Eng. Sci., 2021, **16**, 221–233

Engineered Science

DOI: https://dx.doi.org/10.30919/es8d579

COVID-19 Mortality Prediction among Patients Using Epidemiological Parameters: An Ensemble Machine Learning Approach

Krishnaraj Chadaga,[1] Srikanth Prabhu,[1*] Shashikiran Umakanth,[2] Vivekananda Bhat K,[1] Niranjana Sampathila,[3] Rajagopala Chadaga P[4] and Krishna Prakasha K[5]

Abstract

Coronavirus infection (COVID-19) is a dangerous disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that has quickly spread all around the world, becoming a global pandemic on 11th March 2020. Vaccines have been developed to prevent the spread of this disease and various researches are being conducted to find the cure too. Machine learning (ML) has shown to be useful in battling COVID-19 and various applications have been deployed to comprehend real-world events through the meticulous analysis of data. In this study, we perform a retrospective study of epidemiological parameters to predict the mortality among SARS-CoV-2 patients. The goal of this research is to find important predictive parameters that can indicate the patients who are at the highest risk of death. Supervised ensemble machine learning models were developed that included random forest, catboost, adaboost, gradient boost, extreme gradient boosting and light GBM (Gradient Boosting Machine) for the COVID-19 epidemiology dataset that was obtained from Mexico. Prior to creating the models, Pearson's co-relation and mutual information analysis between various dependent and independent features were used to establish the strength of the association between features in the dataset. Extreme Gradient Boosting achieved the highest results with an accuracy of 96%.

Keywords: COVID-19; Ensemble; Epidemiology; Machine Learning; Mortality.
Received: 11 November 2021; Accepted: 1 December 2021.
Article type: Research article.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus coronavirus infection (COVID-19) is a highly contagious respiratory disorder that has caused chaos all around the world.[1] As of October 13 2021, there have been 237,655,302 confirmed cases and 4,846,981 deaths according to the World Health Organization.[2] COVID-19 patients may suffer from acute respiratory disorder syndrome (ARDS) and may also cause multi organ failure in rare cases.[3] According to recent estimates, the rate of COVID-19 infection related hospitalization ranges from 10.7% to 11.4%. The rate of Intensive Care Unit (ICU) admission varies between 4.9-11.5%.[4] There is also a 2.5% death rate among confirmed cases.[5] The rapid rise in COVID-19 patients has resulted in increased demand for medical supplies and ICU admittance. The COVID-19 patients become critically ill and symptoms progress rapidly and a lot of sudden deaths even after vaccination are being still documented. The significance of this study is in finding the features that help in the prediction of mortality that can be utilized to identify COVID-19 patients whose health-condition is likely to deteriorate early. Recent studies have proved that machine learning (ML) is used in all aspects of medical sciences in generating new clinical knowledge[6-8] as well as in clinical decision support. Numerous research applications using ML techniques for COVID-19 diagnosis, ICU prediction have been developed.[9-10] The number of publications regarding the use of artificial intelligence (AI) in battling COVID-19 has been increasing rapidly.[11-13] There are numerous predictors that can indicate in

[1] *Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, 576104, India.*
[2] *Department of Medicine, Dr. TMA Pai Hospital, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India.*
[3] *Department of Biomedical Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India.*

predicting the severity/mortality of COVID-19 patients. These include chest X-rays, CT-Scans, demographics, hematological parameters, blood and laboratory markers, cough sounds, hypoxia *etc*.

The deadly COVID-19 has symptoms that are extremely similar to the normal flu such as cough, fever and nasal congestion. Other symptoms, including a loss of smell and taste (anosmia) have surfaced as this infection spreads. Severe cases might result in respiratory illness and pneumonia. Older people and people with underlying medical conditions such as diabetes and cardiovascular disease are very vulnerable. More signs and traits that impact patient mortality are becoming apparent as the virus spreads all over the world. The disease impacts many different characteristics and features. Hence, it's difficult to determine which elements have a higher impact on patient death. Machine learning can help by evaluating massive datasets to uncover various trends and generating models that accurately estimate risk variability.

Most ML researches that predicted COVID-19 used only simple ML models. The ensemble models are known to give better results than the conventional ones due to the use of "bagging" and "boosting" techniques. The results were also based on very few parameters and markers. In this research, advanced ensemble algorithms and nineteen epidemiological parameters a have been used.

The objectives addressed in this study are: (a) An in-depth data exploration that finds crucial epidemiological parameters that indicate COVID-19 mortality. (b)Heterogenous ML models that accurately predicts mortality among SARS -CoV-2 patients. (c) The use of explainable AI to obtain feature importance among all the input parameters. It was found that age and the presence of pneumonia increased the risk of patient death. (d) Substantial information about various demographic and epidemiological parameters that are important to prevent in-hospital COVID-19 mortality.

In this research, an epidemiological labelled dataset from Mexico consisting of COVID-19 cases was used to predict the mortality of patients. The blood markers details can be easily collected from patients admitted in hospitals. Furthermore, these attributes are regarded as potential indicators that may be utilized to depict the patient's status in quantitative ways that the ML models can easily learn, making the models more reliable. Unlike many researches, where typical classification algorithms have been used, we use ensemble models for our prediction. The novelty also lies in the usage of feature importance techniques that explains the predictions of the model. Ensemble learning combines the predictions of multiple models that generate a better predictive performance. Xgboost, random forest, adaboost, light gradient boosting machine (GBM), gradient boosting and catboost models have been used for COVID-19 casualty prediction in this paper. The models were evaluated using accuracy, precision, recall, f1-score and area under curve (AUC).

Several researches have been published to predict the severity and mortality in COVID-19 patients using laboratory, hematological and epidemiological parameters with a high level of accuracy. Machine learning can predict better and faster, and the above methods can also be used with X-rays and CT-Scans as a common method of assessment to monitor the progress of the infection. They can also be used as severity indicators of COVID-19 and prediction of mortality can be effectively done in advance so that accurate treatments can be given to the patients to avoid casualties.

A previous study by Shoen *et al*.,[14] explained the use of epidemiological parameters to predict COVID-19 mortality in India. They concluded that the most important factors were age and gender. Other complications such as hypertension, diabetes, cardiovascular and cerebrovascular disease, asthma etc., also played a very significant role. Results concluded that diabetes contributed to 53% casualty followed by hypertension with 33%. Nunung *et al*.,[15] developed a mortality prediction model using tree-based algorithms using routine blood parameters. They used Explainable AI too and chose a set of 11 biomarkers that predicted COVID-19 deaths based on the 1000 blood samples of coronavirus patients from Jakarta, Indonesia. XG boost achieved the best results with an accuracy, precision, recall and F1-score of 98%, 96%, 99% and 98%.

In another study by Hu *et al*.,[16] early mortality risk prediction of SARS -CoV-2 patients using ML and laboratory markers was conducted. Both demographic and laboratory findings were used to develop models on 183 patients (115 survived, 68 died) from Tongji Hospital, Wuhan. Age, CRP (C-reactive protein), D-dimer and lymphocyte count were selected as attributes for the five baseline models. The simplicity of logistic regression made the researchers use it as the final model. The AUC's (Area under curve) of the external dataset was 88%. The sensitivity and specificity obtained were 83% and 79.4% respectively. A Decision support system using ML for early prediction of COVID-19 mortality was developed.[17] Five important features: LDH (lactate dehydrogenase), lymphocytes, neutrophils, CRP and age helped in predicting deaths with 90% accuracy. XG boost was the best performing model since it predicted as early as 16 days before the outcome.

