Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Decision Analytics Journal 7 (2023) 100246

Contents lists available at ScienceDirect

Decision Analytics Journal

journal homepage: www.elsevier.com/locate/dajour

A machine learning and explainable artificial intelligence triage-prediction system for COVID-19

Varada Vivek Khanna [a], Krishnaraj Chadaga [b], Niranjana Sampathila [a,*], Srikanth Prabhu [b], Rajagopala Chadaga P. [c]

[a] Department of Biomedical Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, 576104, India
[b] Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, 576104, India
[c] Department of Mechanical and Manufacturing Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, 576104, India

ARTICLE INFO

Keywords:
Artificial intelligence
Machine learning
COVID-19
Data balancing
Shapley Additive Explanations
Random Forest

ABSTRACT

COVID-19 is a respiratory disease caused by the SARS-CoV-2 contagion, severely disrupted the healthcare infrastructure. Various countries have developed COVID-19 vaccines that have effectively prevented the severe symptoms caused by the virus to a certain extent. However, a small section of people continues to perish. Artificial intelligence advances have revolutionized healthcare diagnosis and prognosis infrastructure. In this study, we predict the severity of COVID-19 using heterogenous Machine Learning and Deep Learning algorithms by considering clinical markers, vital signs, and other critical factors. This study extensively reviews various classifier architectures to predict the COVID-19 severity. We built and evaluated multiple pipelines entailing combinations of five state-of-the-art data-balancing techniques (Synthetic Minority Oversampling Technique (SMOTE), Adaptive Synthetic, Borderline SMOTE, SMOTE with Tomek links, and SMOTE with Edited Nearest Neighbor (ENN)) and twelve heterogeneous classifiers such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, Naïve Bayes, Xgboost, Extratrees, Adaboost, Light GBM, Catboost, and 1-D Convolution Neural Network. The best-performing pipeline consists of Random Forest trained on Borderline SMOTE balanced data that produced the highest recall of 83%. We deployed Explainable Artificial Intelligence tools such as Shapley Additive Explanations and Local Interpretable Model-agnostic Explanations, ELI5, Qlattice, Anchor, and Feature Importance to demystify complex tree-based ensemble models. These tools provide valuable insights into the significance of critical features in the severity prediction of a COVID-19 patient. It was observed that changes in respiratory rate, blood pressure, lactate, and calcium values were the primary contributors to the increase in severity of a COVID-19 patient. This architecture aims to be an explainable decision-support triaging system for medical professionals in countries lacking advanced medical technology and infrastructure to reduce fatalities.

1. Introduction

The Coronavirus infectious disease was initially reported in December 2019 in Wuhan, China [1]. Within a few months, cases emerged from various nations from all continents, making it a global pandemic. Healthcare facilities were flooded with patients complaining of influenza-like symptoms such as myalgia, fever, shortness of breath, and body pain [2]. This virus emerged to be highly deadly, causing more than 6.3 million deaths worldwide [3]. Reverse transcription polymerase chain reaction (RT-PCR) and Rapid Antigen Testing (RAT) were methods deployed for mass testing of this deadly virus [4].

The symptoms of COVID-19 can be easily misdiagnosed with other viral diseases. Although RT-PCR is considered a gold standard for virus detection, there have been reports showing a 20% chance of false negative results [5]. The high probability of false negative results can be fatal to the patient and people in contact. Large-scale vaccination has significantly prevented the spread of COVID-19. However, a portion of the population continues to face deaths (older adults and patients with comorbidities) [6]. Hence, it is crucial to identify such high-risk patients. An early prognosis, appropriate treatment administration, and patient severity insights can avert fatalities. During the peak of the COVID-19 pandemic, the number of vacant Intensive Care Unit (ICU) beds and hospital equipment was limited and had to be rationed judiciously. These resources were scarce in many developing countries. Establishing a balance between optimal resource usage and patient care was essential.

