Machine Learning-based Early Detection and Prognosis of the Covid-19 Pandemic

Ajitha Santhakumari*, R. Shilpa & Hudhaifa Mohammed Abdulwahab

Department of Computer Application, Ramaiah Institute of Technology,
(Affiliated to VTU), Bangalore 560054, Karnataka, India
*E-mail: ajithasankar@gmail.com

Abstract. The outbreak of Covid-19 has caused a global health crisis, presenting numerous challenges to the healthcare system with its severe respiratory symptoms and variable presentation. Early and accurate diagnosis of the virus is critical in controlling its spread and reducing the burden on healthcare facilities. To address this issue and relieve the strain on the healthcare system, this paper proposes a machine learning-based approach for Covid-19 diagnosis. Four algorithms were used for analyzing early Covid-19 detection, i.e., logistic regression, random forest, decision tree, and naive Bayes, using a data set of basic symptoms such as fever, shortness of breath, etc. for predicting positive and negative Covid-19 cases. Furthermore, development of a web portal that provides information on global vaccine distribution and the most widely used vaccines by country along with Covid-19 predictions. Our evaluation results demonstrate that the decision tree model outperformed the other models, achieving an accuracy of 97.69%. This study provides a practical solution to the ongoing Covid-19 crisis through an improved diagnosis method and access to vaccination information.

Keywords: *Covid-19; data preprocessing; healthcare; machine learning; predictive analysis; web tool.*

1 Introduction

Covid-19 is a highly contagious respiratory illness caused by a novel coronavirus. The first instance of the disease was reported in Wuhan, China in December 2019, and the virus is believed to have originated from bats [1]. The virus is highly contagious and has led to a global pandemic, with the virus being transmitted easily through airborne particles exhaled by infected individuals, especially when in close proximity for extended periods as well as through touching contaminated surfaces. Additionally, individuals can spread the virus before showing symptoms, making early detection and prevention crucial. The most common symptoms include cold, cough, fever, and shortness of breath, with older adults being most at risk. Currently, there is no cure for Covid-19, and vaccinations are being developed to curb its spread. The quick spread of the virus has resulted in a lack or delay of diagnosis kits, which has exposed patients, healthcare workers

Received December 16th, 2021, Revised February 26th, 2023, Accepted for publication May, 3rd, 2023.
 ISSN: 2337-5787, DOI: 10.5614/itbj.ict.res.appl.2023.17.2.6

as well as many others in the healthcare sector to the virus. This poses a great threat to populations all over the world.

In recent years, machine learning algorithms have been applied to many different fields to tackle various problems, such as medical diagnosis, treatment prediction, and drug discovery [2]. Machine learning algorithms in particular have shown remarkable performance in the early diagnosis and prediction of Covid-19, enabling early intervention and prevention of the spread of this disease [3]. Numerous research works have been done since the outbreak of Covid-19 using various AI techniques to detect and predict Covid-19 cases. For example, [4] tried to determine the most suitable algorithm for the early detection of Covid-19 based on extensive clinical data. Reference [5] developed a technique based on machine learning to detect the most vital and noteworthy clinical symptoms that can anticipate cases of Covid-19 that are truly positive. Reference [6] proposed a new prediction method for Covid-19 based on artificial intelligence, aimed at accurately predicting outbreaks and the spread of the virus. Although several predictive prognostic models have been proposed for the early detection of individuals at high risk of Covid-19 mortality, a major gap remains in examining the accuracy and reliability of using machine learning algorithms to improve patient categorization and predict Covid-19 cases. This study sought to fill this gap by evaluating the accuracy and effectiveness of four different machine learning algorithms in diagnosing and predicting Covid-19 cases. This study used a data set of Covid-19 patients to extract features, identify symptoms, and develop models to predict Covid-19 cases. The accuracy of these models was then evaluated using statistical measures such as area under the curve (AUC), precision, recall, and F1 score.

The results of this study have the potential to have a significant impact on the healthcare sector by improving the accuracy and reliability of Covid-19 diagnosis and prediction. This could help healthcare workers make better decisions regarding patient treatment and management as well as assist in controlling the spread of the virus. The findings of this study may also serve as a foundation for future research on using machine learning algorithms in diagnosing and predicting other diseases. This study makes several significant contributions. Firstly, it focused on the extraction of features from a data set consisting of Covid-19 patients with the aim of developing models that can accurately predict Covid-19 cases. Secondly, the method identifies and examines the various symptoms exhibited by Covid patients, employing four distinct machine learning approaches for analysis. Thirdly, it developed a machine learning model specifically designed for predicting Covid-19 patients. Additionally, it evaluated the effectiveness of four different machine learning algorithms in the diagnosis and prediction of Covid-19 cases. Finally, the authors conducted a thorough statistical analysis to assess the accuracy of the machine learning models, with precision, recall, and F-measure values as evaluation metrics. These contributions collectively enhance Covid-19 prediction and diagnosis, providing valuable insight for future research in the field.