Moreover, Sanchez-Monteras *et al*.,[18] used both supervised and unsupervised algorithms to predict the survival rate among COVID-19 patients in Madrid, Spain. Important variables such as age and oxygen saturation were considered. Interpretable decision rules for mortality risk was obtained from the decision tree model and it was concluded that pritorization of resources and medicines could be effectively

[4]*Department of Mechanical and Manufacturing Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education. Manipal, Karnataka 576104, India.*

[5]*Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India.*

**E-mail:* srikanth.prabhu@manipal.edu (S. Prabhu)

done using this model. Random forest achieved the best results with an AUC, sensitivity and specificity of 90%, 87% and 79% respectively. In one more study, Li *et al.*,[19] collected clinical data from a hospital in Wuhan, China to predict the mortality in severe COVID-19 patients. Gradient boosting decision tree (GBDT) and logistic regression models were developed for the above scenario. The former exhibited the highest accuracy of 89%. Leukomonocyte, urea, age and oxygen were the most important markers for the above dataset.

A broad learning system (BLS) for mortality prediction in SARS-CoV-2 patients was developed.[20] Three models were developed based on the blood samples of 375 patients. The BLS model achieved a sensitivity and specificity of 94.5% and 94.8% respectively. They also concluded their novel algorithm is better than the xgboost and support vector machine models and offers a more reliable and accurate prediction. To forecast COVID-19 mortality in South Korea, ML algorithms were developed to help decision making.[21] A total of 4004 patients were chosen for training and testing the model. The logistic regression model achieved the best results with AUC, Mathew Co-relation and brier score of 83%, 0.43, 0.036 respectively. Five serum chemistry biomarkers (CRP, blood urea nitrogen, lactic acid, albumin and calcium, were used to predict the mortality in 398 COVID-19 patients from Texas.[22] Support vector machine achieved the highest results with severity, specificity and AUC of 91%, 91% and 93% respectively.

2. Materials and Methods

2.1 Dataset Description

An epidemiology dataset containing the details of both COVID-19 positive and negative cases from Mexico was used in this research. This was made available by the directorate of epidemiology, secretariat of health, Mexico and is freely accessible on their website and Kaggle.[23] The data was acquired from the viral respiratory diseases epidemiological surveillance system and a total of 475 respiratory disease monitoring units (USMER) from all around the country had contributed to this dataset. To totally anonymize the data, best practices and standards were employed. This collection of data provides the findings of RT-PCR testing from different regions all over the country. A total of 263,007 patients were tested. It also contained 41 different clinical and demographic attributes. The entire epidemiological dataset that contained 41 features of COVID-19 patients was saved in a comma separated values (CSV) file in Spanish. Hence, the names of clinical and demographic parameters were translated to English for better understanding. For our research, we chose eight demographic features that included age, sex, type of care, patient death date, pregnancy, type of contact, ICU and Intubation status. Clinical features such as the presence of Pneumonia, diabetes, EPOC (Excess post exercise oxygen consumption), asthma, autoimmune disease, hypertension, obesity, cardiovascular diseases (CVD's), chronic kidney diseases (CKDS) were included. We also included a high-risk factor called "tobacco". RT-PCR results were also included and it consisted of dichotomous values (1- COVID-19 positive, 2-COVID-19 negative). Since our research is based on predicting mortality rates, we created a new feature "Death" from an existing feature "Patient death date". If the patient died, the exact date would be mentioned and if the patient survived, the date "99-99-9999" would be mentioned as default. Hence all the patient columns that did not have the value "99-99-999" were considered dead. These patients were assigned with the value "1". The patients who survived were assigned with the value "0".

The default categorical features in the dataset were encoded as 1 for positive and 2 for negative. However, for this research we encoded 1 as positive and 0 as negative across all attributes. Fortunately, this dataset doesn't suffer from missing values. Fig. 1 shows the age, gender and mortality distribution of the patients. From the diagram, it can be concluded that only a small population died due to coronavirus and most of the patients recovered. However, machine learning algorithms require a balanced dataset. Therefore, we use a technique called borderline synthetic minority oversampling technique (SMOTE) to super sample the minority class instances (deaths). Except for a single column "Age", the dataset was already normalized. In some algorithms, the magnitude of the attribute can have a significant impact on the accuracy of the output. To avoid the impact of variables with different scales, the age parameter was normalized using the minmax scalar available in the python library. After data preprocessing 263,007 rows and 19 columns remained. The final attributes chosen for the machine learning models are described in Table 1.

2.2 Dataset Pre-processing

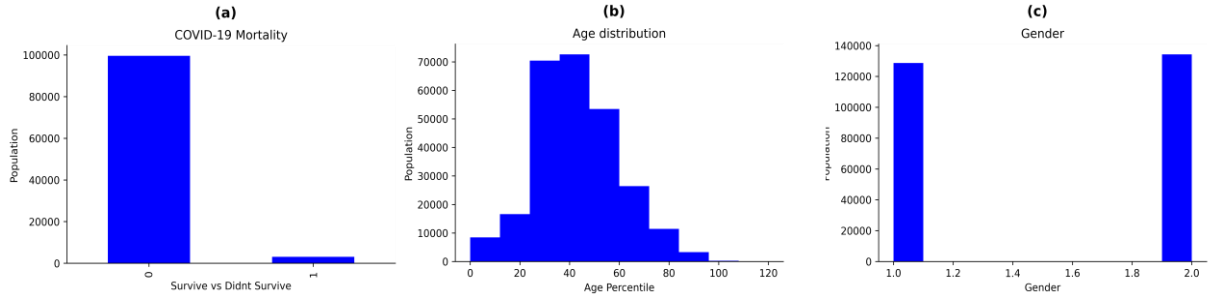

Fig. 1 Distributions of the patients: (a) Age (b) Gender (c) Mortality.

Table 1. Final list of attributes for our Machine Learning models.

Sl. No	Feature	Type of Parameter	Encoding	Data Type
1	Sex	Demographic	0-Female, 1-Male	Int64
2	Type of care	Demographic	0-Goodcare, 1-Excllent care	Int64
3	Death	Label	0-Death, 1-Survived	Int64
4	Intubation	Demographic	0-Negative, 1-Positive	Int64
5	Pneumonia	Clinical	0-Negative, 1-Positive	Int64
6	Age	Demographic	0-Negative, 1-Positive	Int64
7	Pregnant	Demographic	0-Negative, 1-Positive	Int64
8	Diabetes	Clinical	0-Negative, 1-Positive	Int64
9	EPOC	Clinical	0-Negative, 1-Positive	Int64
10	Asthma	Clinical	0-Negative, 1-Positive	Int64
11	Autoimmune disease	Clinical	0-Negative, 1-Positive	Int64
12	Hypertension	Clinical	0-Negative, 1-Positive	Int64
13	Cardiovascular disease	Clinical	0-Negative, 1-Positive	Int64
14	Obesity	Clinical	0-Negative, 1-Positive	Int64
15	Renal disease	Clinical	0-Negative, 1-Positive	Int64
16	Tobacco	Clinical	0-Negative, 1-Positive	Int64
17	Contact	Demographic	0-Negative, 1-Positive	Int64
18	ICU	Demographic	0-Negative, 1-Positive	Int64
19	Other disease	Clinical	0-Negative, 1-Positive	Int64

2.3 Co-relation Analysis

In this study, Pearson's co-relation was used to find the co-relation between the various features and patient mortality. This method is used to analyse whether a strong relationship exists between the dependent and independent variable.[24] Hence, the co-relation co-efficient r is used to measure the strength of the relationship among various variables. This analytical technique is based on the premise that determining the significance of a pertinent attribute in the data can be conducted by analysing the strength of the association between dependent and target variables. In Fig. 2, we see the various co-relations among the demographic and clinical parameters. Type of care, age, cardiovascular disease shows a slight positive co-relation with patient mortality. Intubation, pneumonia and ICU status shows a slight negative co-relation.

The most prominent features were further confirmed using a method called "mutual information", also called "information gain". The reduction in entropy caused by altering a dataset is calculated as information gain. By analyzing the gain of each parameter with the context of target variable, it can also be used for feature selection. This calculation is alluded as mutual information between the two clinical parameters. We used the mutual-info-classif class from the sklearn-feature-selection to compute the information gained from each variable. Fig. 3 concludes that age, intubation, ICU, type of care, pneumonia, diabetes and hyper tension are the most important features that distinguish patient casualty. This is in line with the literature, since most people who died due to COVID-19 were either old or were suffering from other comorbidities.