With the ongoing advancements in clinical informatics, Machine Learning (ML) and Artificial Intelligence (AI) have become an active

* Corresponding author.
E-mail address: niranjana.s@manipal.edu (N. Sampathila).

https://doi.org/10.1016/j.dajour.2023.100246
Received 5 March 2023; Received in revised form 21 April 2023; Accepted 2 May 2023
Available online 6 May 2023

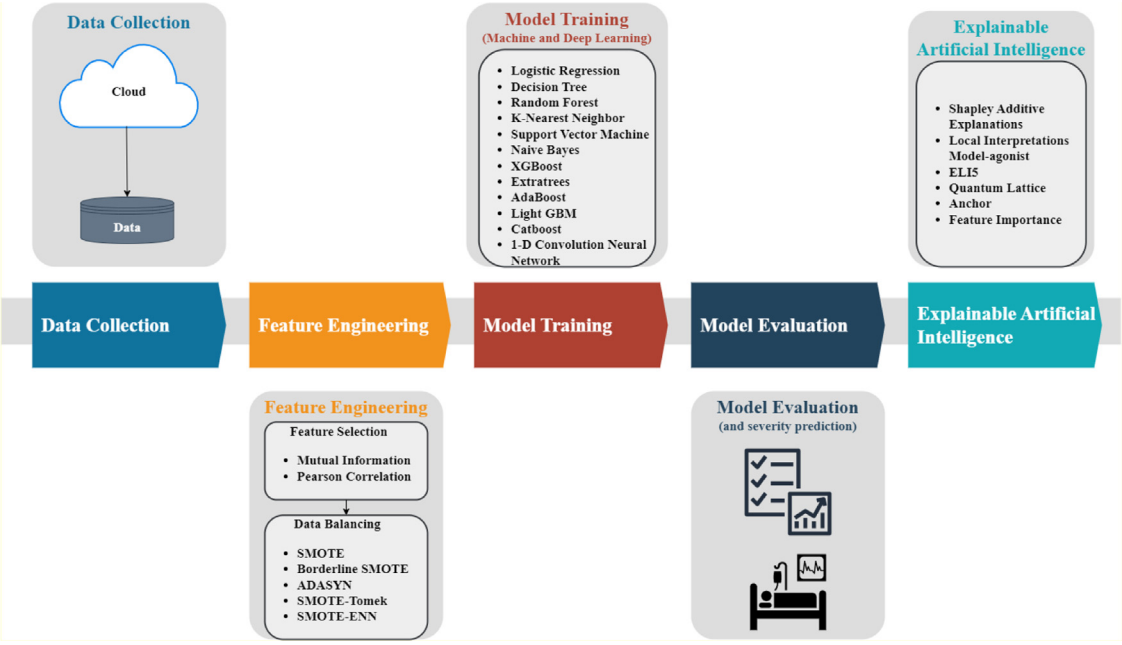

Fig. 1. Process flow of this research.

part of medical technology. Producing statistical estimations that yield in-depth insights has become faster and more accessible than ever. Precise interpretations assist clinicians in the early prediction of a patient's severity. Artificial Intelligence is actively used in drug development and delivery, AI genetic sequencing, robotic surgeries, and others in healthcare [7,8]. Various Laboratory tests monitor crucial biomarkers that assist in COVID-19 severity prediction. Coronavirus induces an immune response in a healthy human circulatory system spiking the levels of biomarkers such as leukocytes, monocytes, bilirubin, lactate dehydrogenase (LDH), C-reactive Protein (CRP), D-Dimer [9,10]. COVID-19 manifestation can alter the levels of various body vitals. Breathlessness, lower oxygen saturation (SpO2) values, elevated respiratory rate, and increased body temperature indicate a COVID-19-positive patient [11,12]. Furthermore, manually monitoring these parameters and assessing the severity is a time-sensitive task and can drastically reduce the efficient functioning of healthcare infrastructure. Machine learning algorithms can automate such triaging processes by recognizing recurring patterns for patients with mild and severe symptoms.