The rest of this paper is organized as follows: related works and their limitations are discussed in Section 2, the methodology used in this study is described in Section 3, the data set analysis is presented in Section 4, the results of the experiments are presented in Section 5, data visualization of vaccination data is discussed in Section 6, the findings and their potential impact on the early detection and prediction of Covid-19 are discussed in Section 7, and finally, conclusions and recommendations for future research are presented in Section 8.

2 Related Work

A great deal of research work has been done since the outbreak of the pandemic to detect and predict Covid-19 cases using various AI (artificial-intelligence) techniques. Reference [4] tried to determine the most suitable algorithm for the early detection of Covid-19 based on extensive clinical data. Reference [5] developed a technique based on machine learning to detect the most important and noteworthy clinical symptoms to anticipate cases of Covid-19 that are truly positive. Reference [6] proposed a new prediction method for Covid-19 based on artificial intelligence, aimed at accurately predicting outbreaks and the spread of the virus. In [7], an epidemiological data set was considered and supervised machine learning models were used for the prediction of Covid infection; a data set from Mexico was used and a model was developed using five different algorithms. In [8], an ML model to recognize early-stage symptoms of coronavirus (Covid-19) infected patients. The features considered in this model were age, gender, fever, cough, and lung infection. Several algorithms were developed. The data was grouped according to the age factor, where the XGBoost algorithm showed the highest accuracy.

In [9], considering an epidemiological data set of Covid patients in South Korea, predictive data processing models of the recovery of people infected with Covid-19 were developed. The model predicted the minimum and maximum number of days for Covid patients to recover from the virus. It also predicted the age group of people who have less or more chance to recover from the disease. Reference [10], modified SEIR and AI prediction of the epidemic trend of Covid in China under public health interventions. This model was more accurate in predicting Covid peaks and outbreak size. Population movement data and epidemiological data were combined together into the SEIR Model to derive the Covid curve. Different machine learning (ML) techniques were used to train data to predict Covid outbreaks. With this approach, the clinical severity of Covid-19 was estimated based on the transmission dynamics in Wuhan, China. The different models developed estimate the probability of dying after developing symptoms of Covid-19 in Wuhan using public and published information.

The study reported in Reference [11] started with examining and analyzing the susceptible-infected-recovered (SIR) model, which is widely used for estimating the number of Covid-19 cases. In [12], a comprehensive parametric model was developed to evaluate and predict the impact of priority and age-specific vaccination policies and non-pharmacological policies. This model includes sub-models for susceptible, suspicious, infected, hospitalized, intensive care, intubated, recovered, and deceased patients, which incorporate these policies. Reference [13] developed a policy-making algorithm that utilizes artificial intelligence to generate phased reopening policies for schools. The algorithm aimed to optimize attendance while minimizing pandemic casualties, even in worst-case scenarios with significant uncertainties. Some more references related to the detection and prediction of Covid-19 can be found in [14 –19].

3 Methodology

In this section, we will first discuss the methodology used in this work and then describe the tools.

3.1 Architecture

This work used machine learning models to analyze and predict Covid-19 symptoms at an early stage. Figure 1 presents the methodology that was followed in this work. Using the training data from the specified data set, we trained four machine learning (ML) algorithms.

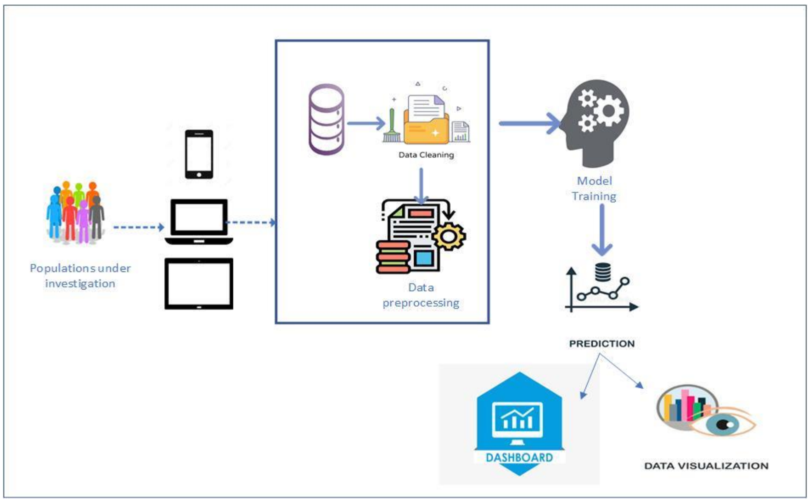


Figure 1 Proposed methodology.

3.2 Description of the Tool

The main objective of this work was to develop a web-based interactive tool. Figure 2 presents the dashboard architecture. The main strategy was to make the dashboard user-friendly for the users of this application. Using a machine learning approach to predict if a patient is Covid positive or negative helps to reduce the burden on healthcare workers and the healthcare system. The data visualization technique used in this application is related to vaccination data. This data visualization helps the user to know more details on different kinds of vaccines used all over the world, the number of people who have been vaccinated per day, and the country that ranks first in getting its people vaccinated. The front end of the tool was designed using FLASK technology, which is a Python framework. Some of snapshots of the tool are shown in Figure 3.