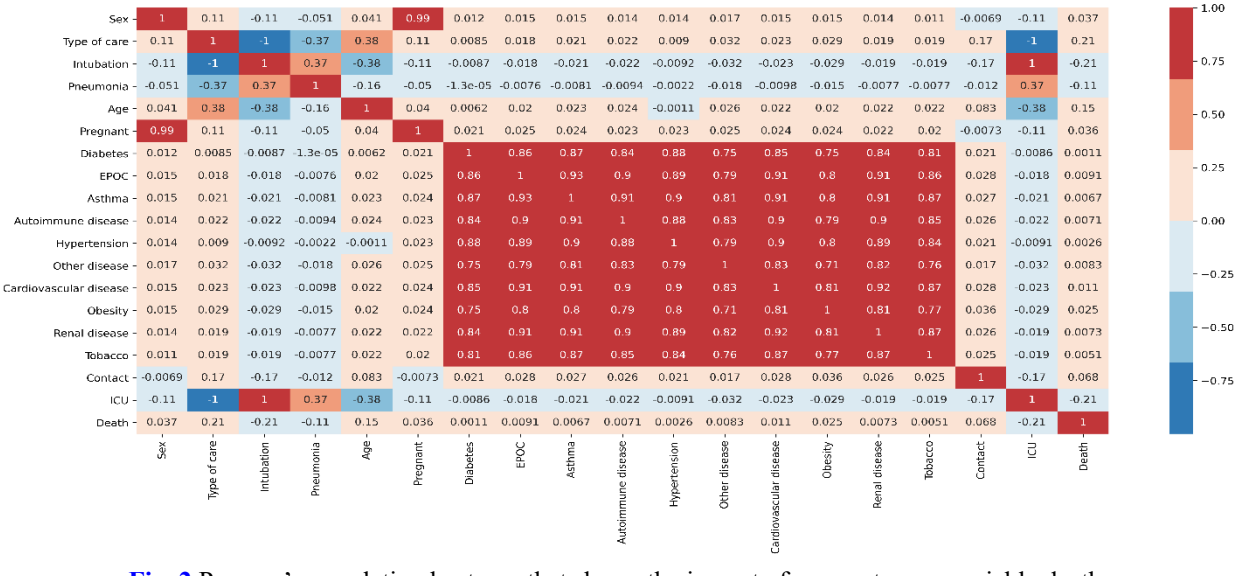

Fig. 2 Pearson's co-relation heatmap that shows the impact of parameters on variable death.

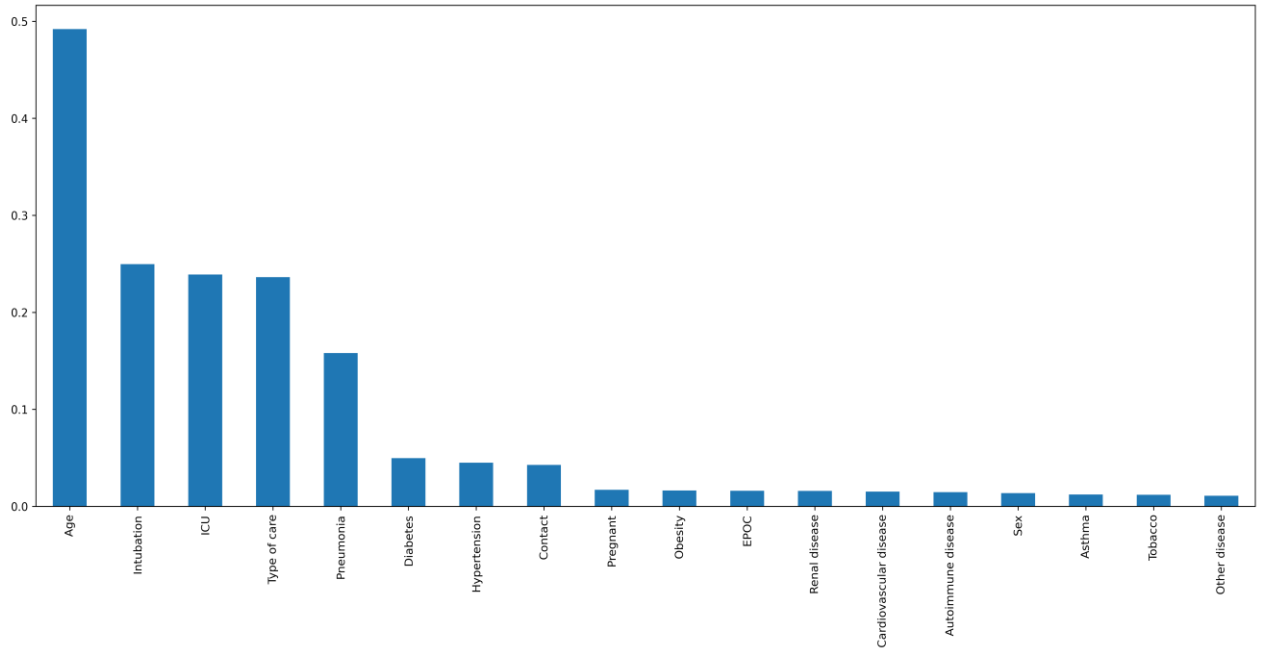

Fig. 3 Bar plot showing the most important features using mutual information.

3. Methodology

The different ensemble algorithms to predict COVID-19 mortality are discussed in this section. We also used Randomized search technique to obtain the best hyperparameters along with the fivefold cross validation technique and Fig. 4 depicts the entire solution pipeline.

3.1 Ensemble Machine Learning

Ensemble methods in statistics and machine learning combine many optimization models to achieve greater predictive performance than any of the individual learning classifier alone. Ensemble algorithms, unlike statistical ensemble in statistical theory, consists of only a specific limited collection of different models, but typically allows for considerable flexible design to exist within those possibilities. When there is a lot of variability among the models, ensembles generally produce better outcomes. As a result, several ensemble approaches aim to create diversification amongst some the models they combine. More random algorithms (like random decision trees) can yield a better ensemble than extremely purposeful algorithms, which may seem counterintuitive (like entropy-reducing decision trees). However, it has been proven that utilizing a range of high learning models are more beneficial than using strategies that aim to deconstruct models in order to encourage heterogeneity. There are further classified into bagging and boosting algorithms.

3.1.1 Random Forest

Random forest belongs to the supervised learning algorithm category. The “forest” refers to the ensemble of primary decision trees, usually combined using a method called bagging. This technique combines multiple models and increases the overall accuracy. This algorithm can be used for both classification and regression.

3.1.2 Adaboost

The Adaboost algorithm, shorthand for Adaptive Boosting, is a boosting approach used in Machine Learning as one of the ensembles methods. The weights are re-allocated to each instance, with higher weights applied to improperly identified instances. This is termed Adaptive Boosting. In supervised learning, boost is used to reduce bias and variation. It is based on the notion of successive learning.

3.1.3 Gradient Boosting

A popular boosting algorithm is gradient boosting. Each predictor in gradient boosting corrects the error of its predecessor. Unlike Adaboost, the training instance weights are not adjusted; instead, each predictor is trained using the predecessor's residual errors as labels. Shrinkage is a crucial characteristic to consider while using this approach. Each tree in the ensemble is multiplied by the learning rate (eta) and ranges from 0 to 1. In each iteration, prediction of the tree in the ensemble is decreased.

3.1.4 Extreme Gradient Boosting

It is an advanced version of the gradient boosted decision trees. In this novel algorithm, numerous numbers of decision trees are assigned sequentially. One of the important factors in xgboost are “weights”. All independent variables are allocated weights and then fed into the classifier (decision tree) and results are calculated. The weight of parameters that the tree predicted incorrectly is increased and the parameters are then loaded into another tree. These various predictors are then combined to create a more powerful and precise model. It can be used to solve problems including classification, regression,

user defined prediction and ranking.