During the pandemic peak, the unpreparedness, inefficiencies, and improper allotment of medical resources lead to multiple fatalities. Hospital resources were exhausting with a global shortage of ventilators, space in hospital ICUs, and clinicians working overtime. During this period, an effective automated severity prediction triaging decision support system could have averted numerous deaths caused by the lack of treatments and hospital resources. In this research, we aim to create a precise and reliable pipeline to automate patient triaging based on severity prediction. We deployed supervised machine learning algorithms on an open-source COVID-19 dataset from Kaggle, containing 1925 medical records of patients admitted to Brazil's Sírio-Libanês hospital [13]. Prior to model training, feature selection techniques such as Mutual Information and Pearson Correlation have been utilized. We have created multiple ML pipelines with classifiers and five data-balancing techniques to evaluate the best-performing architecture for severity prediction. Further, the best pipeline is interpreted with a layer of explainable AI tools. The workflow of this research is described in Fig. 1.

The contributions of this article are:

- Evaluate machine learning research that predicts the severity of a COVID-19 patient using clinical markers.
- Evaluate five data-balancing techniques of Synthetic Minority Oversampling Technique (SMOTE), Adaptive Synthetic, SMOTE with Tomek links, Borderline SMOTE, and SMOTE with Edited Nearest Neighbor.
- Assess eleven classifiers of Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, Naïve Bayes, Extreme Boosting, Light Gradient Boosting, Adaptive Boosting, Extratrees, and Categorical Boosting.
- The ML pipelines are compared with the Deep Learning Classifier 1-Dimensional Convolution Neural Network.
- Validate attributes using Shapley Additive Explanations and Local Interpretable Model-agnostic Explanations, Quantum Lattice, 'Explain Like I am 5', Anchor, and Feature Importance.
- Discuss the significant clinical markers used in predicting COVID-19 prognosis from a medical perspective.

The article is organized as follows: Section 2 discusses the literature review. Various materials and methods are presented in Section 3. The results and discussion are extensively elaborated in Section 4. Finally, the article concludes in Section 5.

2. Literature review

In this research, we conducted a literature review of articles from various databases such as Scopus, medRxiv, PubMed, Crossref, Embase, MEDLINE, and Google Scholar. The search string included keywords such as 'COVID-19','Coronavirus', 'Sars-CoV-2', 'Pandemic', 'Artificial Intelligence', 'Machine learning', 'Severity', 'Prognosis', 'Clinical markers', 'hematology', 'blood tests', 'COVID-19 markers', 'Triage', 'Healthcare', 'AI for COVID-19' and others. The strings were further combined to identify relevant papers.

Aktar et al. [14] demonstrated that values of measurable blood parameters could help distinguish between healthy and severe patients. They chose two datasets of 89 (Zenodo hospital) and 1945 patients (Sirio Libanes, São Paulo hospital). Integration of statistical comparison and correlation techniques such as T-test, chi-square analysis, and Pearson correlation was followed by the usage of Machine learning algorithms such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT). Lactate and granulocyte levels were concluded to be most significant for COVID-19

prognosis prediction. The architecture proposed provided a precision of 90%. Deif et al. [15] developed an automated system for allotting patients to intensive care units. The study proposed extreme gradient boosting (Xgboost) for the decision-making followed by an analytic hierarchy process to prioritize patients according to severity. They utilized supervised ML on the online Kaggle dataset from Sirio Libanês Hospital in Brazil with 1945 data samples. However, only 550 samples were considered during the study. The extraction of 38 features was followed by deploying machine learning models of DT, RF, SVM, KNN, Artificial Neural Network, and the Xgboost classifier achieving a 97% accuracy in predicting mild patients. Statsenko et al. [16] proposed an approach wherein a justifiable threshold for the biomarker levels of a COVID-19 patient could assist predict severity. The data was obtained from Dubai Mediclinic Parkview Hospital, consisting of 520 patients (72 admitted to ICU). The value of markers such as CRP, LDH, Ferritin, and Fibrinogen levels of ICU patients was significantly higher than those of non-ICU patients. They performed multivariable severity classification by deploying classification models such as Adaboost, Gradient Boosting, RF, and Neural Networks. The dataset was imbalanced, which may have changed the predictive threshold values. Nevertheless, they obtained an Area Under the Receiver Operating Characteristic Curve (AUC) score of 0.86.

