi.org/10.2196/21788
35. S. Bhattacharya, P. K. R. Maddikunta, Q. V. Pham, T. R. Gadekallu, C. L. Chowdhary, M. Alazab, et al., Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey, *Sustain. Cities Soc.*, **65** (2021), 102589. https://doi.org/10.1016/j.scs.2020.102589
36. D. Ferrari, J. Milic, G. Tonelli, G. Guaraldi, Machine learning in predicting respiratory failure in patients with COVID-19 pneumonia: Challenges, strengths, and opportunities in a global health emergency, *PLoS One*, **15** (2020), e0239172. https://doi.org/10.1371/journal.pone.0239172
37. F. Prinzi, C. Militello, N. Scichilone, S. Gaglio, S. Vitabile, Explainable machine-learning models for COVID-19 prognosis prediction using clinical, laboratory, and radiomic features, *IEEE Access*, **11** (2023), 121492–121510. https://doi.org/10.1109/ACCESS.2023.3327808
38. P. Soda, N. Claudia D'Amico, J. Tessadori, G. Valbusa, V. Guarrasi, C. Bortolotto, et al., AIforCOVID: Predicting the clinical outcomes in patients with COVID-19 applying AI to chest X-rays. An Italian multicentre study, *Med. Image Anal.*, **74** (2021), 102216. https://doi.org/10.1016/j.media.2021.102216
39. D. Wang, C. Huang, S. Bao, T. Fan, Z. Sun, Y. Wang, et al., Study on the prognosis predictive model of COVID-19 patients based on CT radiomics, *Sci. Rep.*, **11** (2021), 11591. https://doi.org/10.1038/s41598-021-90991-0
40. H. Abdeltawab, F. Khalifa, Y. ElNakieb, A. Elnakib, F. Taher, N. S. Alghamdi, et al., Predicting the level of respiratory support in COVID-19 patients using machine learning, *Bioengineering*, **9** (2022), 536. https://doi.org/10.3390/bioengineering9100536
41. T. Liu, E. Siegel, D. Shen, Deep learning and medical image analysis for COVID-19 diagnosis and prediction, *Annu. Rev. Biomed. Eng.*, **24** (2022), 179–201. https://doi.org/10.1146/annurev-bioeng-110220-012203

42. K. Moulaei, M. Shanbehzadeh, Z. Mohammadi-Taghiabad, H. Kazemi-Arpanahi, Comparing machine learning algorithms for predicting COVID-19 mortality, *BMC Med. Inform. Decis. Mak.*, **22** (2022), 1–12. https://doi.org/10.1186/s12911-021-01742-0
43. Y. Xiong, Y. Ma, L. Ruan, D. Li, C. Lu, L. Huang, et al., Comparing different machine learning techniques for predicting COVID-19 severity, *Infect. Dis. Poverty*, **11** (2022), 19. https://doi.org/10.1186/s40249-022-00946-4
44. A. AlMoammar, L. AlHenaki, H. Kurdi, Selecting accurate classifier models for a MERS-CoV dataset, *Proc. SAI Intell. Syst. Conf.*, Springer, Cham, (2018), 1070–1084. https://doi.org/10.1007/978-3-030-01054-6_74
45. X. Jiang, M. Coffee, A. Bari, J. Wang, X. Jiang, J. Huang, et al., Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity, *Comput. Mater. Continua*, **63** (2020), 537–551. https://doi.org/10.32604/cmc.2020.010691
46. R. N. Reddy, COVID-19 detection using SVM classifier, *Engpaper J.*, (2020).
47. C. Iwendi, S. Mohan, S. khan, E. Ibeke, A. Ahmadian, T. Ciano, COVID-19 fake news sentiment analysis, *Comput. Electr. Eng.*, **101** (2022), 107967. https://doi.org/10.1016/j.compeleceng.2022.107967
48. E. Besson, A. Norris, A. Bin Ghouth, T. Freemantle, M. Alhaffar, Y. Vazquez, et al., Excess mortality during the COVID-19 pandemic: A geospatial and statistical analysis in Aden governorate, Yemen, *BMJ Glob. Health*, **6** (2021), e004564. https://doi.org/10.1136/bmjgh-2020-004564
49. R. H. Shumway, D. S. Stoffer, *Time Series Analysis and Its Applications*, Springer, New York, 2000. https://doi.org/10.1007/978-1-4757-3261-0
50. R. J. Hyndman, G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed., OTexts, Melbourne, Australia, 2021.
51. D. J. Bartholomew, Time series analysis, forecasting, and control, *J. Oper. Res. Soc.*, **22** (1971), 199–201. https://doi.org/10.1057/jors.1971.52
52. J. M. Hilbe, *Logistic Regression Models*, 1st ed., Chapman and Hall/CRC, New York, 2009. https://doi.org/10.1201/9781420075779
53. T. Hastie, R. Tibshirani, J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, New York, 2009. https://doi.org/10.1007/978-0-387-84858-7
54. S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed., Prentice Hall, 2010.
55. S. Haykin, *Neural Networks and Learning Machines*, 3rd ed., Pearson Education, 2009.
56. A. S. Bin-Ghouth, S. Al-Shoteri, N. Mahmoud, A. Musani, N. M. Baoom, A. A. Al-Waleedi, et al., SARS-CoV-2 seroprevalence in Aden, Yemen: A population-based study, *Int. J. Infect. Dis.*, **115** (2022), 239–244. https://doi.org/10.1016/j.ijid.2021.12.330
57. R. Blagus, L. Lusa, Gradient boosting for high-dimensional prediction of rare events, *Comput. Stat. Data Anal.*, **113** (2017), 19–37. https://doi.org/10.1016/j.csda.2016.07.016
58. S. Nusinovici, Y. C. Tham, M. Y. C. Yan, D. S. W. Ting, J. Li, C. Sabanayagam, et al., Logistic Regression was as good as machine learning for predicting major chronic diseases, *J. Clin. Epidemiol.*, **122** (2020), 56–69. https://doi.org/10.1016/j.jclinepi.2020.03.002
59. V. G. Costa, C. E. Pedreira, Recent advances in Decision Trees: An updated survey, *Artif. Intell. Rev.*, **56** (2022), 1–36. https://doi.org/10.1007/s10462-022-10275-5