3.1.5 Light gbm

Light gbm is also a boosting framework that uses multiple decision trees to make prediction. However, it uses two new techniques: Exclusive feature bundling (EFB) and gradient - based one side sampling. These two features make the classifier work better than the other boosting frameworks.

3.1.6 Catboost

Yandex developed catboost, or categorical boosting, an open-source boosting library. They can be used in ranking, recommendation systems, forecasting, and even personal assistants, in addition to regression and classification. Categorical characteristics are common in datasets and there are several ways for handling categorical features in boosted trees. Catboost automatically handles categorical features, unlike other gradient boosting methods (which require numeric data). One-hot encoding is one of the most frequent strategies for dealing with categorical data, however it becomes infeasible with more characteristics. To address this, features are classified into categories based on the target statistics (estimate target value for each category). Greedy, hold out, leave one out, and ordered are all approaches to calculate target statistics.

3.2 Shapley Additive features (SHAP)

SHAP stands for SHapley Additive exPlanations, and was developed from Shapley values. It was first introduced by Lloyd shapley in 1951 as a solution concept for cooperative game theory. SHAP is compatible with any deep learning or machine learning model. The 'Tree Explainer' method is utilised in a variety of tree-based models, including random forests, lightgbm and gradient boosting algorithms. SHAP uses different visuals to show the value of features and how they contribute to predictions. SHAP values are used to compare the impact of having a certain value for a certain feature to the forecast we'd make if that feature had a baseline estimate.

3.3 Local Interpretable model-agnostic explanations (LIME)

LIME is model-independent, which means it may be used with any machine learning model. The technique tries to figure out what the model is doing by changing the input of data samples and seeing how the predictions change. Model-specific approaches examine the core components of the black model machine learning model and how they interact in order to gain a better understanding of it. Local model interpretability is provided by LIME. It also tweaks the feature values in a single data sample and then evaluates the influence on the output.

3.4 K-fold cross validation technique

Any of several related model validation strategies for determining how the findings of a statistical methods will generalise to an independent data set is known as cross-validation. Cross-validation is a resampling process that tests and trains a model using various chunks of the data on successive rounds. It's most commonly employed in situations when the goal is prediction and the user wants to know how well a predictive model will perform in practise.

4. Results

4.1 Evaluation Metrics

A range of key evaluation metrics are used for our ensemble models. The following measures were used to test model reliability: accuracy, precision, recall, F1-score, AUC, PR-score and confusion matrix.

- **Accuracy:** The percentage of correctly predicted mortality status of patients in the entire dataset. It is calculated using the following equation (1).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{1}$$

where TP (True positive) and TN (True negative) are the correct predictions. FP (False positives) and FN (False negatives) are incorrect predictions. Our goal is to reduce a substantial number of FP's and FN's.

- **Precision:** The percentage of patients who did not die and were correctly identified by the ML models. It emphasizes on false positives. A model with good precision has very few false positive cases. It is calculated using the following equation (2).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{2}$$

- **Recall (Sensitivity):** The percentage of patients who actually survived and were correctly identified by the ensemble classifiers. It emphasizes on the false negatives. A model with good recall has very false negatives cases. It is calculated using the following equation (3).

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

- **F1-score:** It is a metric which considers both precision and recall. It emphasizes on both false positive and false negatives. A higher F1-score means that there are very false positive and negative cases and the models classify correctly. It is calculated using the following equation (4).

$$\text{F1} - \text{score} = 2 * \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \tag{4}$$

- **AUC:** The relationship between true positive rate (TPR) and false positive rate (FPR) is described in the ROC curve. AUC is the region under the ROC curve and it shows how well the ensemble classifier distinguishes between the binary classes (Dead and Survived).

Average Precision: At various thresholds, the precision-recall curve shows the trade-off between precision and recall.

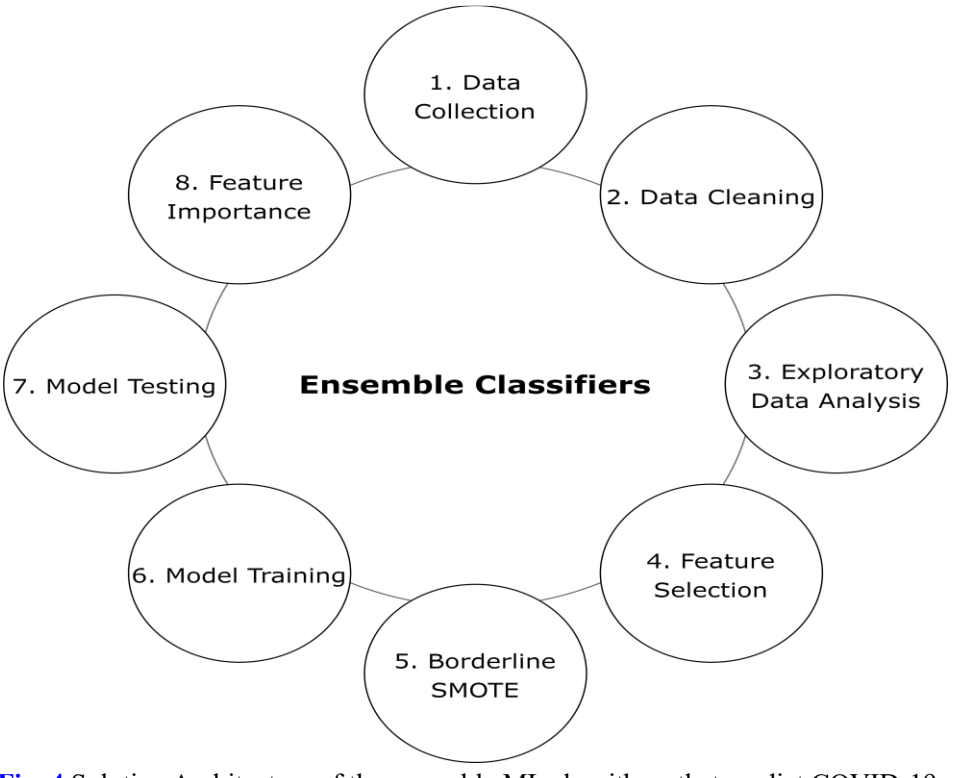

Fig. 4 Solution Architecture of the ensemble ML algorithms that predict COVID-19 mortality.

Average precision is the weight mean of precision and recall achieved at each threshold with the increase in recall from the preceding threshold used as weights. The formula for AP is given in equation (5) below.

$$\text{Average Precision} = \sum_p (R_k - R_{k-1}) P_k \quad (5)$$

where P_k and R_k are the precision and recall at the nth threshold.

- **Confusion Matrix**: Confusion matrix is a square matrix that gives a pictorial representation of the instances correctly classified by the model. The diagonal values indicate the correct prediction (Both true positives and true negatives) and the other columns indicate the misclassified instances (Both false positives and false negatives).

4.2 Model Evaluation

By aiding in the advanced prediction of mortality among COVID-19 patients, special care can be given to such patients who are at a high risk and prevention of unwanted casualties can be efficiently monitored. In this paper, ensemble algorithms such as random forest, adaboost, catboost, gradient an boost, xgboost and light gbm were used as the state-of-art classifiers since they are known to give better results. The performance of these models is summarized in Table 2. To obtain the best parameters, randomized search hyperparameter tuning technique was used. The SMOTE technique was also used prior to model evaluation. This method super sampled the minority class (COVID-19 casualties) and a balance between the two classes was obtained. Afterwards, the models were evaluated using the 5-fold cross validation method. The train-test-split method from the python library was used to divide the data into training and testing (80% for training and 20% for testing). The confusion matrix is used to observe the correct positive and negative outcomes predicted by the model. In this matrix, the number of false positives and false negatives (inaccurate predictions) can be observed too. The normalized confusion matrices of all models are depicted in Table 3 and an example is pictorially described in Fig. 5. The diagonal elements are the cases that have been identified correctly. The other elements display the error in prediction. It is noticeable that most of the cases have been identified accurately. We can calculate the accuracy, precision, recall and other metrics using the confusion matrix

Table. 2. Results obtained from the various ensemble models.