Xiong et al. [17] evaluated various ML classifiers, such as RF, SVM, and Logistic regression (LR), processed on a dataset of 287 patients. They selected the best 23 parameters from the original 86 by Least Absolute Shrinkage and Selection Operator with Cross Validation estimator and Spearman's rank correlation. Feature importance was further assessed by measuring the performance of each model. The highest AUC score of 0.97 was obtained from the RF model. They concluded that neutrophil to lymphocyte ratio, D-dimer, and LDH value significantly influence severity prediction.

Schöning et al. [18] proposed a method of predicting risk scores using an ML-based approach. Analysis of COVID-19 data from February to August 2020 (first wave) for training the models and COVID-19 data from September to November 2020 (Second Wave) for model validation was conducted. The dataset consisted of 657 patients. Features like gender, CRP, sodium, and hemoglobin levels were the most prominent parameters in predicting the risk score. The highest AUC score obtained was 0.98. Karlafti et al. [19] proposed an unsupervised ML approach for the prognosis prediction of COVID-19 patients. Data from 268 patients were obtained from AHEPA University Hospital of Thessaloniki. The important features were chosen using principal component analysis for Numerical features and multiple correspondence analysis for categorical features. Further, using Gaussian Mixture models reduced the overall clustering. Aljameel et al. [20] analyzed data from 287 patients admitted to King Fahad University Hospital, Saudi Arabia. They aimed to create a model to trigger early caution for all at-risk patients. LR, RF, and Xgboost were deployed in this study. Additionally, they used 10-fold cross-validation for training and testing. SMOTE balancing techniques alleviated data imbalance. They concluded that RF was the best classifier since it obtained an accuracy of 95%. The remaining articles are tabulated in Table 1.

3. Materials and methods

This study was conducted on a six-core i7, 8th Gen processor, 16 GB RAM, NVIDIA GeForce GTX 1050 Ti- GPU. Jupyter-notebook version: 6.1.4 was used to build ML architectures. This section encompasses the data description, followed by components of feature engineering, data-balancing techniques, and a detailed view of the machine learning models deployed. These methods, in tandem, assisted in creating multiple frameworks for a holistic severity prediction system.

3.1. Data description

We used an open-source dataset from the Kaggle repository to extract the data for this study. A Data Intelligence team from Sírio-Libanês Hospital (São Paulo and Brasilia) uploaded this dataset [13]. It contains records of 1925 patients along with 231 recorded parameters. The features in the dataset can be categorized as follows: Three demographic parameters, nine grouped diseases, thirty-six blood parameters, and six vital signs of COVID-19-positive patients. This study aims to predict the severity of COVID-19 by assessing the 'ICU' feature by a binary classification with values '1' and '0'. '1' indicates that the patients were admitted to ICU. Out of 1925 patients, 515 were admitted to the ICU. On data retrieval, we observed the original dataset was normalized between −1 and 1.

3.2. Data pre-processing

During data pre-processing, the missing values were replaced with the statistical median of the column. Among the 231 recorder parameters, we initially selected 47 features. The categorical variables were then converted into numerical ones by one-hot feature encoding. For feature extraction, we aimed to drop all features having a low correlation with the target variable. The two feature selection methods used were Pearson correlation and Mutual information [28]. Mutual information (MI) is a part of probability and information theory wherein the dependence of one feature is obtained by studying the remaining random feature. Manual extraction for the top 15 features was conducted based on Fig. 2, representing a decreasing dependency plot obtained with MI. The formula used to calculate mutual information is described in Eq. (1).

$$\begin{aligned} MI\,(Feature;Target) &= MI\,(Feature;ICU) \\ &= H\,(Feature) - H\,\{Feature|ICU\} \end{aligned} \tag{1}$$

Here, H indicates the entropy.