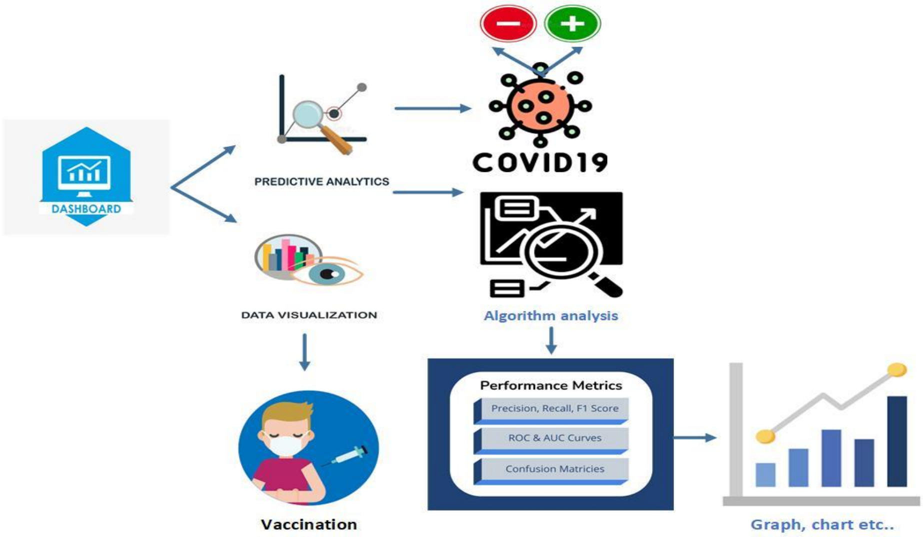


Figure 2 Dashboard architecture.

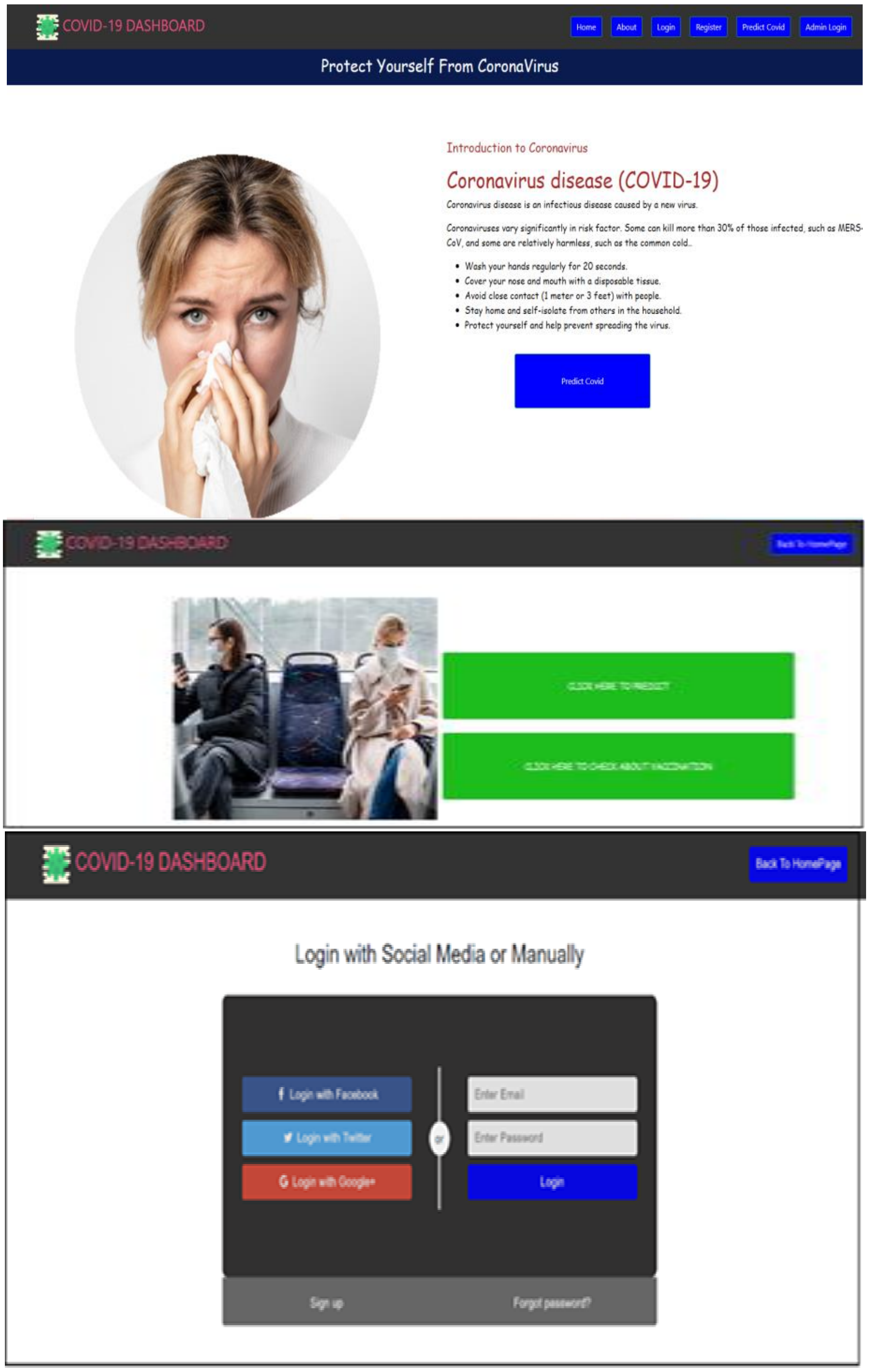


Figure 3 Ssnapshots of the tools.