Sl. No	Model	Accuracy	Precision	Recall	F1-score	AUC	AP
1	Random Forest	93%	93%	94%	93%	94%	0.98
2	XGBoost	96%	95%	95%	95%	96%	0.99
3	LightGBM	91%	93%	88%	91%	91%	0.98
4	CatBoost	93%	93%	92%	93%	93%	0.99
5	AdaBoost	83%	86%	78%	82%	83%	0.88
6	Gradient Boost	86%	91%	75%	84%	85%	0.91

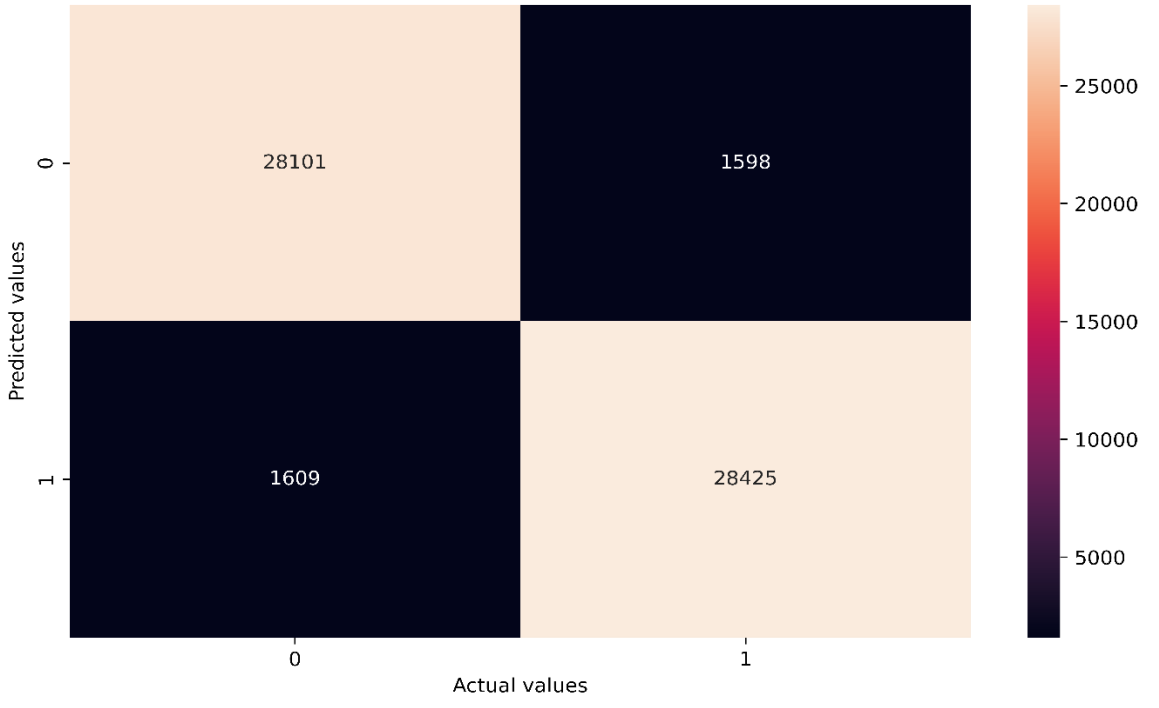

Fig. 5 Confusion Matrix obtained by the classifier XGBoost.

Table 3. Normalized confusion matrices of ensemble algorithms.

(a) Random forest		Actual	
		Negative	Positive
Predicted	Negative	0.94	0.06
	Positive	0.07	0.93
(b) AdaBoost		Actual	
		Negative	Positive
Predicted	Negative	0.77	0.23
	Positive	0.12	0.87
(c) CatBoost		Actual	
		Negative	Positive
Predicted	Negative	0.92	0.08
	Positive	0.06	0.94
(d) Gradient Boost			
		Actual	
		Negative	Positive
Predicted	Negative	0.93	0.07
	Positive	0.08	0.92
(d) XGBoost			
		Actual	
		Negative	Positive
Predicted	Negative	0.95	0.05
	Positive	0.05	0.95
(d) Light GBM			
		Actual	
		Negative	Positive
Predicted	Negative	0.88	0.12
	Positive	0.94	0.06

itself. The AUC and precision-recall (PR) curves are described in Fig. 6 and Fig. 7 respectively.

The xgboost model the best results among all models. The percentage of correct classifications made by the ensemble algorithms is called accuracy. The accuracy obtained by the xgboost model was 96%. Random forest, adaboost, catboost, gradient boost and light gbm obtained an accuracy of 93%, 83%, 93%, 86% and 91% respectively. Precision is also a metric that indicates model performance. It is the ability of the classifier to reveal accurate data points. In terms of mathematics, it is the number of true positives divided by the number of true positives and the number of false positive together. XG boost achieved the highest precision with 95%. Random forest, adaboost, catboost, gradient boost and xgboost obtained a precision of 93%, 86%, 93%, 91% and 93% respectively. The percentage of successful identification of patients who survived COVID-19 can be called as Recall/Sensitivity.

The recall obtained by random forest, xgboost, light gbm, catboost, adaboost, gradient boost models were 94%, 95%, 88%, 92%, 78% and 75% respectively. F1-score is a metric that combines both precision and recall. It is the harmonic mean of the above parameters mentioned. The F1-scores of the random forest, XG boost, light BGM, catboost, adaboost and gradient boost were 94%, 95%, 91%, 93%, 83% and 85% respectively. The relationship between the true positive rate and the false positive rate is depicted by the receiver operating characteristic (ROC) curve. The area under the curve (AUC) is the region under the ROC curve that shows how well the classifier differentiates between the binary groups. The higher the AUC, the better the performance of the models. The AUC's obtained by the random forest, xgboost, light gbm, catboost, adaboost and gradient boost were 94%, 95%, 91%, 93%, 83% and 85% respectively.

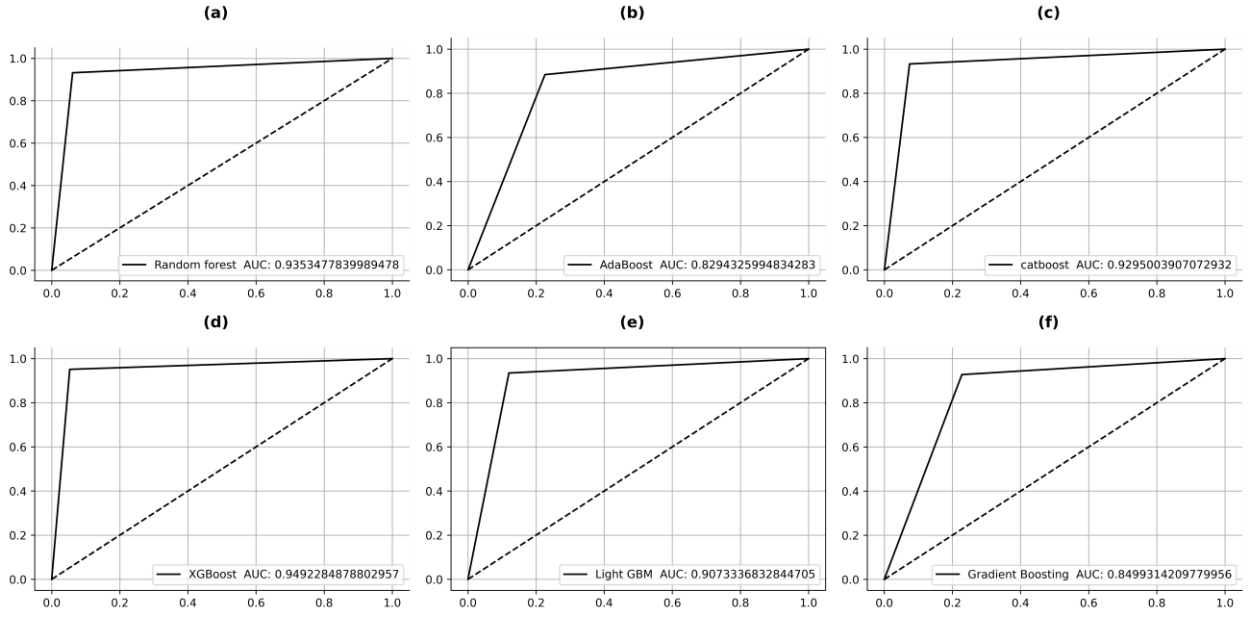

Fig. 6 AUC curves of the ensemble algorithms: (a) Random Forest (b) AdaBoost (c) CatBoost (d) XGBoost (e) Light GBM (f) Gradient Boosting.