In Pearson correlation, the linear dependence of the features is calculated, producing a heatmap. This graphically represents the degree of correlation between features [29]. Pearson correlation heatmap is described in Fig. 3. The calculation of Pearson's correlation coefficient is described in Eq. (2).

$$\rho = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sqrt[2]{\sum(y-\overline{y})^2\sum(y-\overline{y})^2}} \tag{2}$$

where x = x_i = *value of x- variables of the feature*
$y = y_i$ = *value of y- variables of the feature*
$\overline{x}$ = *average of x-variables of the feature*
$\overline{y}$ = *average of the y-variables of the feature*

The value of the Pearson Correlation Coefficient is zero when both the features under consideration are independent. Features with Pearson Correlation Coefficient lower than 0.04 were removed from the dataset. Features namely, PC02_VENOUS_MEAN, PC02_ARTERIAL_MEAN, OXYGEN_SATURATION_MEAN, HEART_RATE_MEAN and PC02_ARTERIAL_MEAN were dropped. The final dataset contained ten features, a list of which is tabulated in Table 2.

3.3. Data balancing

Out of the 1925 patients, 515 were admitted to the ICU. This data is disproportionate as the number of non-ICU patients is significantly higher than that of ICU patients. Data imbalance can significantly impact the accuracy and performance metrics. Two widely used approaches to correct data imbalance are: (1) The majority class is under-sampled. However, there is a loss of valuable information. (2) The minority class can be oversampled, producing multiple synthetic points. However, there can be multiple identical points. For this study, oversampling of the majority class was conducted. The data was split by the 70:30 train-test ratio, wherein 70% of the data is given to the models for training, and 30% is used for validating and evaluating the models. Data balancing was performed on the training data. We created five pipelines deploying the following balancing techniques: ADASYN, SMOTE, SMOTE-Tomek, Borderline SMOTE, and SMOTE-ENN. Further,

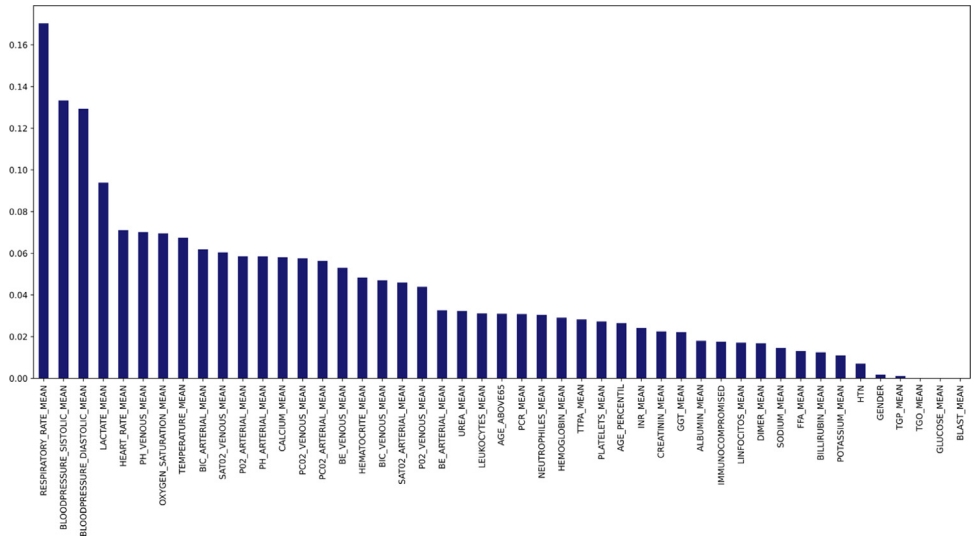

Fig. 2. Mutual information decreasing dependency graph.

Table 1
Related researches which predict COVID-19 severity using machine learning.