4 Analysis of Covid-19 Data Set

4.1 Data Cleaning and Pre-Processing

The data set that was utilized for the prediction of Covid-19 disease was taken from the GitHub repository. The data set was a continuous data set and was therefore well suited for regression analysis as it needs to make predictions based on continuous dependent variables from several independent ones. The relation between dependent and independent variables can be defined by the coefficient of both variables in a regression mathematical statement. The mutual information method was used for feature selection. We used the mutual information score to measure the relevance of each feature to the target variable [20]. We then selected the top features based on their scores [21]. We used the following features: age, gender, and the following medical features: cough, fever, sore throat, shortness of breath, headache, diabetics, asthma, and sinusitis. The data set encoded ‘0’ for negative results and ‘1’ for positive results.

The original data set contained a gender string. For the purpose of this work, it was converted into integers (0,1). The original data set was encoded with ‘0’ and ‘1’ for clinical values, ‘0’ indicating the absence of symptoms and ‘1’ indicating the presence of symptoms. The data set also had a value ‘NONE’, which also indicated the absence of symptoms. For the purpose of this work, it was replaced with ‘0’. The gender field also considered ‘NONE’, which was randomly replaced with ‘male’ and ‘female’.

The age field only contained ‘yes’ and ‘no’, where ‘yes’ stood for people above the age of 60 and ‘no’ stood for people below the age of 60. This field also had ‘NONE’, which was randomly replaced with ‘yes’ and ‘no’ values. The date field and the test indication field from the data set were removed as no work was done using the date field. There were no missing values in the data set. Lastly, we produced the final data set, which had the following features: age, gender, cough, fever, shortness in breath, sore throat, corona results, other problems (diabetics, asthma, sinusitis).

4.2 Dataset

Table 1 presents the features that were considered for building the machine learning model to predict Covid-19. These features are crucial for analyzing and predicting Covid-19 cases. The features were categorized into two types: string and Boolean. The string features were gender, or sex, and test indication. The Boolean features were age group, high fever, dry cough, sore throat, shortness of breath, and headache.

Table 1 Description of the features.

Features considered	Type of the feature	Description of the feature
Gender, or sex	String	Similar proportions of male and female patients were considered
Age group	Boolean	Age group factor for patients above 60 years and below 60 years (yes or no).
High fever	Boolean	Patients developing a high fever or body temperature.
Dry cough	Boolean	Patients developing a dry cough
Sore throat	Boolean	Patients developing symptoms of a sore throat (no taste).
Shortness of breath	Boolean	Patients developing symptoms of a breathing problem or low oxygen level.
Headache	Boolean	Patients developing the symptom of a headache.
Test indication	String	Condition how a person was detected positive (travel history, in contact with Covid-19 infected patient).
Other problems	String	Patients who also have other health issue like asthma, diabetes, sinusitis).

4.3 Evaluation Criteria Used for Comparison of Different Models

The various parameters considered for evaluation purposes in the different models used were the following.

4.3.1 Precision

Precision is a recognized way of finding an assessment metric when there is extreme positivity about the prediction. It is calculated by measuring the ratio of predicted positives that are true positives. Therefore, it depends on true positive values and false positive values. Precision is calculated as given by Eq. (1):

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

4.3.2 Recall

Recall was also considered as an evaluation criterion in this work, i.e., allowable decision on the metric when there is an actual need to identify the number of positives that can be expected. This shows us the proportion of real positives that are correctly and accurately classified. The numbers of real positives and false negatives are used to measure recall. Recall is calculated as given in Eq. (2):

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

4.3.3 F1 Score

The F1 score balances the consonance between the precision values obtained and the recall values obtained for the classification. The F1 score lies between the range of 0 and 1 and is compatible with the precision and recall values. The F1 score is calculated as given in Eq. (3):

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

4.3.4 Area Under the Curve (AUC)

AUC and ROC were used to measure the performance for the classification models using various threshold settings. The probability was measured by ROC while AUC measured the degree of separability. If the value of AUC is higher, the model is better at predicting the classes, i.e., it predicts the classes of ‘0’ as 0 and classes of ‘1’ as 1. True positive rate is the only range of true values that are utilized. The AUC was calculated as given in Eq. (4).