This unpresented pandemic has spurred various research groups to develop machine learning applications with the purpose of predicting COVID-19 diagnosis and prognosis. However, only a few applications based on laboratory markers and epidemiological information have been devised. Qomariyal *et al*.,[15] used three classifiers: decision tree, random forest and xgboost for causality prediction due to coronavirus. However, in this study, only biomarkers were used. The maximum accuracy obtained was 98%. An accuracy of 94% was obtained by the decision tree model.[25] Eleven epidemiological parameters were selected for the final models. However, these models were used only for preliminary diagnosis. Five serum chemistry laboratory markers were used to predict patient casualty.[22] A 91% sensitivity and specificity were obtained by support vector machine model. Survival prediction among ICU admitted COVID-19 patients was experimented.[20] Many machine learning models were used and the maximum accuracy obtained was 90%. In comparison with other researches, this study gave preference to more attributes since it gives a better outlook. The ensemble algorithms are better than the conventional ML algorithms and we made efficient use of them in this research. The xgboost achieved optimal results with an accuracy, precision, recall, F1-score and AUC of 95%, 95%, 95%, 95% and 96% for the Mexican dataset. The summary of various researches is described in Table 4.

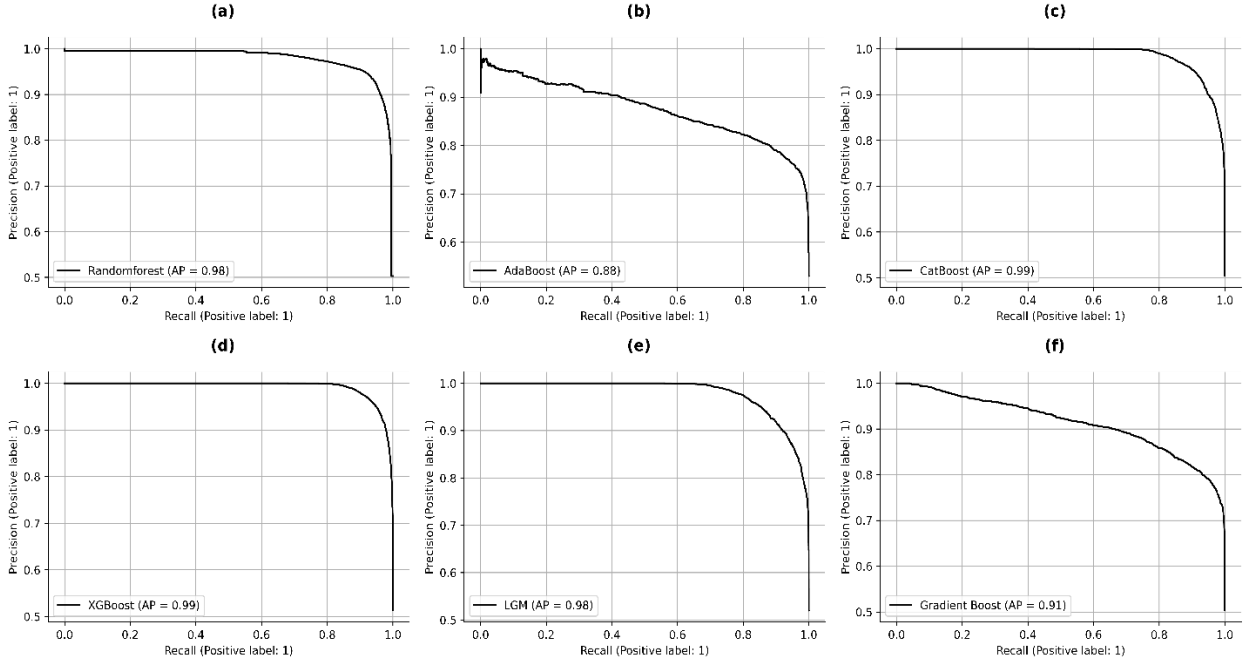

Fig. 7 PR curves of the ensemble algorithms: (a) Random Forest (b) AdaBoost (c) CatBoost (d) XGBoost (e) Light GBM (f) Gradient Boosting.

Table. 4. Comparison of related studies in predicting COVID-19 mortality.

Reference	Dataset Source	Total features	Models used	Maximum Accuracy
Qomariyah *et al.*, [15]	1000 COVID-19 patients, Jakarta	11 biomarkers	Decision tree, random forest and xgboost	98%
Muhammad *et al.*; [25]	263,007 patients from Mexico	11 epidemiological parameters	Five models	95%
Han *et al.*, [20]	375 patients from Wuhan	19 paramters	Broad Learning System	95%
Booth *et al.*, [22]	398 patients from Texas	12 laboratory markers	Many models	91%-sensitivity 91%-specificity
Present study	263,007 patients from Mexico	19 epidemiological paramters	Five Ensemble Algorithms	96%

4.3 Feature Importance

Clinical judgments made with the help of machine learning models in health care settings may have an impact on patient's lives, in addition to numerous legal and ethical implications. As a result, models that are interpretable and precise are in great demand in these applications. Model interpretability in the medical field refers to the ability of medical professionals to comprehend how the model uses input information to produce predicts, to check predictive modelling before relying on them, and to defend treatment decisions based on the results of the ensemble algorithms. As a result, casualty-based feature relevance estimates are critical for predictive model interpretability and reliability. We used the SHAP and LIME technique to analyze each feature's value in affecting he anticipated output in order to comprehend the suggested ensemble models.

SHAP analyzes each feature and its importance to the model output based on Shapley data. Fig. 8 depicts the importance of each feature in predicting COVID-19 mortality in descending order. In this research, age is the most important parameter followed by pneumonia, contact, intubation and ICU. Patients who are pregnant are also very vulnerable. People with underlying medical conditions such as diabetes, obesity, hyper tension, cardiovascular disease, asthma, and renal disease are also at high risk. The beeswarm plot, Fig. 8b indicates how the value of each feature impacts model output. Features are arranged based on the order of importance. The color on the right side represents the feature value, red denotes a greater value, while blue represents a lower value. From the diagram it can be observed that there is a greater mortality risk for the elderly.

Fig. 9 describes feature importance using LIME. Fig. 9a represents a patient who survived and Fig. 9b represents a patient who died due to COVID-19. LIME works by producing new training samples in the vicinity of the instance to be explained and using the older model to forecast other samples. The sample is then weighed depending on its proximity to the given instance and a linear regression is formed utilizing the newer samples and the considered instance. The learnt linear model is validated on a small scale using this strategy. Blue indicates patient who survived and orange indicates patient that died. In Fig. 9a we see that most of the features are blue and the model interprets this result based on majority voting. In Fig. 9b most of the features are orange and the model concludes by predicting the patient status as dead.