References	Sources	Data size	Total features	Model used	Accuracy	Sensitivity	Specificity	AUC	Novelty
[21]	34 medical centers nationwide	1,926,526 patients	–	RF, Xgboost, LR, SVM	71.10%	46.10%	–	0.86	Analysis of a large dataset with a comprehensive evidence-based review.
[22]	Taizhou Hospital	841	124	LightGMB (Gradient Boosting Machine), Catboost, Random Forest, Adaboost, Nearest Neighbors and Decision Tree	98%	64%	98%	1	Targeted eleven most significant features in routine testing.
[23]	Multiple sources	759	20	SVM, KNN, Xgboost, Extra-tree, Logistic Regression	–	95%	–	0.99	Targeted age, Lymphocyte count, D-dimer, CRP and Creatinine, achieving high sensitivity.
[24]	Tongji Hospital	151	33	Artificial Neural Networks	–	100%	86%	0.95	Demonstrated the correlation between blood urea nitrogen, albumin levels and severity of COVID-19.
[25]	Tongji Hospital, wuhan	183	–	Multiple models	–	83.90%	–	0.88	Provided a comprehensive analysis of blood parameter levels in diseased and surviving patients.
[26]	Tertiary medical centre with designated departments for patients with COVID-19	6995	–	Artificial Neural Networks, RF, DT, *Acute Physiology and Chronic Health Evaluation-II*	92.00%	88.00%	92.70%	0.9	Identification of risk factor trends six days before the patient's condition becoming critical.
[27]	Tongji hospital	375	300+	Xgboost	98.00%	100%	–	–	Converted patient classification into mild to severe to critical condition rather than focusing on endangered patients.

one distinctive set of ML pipelines was dedicated to models trained on unbalanced data. The sklearn and imblearn python libraries were used for balancing data. We created six sets of frameworks. Fig. 4 represents the count of the negative and positive data points in the train data before and after data balancing.

ADASYN (Adaptive Synthetic) adapts and varies the weights of different data points in the minority class to compensate for the imbalance [30]. This technique calculates the ratios of each majority (non-ICU class) by its K-nearest neighbor. All such ratio values are normalized. This process is looped, wherein a minority point is chosen from the neighborhood of each ICU data point, followed by the generation of new data points.

The following equations define the ADASYN balancing:

$$Degree\ of\ class\ imbalance = d = \frac{Number\ of\ minority\ class}{Number\ of\ majority\ class} = \frac{ml}{ms} \quad (3)$$

$$Count\ of\ synthetic\ classes\ to\ be\ produced\ for\ ICU\ class = G = (ml - ms) \times \beta \quad (4)$$