60. M. A. Araaf, K. Nugroho, D. R. I. M. Setiadi, Comprehensive analysis and classification of skin diseases based on image texture features using K-nearest Neighbors algorithm, *J. Comput. Theor. Appl.*, **1** (2023), 31–40. https://doi.org/10.33633/jcta.v1i1.9185
61. J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, A. Lopez, A comprehensive survey on support vector machine classification: Applications, challenges, and trends, *Neurocomputing*, **408** (2020), 189–215. https://doi.org/10.1016/j.neucom.2019.10.118
62. F. El Barakaz, O. Boutkhoum, A. El Moutaouakkil, A hybrid naïve Bayes based on similarity measure to optimize the mixed-data classification, *TELKOMNIKA Telecommun. Comput. Electron. Control*, **19** (2021), 155–162. https://doi.org/10.12928/telkomnika.v19i1.18024
63. A. Shahraki, M. Abbasi, Ø. Haugen, Boosting algorithms for network intrusion detection: A comparative evaluation of Real AdaBoost, Gentle AdaBoost, and Modest AdaBoost, *Eng. Appl. Artif. Intell.*, **94** (2020), 103770. https://doi.org/10.1016/j.engappai.2020.103770
64. K. E. ArunKumar, D. V. Kalaga, C. M. S. Kumar, M. Kawaji, T. M. Brenza, Forecasting of COVID-19 using deep layer recurrent neural networks (RNNs) with gated recurrent units (GRUs) and long short-term memory (LSTM) cells, *Chaos Solitons Fractals*, **146** (2021), 110861. https://doi.org/10.1016/j.chaos.2021.110861
65. A. I. Shahin, S. Almotairi, A deep learning BiLSTM encoding-decoding model for COVID-19 pandemic spread forecasting, *Fractal Fract.*, **5** (2021), 175. https://doi.org/10.3390/fractalfract5040175
66. D. Bergh, Sample size and Chi-squared test of fit: A comparison between a random sample approach and a Chi-square value adjustment method using Swedish adolescent data, in *Pacific Rim Objective Measurement Symposium (PROMS) 2014 Conference Proceedings*, Springer, (2015), 197–211. https://doi.org/10.1007/978-3-662-47490-7_15
67. L. Antonicelli, C. Tontini, G. Manzotti, L. Ronchi, A. Vaghi, F. Bini, et al., Severe asthma in adults does not significantly affect the outcome of COVID-19 disease: Results from the Italian Severe Asthma Registry, *Allergy*, **76** (2020), 902–905. https://doi.org/10.1111/all.14558
68. J. Hartmann-Boyce, J. Gunnell, J. Drake, A. Otunla, J. Suklan, E. Schofield, et al., Asthma and COVID-19: Review of evidence on risks and management considerations, *BMJ Evidence-Based Med.*, **26** (2021), 195–195. http://dx.doi.org/10.1136/bmjebm-2020-111506
69. W. M. S. Yafooz, A. H. M. Emara, M. Lahby, Detecting fake news on COVID-19 vaccine from YouTube videos using advanced machine learning approaches, in *Combating Fake News with Computational Intelligence Techniques*, Springer, Cham, (2022), 421–435. https://doi.org/10.1007/978-3-030-90087-8_21
70. J. N. Mandrekar, Receiver operating characteristic curve in diagnostic test assessment, *J. Thorac. Oncol.*, **5** (2010), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d
71. M. S. Alie, Y. Negesse, K. Kindie, D. S. Merawi, Machine learning algorithms for predicting COVID-19 mortality in Ethiopia, *BMC Public Health*, **24** (2024), 1728. https://doi.org/10.1186/s12889-024-19196-0
72. F. S. H. De Souza, N. S. Hojo-Souza, E. B. Dos Santos, C. M. Da Silva, D. L. Guidoni, Predicting the disease outcome in COVID-19 positive patients through machine learning: A retrospective cohort study with Brazilian data, *Front. Artif. Intell.*, **4** (2021), 579931. https://doi.org/10.3389/frai.2021.579931

73. K. Puttegowda, S. K. DS, S. Mallu, V. CP, V. Ravi, S. BC, Automatic COVID-19 prediction with comprehensible machine learning models, *Open Public Health J.*, **17** (2024). https://doi.org/10.2174/0118749445286599240311102956