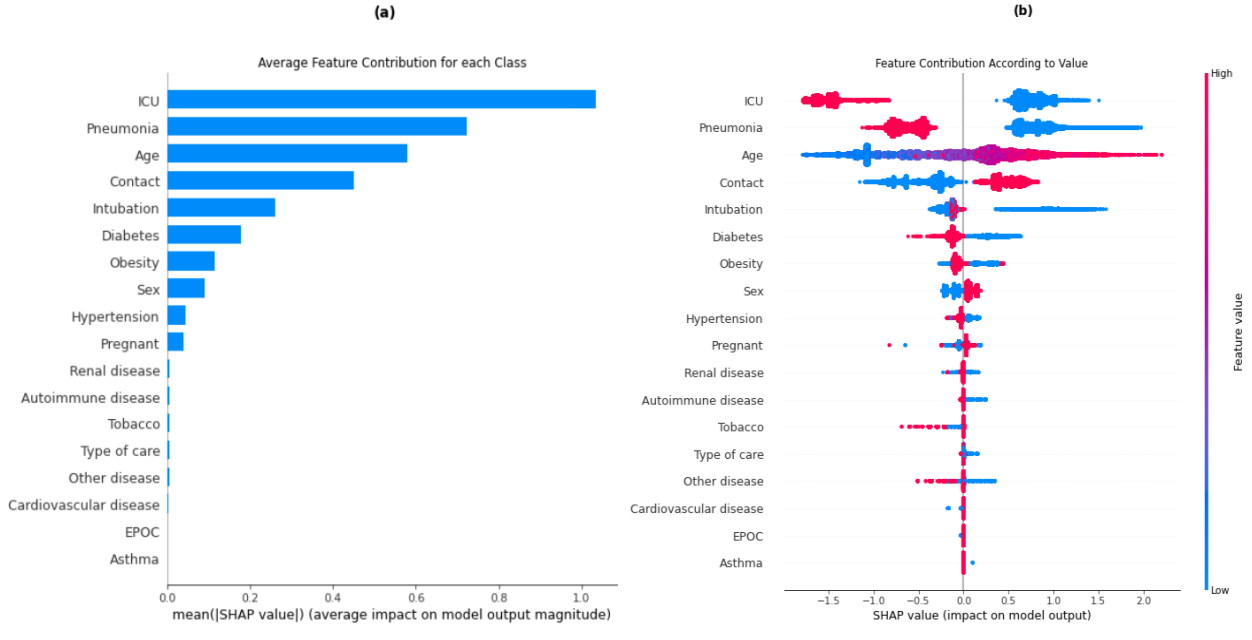

Fig. 8 SHAP plots showing feature importance in descending order (a) density plot (b) Beeswarm plot.

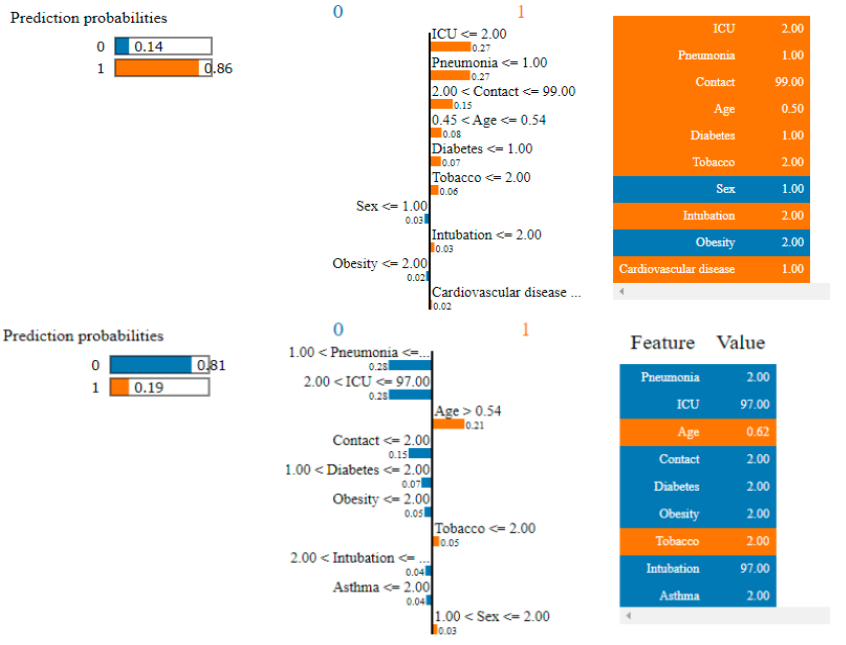

Fig. 9 Feature Importance using LIME (a) Patient death probabilities (b) patient survival probabilities.

5. Conclusion

In this research, epidemiological parameters were used to predict mortality risk in COVID-19 patients. These models are very useful since we can predict the mortality in advance. The most vulnerable patients can be given appropriate treatments and lives could be saved. This research can also save scarce hospital resources such as ventilators and ICU beds. These models can be used in almost all health care facilities since they require very minimal hardware and resources. The retrospective study was performed on 263,007 patients from Mexico. Nineteen demographic and clinical parameters were used as attributes for the six ensemble machine learning models. A comprehensive literature review was performed in the beginning and the various methods were compared with our models. After describing the dataset, pre-processing was performed that included exploratory data analysis and data cleaning. Various plots were visualized to understand more about input data. Pearson co-relation was used to understand various hidden co-relations that effected the target output. The SMOTE technique was used to balance the dataset and all models were tested using the 5-fold cross validation technique. Our models were evaluated using six performance metrics. Among all the algorithms, xgboost achieved the optimal results. Xgboost is highly flexible and it uses the power of parallel processing. In comparison to the gradient boosting algorithms, it is even faster. Regularization is also supported by this efficient algorithm. It also works well with imbalanced datasets. Hence it performed better than all the ensemble algorithms. In any medical research, feature importance is extremely important. We used SHAP and LIME methods to understand the most important epidemiological parameters that resulted in COVID-19 mortality. Older age and the presence of pneumonia, diabetes, obesity, cardiovascular and renal diseases indicated that the patient might succumb to COVID-19.

The applications of ensemble machine learning in this domain has a tremendous potential for aiding doctors and health care professionals in decision making, and predicting complications and in-patient mortality. Patients who are more vulnerable can be given appropriate treatments and lot of precious lives can be saved using medical intervention and Artificial Intelligence. Medical validation can be performed by clinical experts and these models can be deployed in various hospitals and health care facilities.

6. Challenges and Future Directions

This section highlights the different issues and the apparent suggestions for further researches in predicting COVID-19 mortality.

6.1 Challenges

• **Lack of important blood markers**: There are distinct blood markers such as Lactate Dehydrogenase (LDH), lymphocytes, CRP (C-reactive protein) and D-Dimer that indicates COVID-19 severity. However, this dataset didn't contain any blood reports. For better accuracy and validation, it is important to consider all the laboratory markers.

• **Single Center Data**: For accurate validation, it is important to consider from different sources and geographical resources. However, the data available in this research is only from Mexico. The results might be biased if we do not consider patients from all around the globe.

• **Data Balance**: The data in this research was very imbalanced. For ML algorithms to be reliable, it is very important to have balanced data. We use the borderline SMOTE technique in this research to reduce the impact of data imbalance. However, for further research, we can consider a more balanced data. In medical ML, most of the datasets are imbalanced since the

number of patients who are diagnosed with the disease are very few when compared to the normal population.

• **Statistical Analysis**: Since the data was already normalized, the scope to perform accurate statistical analysis was very slim. It is very important to know the exact values for advanced statistical investigation. Descriptive statistics (Mean, median, mode, *etc.*) and inferential statistics (t-tests, z-tests, chi-square tests) are very important to understand the input features.

6.2 Future Directions

• **Improving the dataset**: Collection of data that includes demographic, epidemiological, hematological, biochemistry and biomarkers must be done in subsequent researches. It is very important to consider these parameters for better classification.