$$Sensitivity = TPR(True\ Positive\ Rate) = Recall \quad (4)$$

5 Experiment Using Different Methodology

5.1 Logistic Regression

The most popular machine learning algorithm amongst the supervised machine learning techniques is the logistic regression algorithm. Given a set of independent variables, categorical dependent variables can be easily predicted using this technique. A categorical variable is usually discrete in nature and its output can easily be predicted. The output of this contains a Boolean value, which can be either ‘yes’ or ‘no’, ‘0’ or ‘1’, ‘true’ or ‘false’, etc., but this does not give an exact value of 0 or 1. Rather, the output obtained will be the probability of the data being between 0 and 1. The only difference between logistic regression and linear regression is that the linear regression model is used only for solving regression problems while the logistic regression model is used for solving classification or categorical problems. The function produced in logistic regression is S shaped, i.e., a sigmoid function, which is used as a fitting curve whose output lies between 0 and 1.

Logistic regression is calculated as given in Eq. (5):

$$f(x) = \frac{\mathbf{1}}{\mathbf{1+e-(x)}} \quad (5)$$

For the developed logistic regression model, we obtained an accuracy of 96.33%, followed by precision, recall, F1 score with 96.05%, 99.73% and 97.86%, respectively.

Figures 4 and 5 show a pictorial representation of the above-mentioned values:

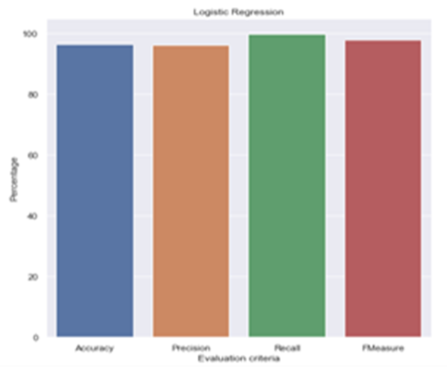


Figure 4 Values of evaluation criteria for logistic regression.

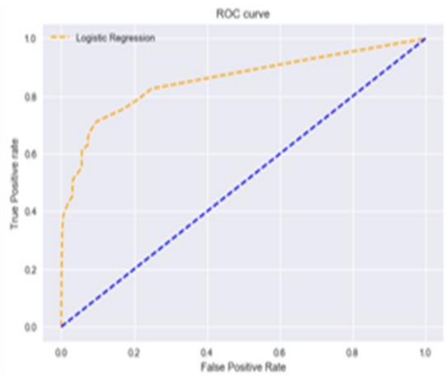


Figure 5 ROC curve for logistic regression.

5.2 Decision Tree

Decision tree is one of the machine learning algorithms that belong to the family of supervised machine learning algorithms. The decision tree technique is used to partition activities where a tree is constructed by portioning the data set into smaller units. The data set is divided into smaller or minute chunks until each partition or smaller chunk is clean as well. Classification of the data depends on the type or format of the data being used.

Decision tree algorithms can be used for classification and data regression problems. The technique uses a tree-like representation; this tree format consists of leaf nodes that correspond to a group of attributes and branches that correspond to a value. The technique is developed in an iterative manner. Statistical properties measure how well the given attribute divides the training data set, which is known as information gain. The attribute that has high information gain can split the training data to improve the accuracy.

For the developed decision tree model, we obtained an accuracy of 96.98%, followed by precision, recall, F1 score at 96.17%, 99.71% and 97.91% respectively.

Figures 6 and 7 show a pictorial representation of the above-mentioned values:

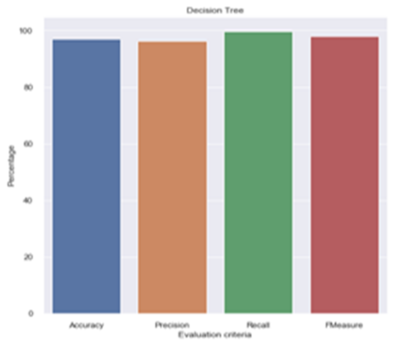


Figure 6 Values of evaluation criteria for decision tree.

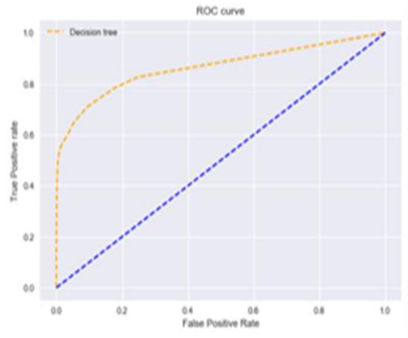


Figure 7 ROC curve for decision tree.

5.3 Random Forest

Random forest is another machine learning algorithm that is flexible and easy to use. Random forest is a combination of both regression as well as classification techniques, which has the capacity to train similar size data sets called bootstraps, and finally put them together for a more accurate result. Re-sampling of the data set enables us to create the bootstraps. The performance of random forest is much better than a single tree. Random forest works with large data sets with greater and better accuracy. For the developed random forest model, we obtained an accuracy of 96.36%, followed by precision, recall, F1 score with 95.73%, 99.88% and 97.76%, respectively.