• **Deep learning**: When the data is large, deep learning models work better than the traditional machine learning models. They also require very little data cleaning and preprocessing. Deep learning models are also very fast when the data is huge. The accuracy obtained by these models are also superior.

• **Medical validation**: Validation must be done by medical experts so that the models can be deployed in real time. Once the models deliver trust worthy results, they can be deployed in real time to assist medical personnel.

• **Combining multiple machine learning models**: COVID-19 severity can also be diagnosed using CT scans and X-rays. Ensemble of the above methods combined with our models can be performed to obtain optimal accuracy. False negatives obtained by one of these models can be easily prevented.

Conflict of interest

There are no conflicts to declare.

Supporting information

Not applicable.

References

[1] M E El Zowalaty, J D Järhult, *One Health Amsterdam.*, 2020, **9**, 100124, doi: 10.1016/j.onehlt.2020.100124.

[2] WHO Coronavirus disease situation reports, 16 April 2020.https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.

[3] F Yang, S Shi, J Zhu, J Shi, K Dai, X Chen, *J Med Virol*., 2020, **92**, 2511-2515, doi: 10.1002/jmv.25891.

[4] C D Covid, R Team, C COVID, R Team, C COVID C, R Team, S Bialek, E Boundy, V Bowen, N Chow, A Cohn, *Morbid. Mortal. Weekly Rep.*, 2020, **69**, 343-346, doi: 10.15585/mmwr.mm6912e2.

[5] C S Tyrrell, O T Mytton, S V Gentry, M Thomas-Meyer, J L Allen, A A Narula, B. McGrath, M Lupton, J Broadbent, A Ahmed, Mavrodaris, *Thorax.* 2021, **76**, 302-312, doi: 10.1136/thoraxjnl-2020-215518.

[6] C Yang, C. Delcher, E. Shenkman, S. Ranka, *CRC Press*, 2019, Boca Raton, FL, USA.

[7] E. Topol, Basic Books: New York, 2019, NY, USA.

[8] A.C Chang, *Academic Press*,2020, New York, NY, USA.

[9] P. Timsina, A. Kia, *J. Clin. Med.*, 2020, **9**, 1668-1674, doi: 10.3390/jcm9061668

[10] J. L Izquierdo, J. Ancochea, J. B. Soriano, *J. Med. Internet Res*., 2020, doi: 10.2196/21801.

[11] K Chadaga, S Prabhu, B K Vivekananda, S Niranjana, S. Umakanth, *Cogent Eng.*, 2021, **8**, 1958666, doi: 10.1080/23311916.2021.1958666.

[12] S Kushwaha, S Bahl, A K Bagha, P Kulwinder Singh, M Javaid, A Haleem, R P Singh, *J. Ind. Inf. Integr.*, 2020, 5,453-479, doi: 10.1142/S2424862220500268

[13] A Alimadadi, S Aryal, I Manandhar, P B Munroe, B Joe, X Cheng, *Physiol Genomics*, 2020, **52**, 200–202, doi : 10.1152/physiolgenomics.00029.2020

[14] F Shoeb, I Hussain, G Afrin, S T Mufti, S T Raza, F Mahdi F, *MedRxiv*, 2021.

[15] N N Qomariyah, A A Purwita, S D Asri, D Kazakov, *International Conference on ICT for Smart Society (ICISS)* 2021, IEEE, doi: 10.1109/ICISS53185.2021.9533219.

[16] C Hu, L Liu, Y Jiang, O Shi, X Zhang, K Xu, C Suo, Q Wang, Y Song, K Yu, X Mao, *Int. J. Epidemiol.*, 2020, **6**, 1918-1929, doi: 10.1093/ije/dyaa171.

[17] A Karthikeyan, A Garg, P K Vinod, U D Priyakumar, *Front. public health*, 2021, **9**, 313-346, doi: 10.3389/fpubh.2021.626697.

[18] M Sánchez-Montañés, P Rodríguez-Belenguer, A J Serrano-López, E Soria-Olivas, *Int. J. Environ. Res. Public Health,* 2020, **22**, 83-86, doi: 10.3390/ijerph17228386.

[19] S Li, Y Lin, T Zhu, M Fan, S Xu, W Qiu, C Chen, L Li, Y Wang, J Yan, J Wong, *Neural Comput. Appl.*, 2021, **9**,1-10, doi: 10.1007/s00521-020-05592-1.

[20] R Han, Z Liu, C L Philip Chen, L Xu, G Peng, *7th International Conference on Information, Cybernetics, and Computational Social Systems* (ICCSS) 2020, 837-842, doi: 10.1109/ICCSS52145.2020.9336835.

[21] A K Das, S Mishra, S S Gopalan, *PeerJ.*, 2020, **8**, e10083, doi: 10.7717/peerj.10083

[22] A L Booth, E Abels, P McCaffrey, *Modern Pathol.*, 2021, **34**, 522-531, doi: 10.1038/s41379-020-00700-x.

[23] Mariana R Franklin, Mexico COVID-19 clinical data, https://www.kaggle.com/marianarfranklin/mexico-covid19-clinical-data/metadata.

[24] P. Sedgwick, *Bmj.*, 2012, **345**, 375-402 doi: 10.1136/bmj.e4483.

[25] L J Muhammad, E A Algehyne, S S Usman, A Ahmad, C Chakraborty, I A Mohammed, *SN computer science*, 2021, 2, 1-3, doi: 10.1007/s42979-020-00394-7.

Author information

Krishnaraj Chadaga *received the MTech degree in Computer Science and Engineering from Manipal Institute of Technology respectively. He is currently pursuing his Ph.D. degree in Manipal*

Institute of Technology. He has also worked in various software companies such as Dell EMC, Informatica and Diya Systems. His research interests include Machine Learning and Deep learning for medical diagnosis and prognosis, bioinformatics, etc.

Srikanth Prabhu *is an Associate Professor (Senior-Scale) in the Department of Computer Science and Engineering Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India. He received his MTech and Ph.D degree from IIT Karagpur, India. His research interests include medical diagnosis and molecular bioinformatics.*

Shashikiran Umakanth *is Professor & Head of Internal Medicine, and Medical Superintendent at Dr TMA Pai Hospital, Udupi, of Manipal Academy of Higher Education. A clinician and medical teacher, he has been awarded for excellence in clinical teaching many times by undergraduate medical students. He has more than 50 research publications and has presented papers in many international conferences. His areas of interest include infectious diseases, diabetes and related metabolic disorders, medical education and technology. He has also served as nodal head during this COVID-19 pandemic.*

Vivekananda Bhat K *received the Ph.D. degree in computer science and engineering from IIT Kharagpur, India and the M.Tech. degree from National Institute of Technology Karnataka, Surathkal, India. He is currently an Associate Professor (Senior-scale) with the Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India. His research interests includes cyber security and machine learning. He has published several papers in reputed international journals and conferences. He is a senior Member of IEEE.*

Niranjana Sampathila*, PhD, is currently working as Associate Professor-Senior Scale in the Department of Biomedical Engineering, Manipal Institute of Technology (MIT), Manipal Academy of Higher Education, Manipal. Has more than 24 years of teaching and resaerch experience. He is a senior member of IEEE (USA) and Fellow of IE (India). Research interest includes: Pattern reconition and AI in healthcare, biomedical engineering and miniaturised systems and nanotechnology.*

Rajagopala Chadaga *is an Associate Professor (Senior -Scale) in Department of Mechanical and Manufacturing Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education. His researches include the application of machine and deep learning in mechanical Engineering domain.*

Krishna Prakasha K *received the B.E. and M.Tech. degrees from Viswesvaraya Technological University, Belagavi, and a Ph.D. degree in Network Security from the Manipal Academy of Higher Education, Manipal, India. He is currently an Assistant Professor (Senior) with the Department of Information and Communication Technology. His research interests include information security, network security, algorithms, real-time systems, wireless sensor networks and machine learning*

Publisher's Note: Engineered Science Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.