Figures 8 and 9 show a pictorial representation of the above-mentioned values.

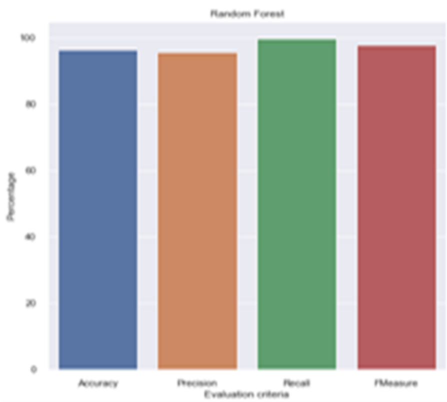


Figure 8 Values of evaluation criteria for random forest.

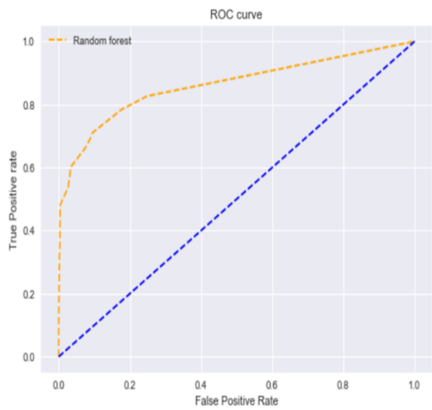


Figure 9 ROC curve for random forest.

5.4 Naïve Bayes

Another machine learning algorithm considered in this work was the naïve Bayes algorithm, which is used on tasks or work based on classification learning. The tasks performed are ones in which occurrences of the data set are separated based on features. The naïve Bayes model is based on Bayes' theorem and is probabilistic in nature. The mathematical equation involved in naïve Bayes is:

$$P(A|C) = \frac{P(C|B)P(B)}{P(C))} \quad (6)$$

For the developed naïve Bayes model, we obtained an accuracy of 94.32%, followed by precision, recall, F1 score with 96.27%, 99.30% and 97.76%, respectively.

6 Data Visualization for Vaccination Data

Figures 10 and 11 show a pictorial representation of the above-mentioned values:

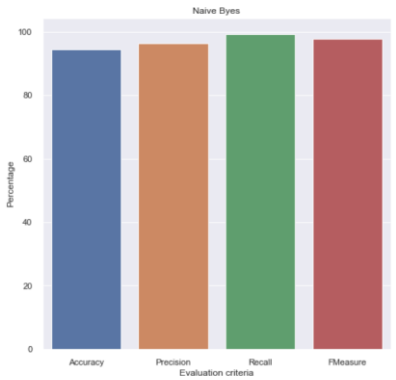


Figure 10 Values of evaluation criteria for naïve Bayes.

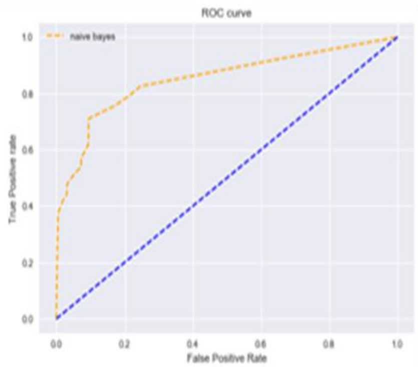
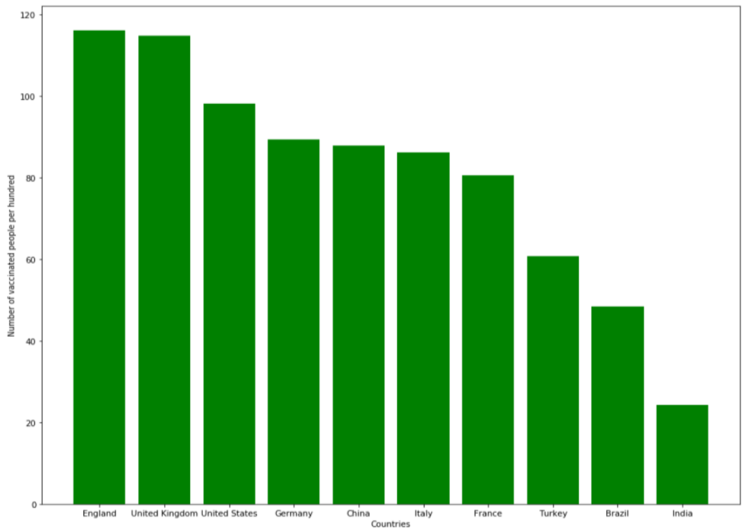


Figure 11 ROC curve for naïve Bayes.

Data visualization using machine learning plays an important role in understanding how the data is used for a specific machine learning model, which helps in analyzing the data. With a huge amount of data available, taking useful information from the available data will help people in getting clear insight into different areas. Data visualization using machine learning gives a high-level statistical overview on the data and its attributes. In this work, data analysis and visualization were done for the available data.

Figures 12-14 display the results obtained from the vaccination data [December 2020-July 2021].



Total number of vaccinated people per hundred.

Figure 12 Number of vaccinated people across the world.

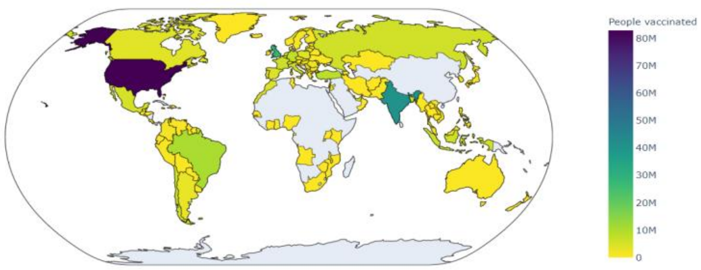


Figure 13 Number of vaccinated people across the world.

Figure 14 Countries using different vaccines.

7 Results and Discussion

Early-stage diagnosis and precautionary measures of Covid-19 disease are helpful in reducing the huge burden on the health, healthcare, and economic sectors in different parts of the world. This can be done by diagnosing coronavirus tested patients (positive-tested patients). In this work, decision tree, logistic regression, naïve Bayes, and random forest algorithms were used for prediction of Covid-19 infection.

The data set that was considered, which contained positive as well as negative Covid cases, was divided into training and testing sets. The developed models were trained with 80% of the data and tested with the remaining 20% of the data. Four different types of models were developed and trained using different machine learning (ML) approaches to predict whether patients were infected with Covid or not.

The models were assessed using a performance evaluation metric and also the precision, recall and F1 score values were determined to evaluate their efficiency and performance. Table 2 illustrates the final outcome produced by all four algorithms.

Table 2 Evaluation results.

SL NO.	Model	Accuracy	Precision	Recall	F-Measure
1	Decision tree	96.98	96.17	99.71	97.91
2	Logistic regression	96.33	96.05	99.73	97.86
3	Random forest	96.36	95.73	99.88	97.76
4	Naïve Bayes	94.32	96.27	97.30	97.76

Among the different models developed, the model developed with decision tree was the best model, with an accuracy of 96.98% when compared with the other three models (logistic regression, naive Bayes, and random forest), which had an accuracy of 96.33%, 94.32%, and 96.36%, respectively.

8 Conclusion

Covid-19 cases are rapidly increasing, which constitutes a great threat to global health. The major key to stop the pandemic and the spread of the virus is to identify the affected individuals as early as possible. Machine learning algorithms provide good and efficient results that can be applied to the current situation (Covid-19 situation). We developed a web-based interactive tool that shows visualizations of global vaccination data. In the future, we would like to make this website available to the public with the statistics dynamically updated according to the latest data available.

References

[1] Bo, W., Ahmad, Z., Alanzi, A.R., Al-Omari, A.I., Hafez, E.H. & Abdelwahab, S.F., *The Current Covid-19 Pandemic in China: An Overview and Corona Data Analysis*, Alexandria Engineering Journal, **61**(2), pp. 1369-1381, 2022. DOI: 10.1016/j.aej.2021.06.025.

[2] Vaishya, R., Javaid, M., Khan, I.H. & Haleem, A., *Artificial Intelligence (AI) Applications for Covid-19 Pandemic*, Diabetes & Metabolic Syndrome: Clinical Research & Reviews, **14**(4), pp.337-339, 2020. DOI: 10.1016/j.dsx.2020.04.012.

[3] Mei, X., Lee, H.C., Diao, K.Y., Huang, M., Lin, B., Liu, C., Xie, Z., Ma, Y., Robson, P.M., Chung, M. & Bernheim, A., *Artificial Intelligence–enabled Rapid Diagnosis of Patients with Covid-19*, Nature Medicine, **26**(8), pp.1224-1228, 2020. DOI: 10.1038/s41591-020-0931-3.

[4] Sun, N.N., Yang, Y., Tang, L.L., Dai, Y.N., Gao, H.N., Pan, H.Y. & Ju, B., *A Prediction Model Based on Machine Learning for Diagnosing the Early Covid-19 Patients*, MedRxiv, pp. 2020-06, 2020. DOC: 10.1101/2020.06.03.20120881.

[5] Ahamad, M.M., Aktar, S., Rashed-Al-Mahfuz, M., Uddin, S., Liò, P., Xu, H., Summers, M.A., Quinn, J.M.W. & Moni, M.A., *A Machine Learning Model to Identify Early-Stage Symptoms of SARS-Cov-2 Infected Patient*, Expert Systems with Applications, **160**, 113661, 2020. DOI: 10.1016/J.Eswa.2020.113661.

[6] Jing, S., Qian, Q., She, H., Shan, T., Lu, S., Guo, Y. & Liu, Y., *A Novel Prediction Method Based on Artificial Intelligence and Internet of Things for Detecting Coronavirus Disease (Covid-19)*, Security and Communication Networks, **2021**, pp. 1-11, 2021. DOI: 10.1155/2021/7812223.

[7] Muhammad, L.J., Algehyne, E.A., Usman, S.S., Ahmad, A., Chakraborty, C. & Mohammed, I.A., *Supervised Machine Learning Models for Prediction of Covid-19 Infection Using Epidemiology Dataset*, SN Computer Science, **2**, pp. 1-13, 2021. DOI: 10.1007/s42979-020-00394-7.

[8] Ahamad, M.M., Aktar, S., Rashed-Al-Mahfuz, M., Uddin, S., Liò, P., Xu, H. & Moni, M.A., *A Machine Learning Model to Identify Early-Stage Symptoms of SARS-Cov-2 Infected Patients*, Expert Systems with Applications, **160**, 113661, 2020. DOI: 10.1016/j.eswa.2020.113661.

[9] Jibril, M.L., Islam, M.M., Sharif, U.S. & Ayon, S.I., *Predictive Data Mining Models for Novel Coronavirus (Covid-19) Infected Patients' Recovery*, SN Comput. Sci., **1**(4), 206, 2020. DOI: 10.1007/s42979-020-00216-w.

[10] Yang, Z., Zeng, Z., Wang, K., Wong, S.S., Liang, W., Zanin, M. & He, J., *Modified SEIR and AI prediction of the epidemics trend of Covid-19 in China Under Public Health Interventions*, Journal of Thoracic Disease, **12**(3), 165, 2020. DOI: 10.21037/jtd.2020.02.64.

[11] Tutsoy, O., Çolak, Ş., Polat, A., & Balikci, K., *A Novel Parametric Model for the Prediction and Analysis of the Covid-19 Casualties*, IEEE Access, **8**, pp. 193898-193906, 2020. DOI: 10.1109/ACCESS.2020.3033146.

[12] Tutsoy, O. & Tanrikulu, M.Y., *Priority and Age Specific Vaccination Algorithm for the Pandemic Diseases: A Comprehensive Parametric Prediction Model*, BMC Medical Informatics and Decision Making, **22**(1), 4, 2022. DOI: 10.1186/s12911-021-01720-6.

[13] Tutsoy, O., *Covid-19 Epidemic and Opening of the Schools: Artificial Intelligence-Based Long-Term Adaptive Policy Making to Control the Pandemic Diseases*, IEEE Access, **9**, pp. 68461-68471, 2021. DOI: 10.1109/ACCESS.2021.3078080.

[14] Naz, R., & Al-Raeei, M., *Analysis of Transmission Dynamics of COVID-19 via Closed-form Solutions of a Susceptible-infectious-quarantined-diseased Model with a Quarantine-adjusted Incidence Function*, Mathematical Methods in the Applied Sciences, **44**(14), pp. 11196-11210, 2021. DOI: 10.1002/mma.7481.

[15] Aktar, S., Ahamad, M.M., Rashed-Al-Mahfuz, M., Azad, A.K.M., Uddin, S., Kamal, A.H.M., Alyami, S.A., Lin, P.-I., Islam, S.M.S., Quinn, J. M., Eapen, V. & Moni, M.A, *Machine Learning Approach to Predicting Covid-19 Disease Severity Based on Clinical Blood Test Data: Statistical Analysis and Model Development*, JMIR Medical Informatics, **9**(4), E25884, 2021. DOI: 10.2196/25884.

[16] Auwul, M. R., Rahman, M. R., Gov, E., Shahjaman, M., & Moni, M. A., *Bioinformatics and Machine Learning Approach Identifies Potential Drug Targets and Pathways in Covid-19*, Briefings in Bioinformatics, **22**(5), bbab120, 2021. DOI: 10.1093/bib/bbab120.

[17] Kivrak, M., Guldogan, E. & Colak, C., *Prediction of Death Status on the Course of Treatment in SARS-COV-2 Patients with Deep Learning and Machine Learning Methods*, Computer Methods and Programs in Biomedicine, **201**, 105951, 2021. DOI: 10.1016/j.cmpb.2021.105951.

[18] Aydin, N., & Yurdakul, G., *Assessing Countries' Performances Against Covid-19 Via WSIDEA And Machine Learning Algorithms*, Applied Soft Computing, **97**, 106792, 2020. DOI: 10.1016/j.asoc.2020.106792.

[19] Yadav, M., Perumal, M., & Srinivas, M., *Analysis on Novel Coronavirus (Covid-19) Using Machine Learning Methods*, Chaos, Solitons & Fractals, **139**, 110050, 2020. DOI: 10.1016/j.chaos.2020.110050.

[20] Peng, H., Long, F. & Ding, C., *Feature Selection Based on Mutual Information Criteria of Max-Dependency, Max-Relevance, and Min-*

Redundancy, IEEE Transactions on Pattern Analysis and Machine Intelligence, **27**(8), pp. 1226-1238, 2005. DOI: 10.1109/TPAMI.2005.159

[21] Abdulwahab, H.M., Ajitha, S. & Saif, M.A.N, *Feature Selection Techniques in the Context of Big Data: Taxonomy and Analysis*, Applied Intelligence, **52**(12), pp. 13568-13613, 2022. DOI: 10.1007/s10489-021-03118-3.