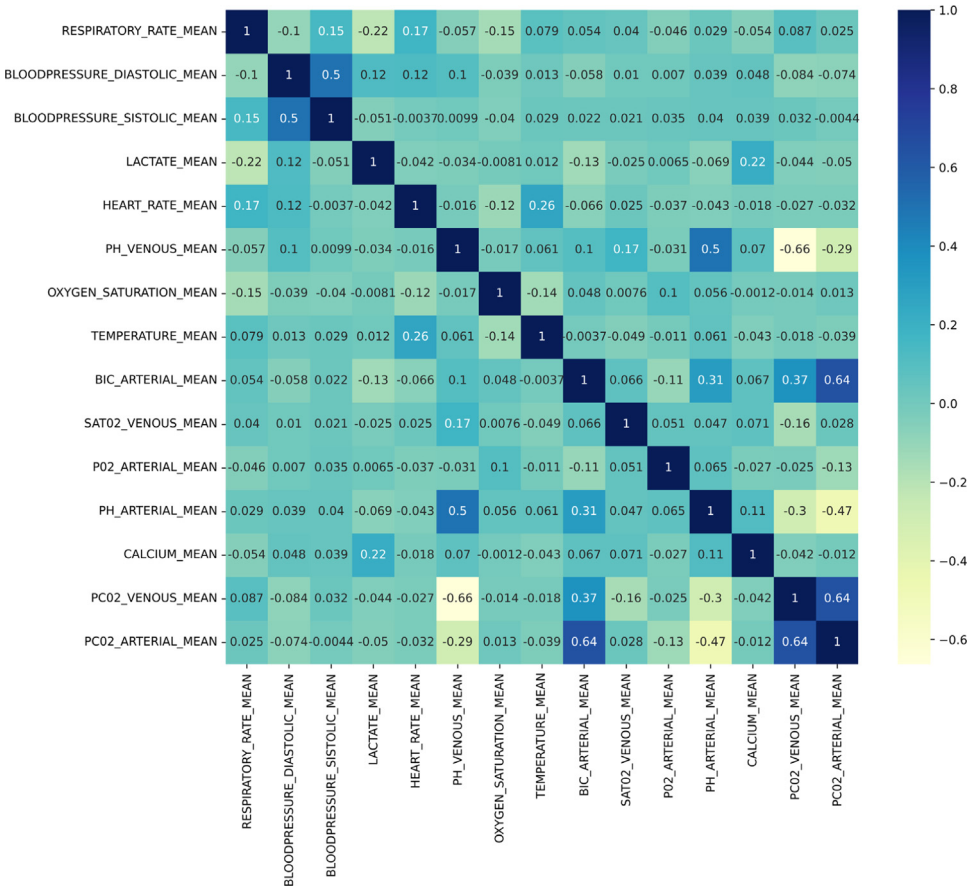

Fig. 3. Pearson correlation heatmap.

Table 2
Description of the top 14 features chosen.

Number	Feature name	Feature description	Criteria
(1)	RESPIRATORY_RATE_MEAN	Number of breaths taken in a minute	Vital sign
(2)	BLOODPRESSURE_DIASTOLIC_MEAN	Diastolic (Lower limit) value of blood pressure measurement	Vital sign
(3)	BLOODPRESSURE_SYSTOLIC_MEAN	Systolic (Higher limit) value of blood pressure measurement	Vital sign
(4)	LACTATE_MEAN	Measure of lactic acid in the blood	Blood parameter
(5)	PH_VENOUS_MEAN	Blood gas measure Base excess (BE) indicates that all bases of gases in the arterial blood.	Blood parameter
(6)	TEMPERATURE_MEAN	Patient's body temperature	Vital sign
(7)	BIC_ARTERIAL_MEAN	Blood gas measure BE Base excess (BE) indicates that all bases of gases in the venous blood.	Blood parameter
(8)	SAT02_VENOUS_MEAN	Measure of the oxygen content of in the venous blood.	Blood parameter
(9)	CALCIUM_MEAN	Calcium mineral content in the blood, usually 1% of the body's calcium circulates in the blood.	Blood parameter
(10)	PC02_VENOUS_MEAN	Measure of partial pressure of carbon dioxide content in the venous blood	Blood parameter

$(\beta = depicting\ the\ balance\ level\ following\ synthetic\ data\ production)$

$$r_i = \frac{\Delta_i}{K_i}, \quad i = 1, 2, \ldots, ms \tag{5}$$

$$Normalize\, r_i = \hat{r}_i = \frac{r_i}{\sum_{i=1}^{ms} r_i} \tag{6}$$

$$Data\ to\ be\ generated\ for\ each\ ICU\ sample x_i = g_i = \hat{r}_i x G \tag{7}$$

SMOTE (Sample Minority Oversampling Technique) produces artificial points by moving the original closer to its Nearest Neighbor [31]. A random sampling of the minority class is prone to issues of overfitting. However, SMOTE deploys principles of interpolation. This algorithm selects a point randomly from the parameter space and one of its k nearest neighbors; these two points are connected randomly, forming a segment. A combination of these two points generates the new synthetic point.

K-nearest neighbor's principle is based on the Minkowski Distance [31]:

$$Distance\ between\ two\ points = d(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}} \tag{8}$$

Borderline SMOTE is an advancement in the traditional SMOTE technique. SMOTE often creates line bridges of synthetic values between the minority and majority classes in conditions where outliers are present in the dataset. This can cause errors in the classification of samples. Borderline SMOTE alleviates these issues and does not consider such outliers while creating artificial points [31].

SMOTE-Tomek conducts over and undersampling in tandem. The SMOTE technique involves oversampling the Tomek links, identifying all the nearest neighbor points across the classes, and removing them (Undersampling) [31].

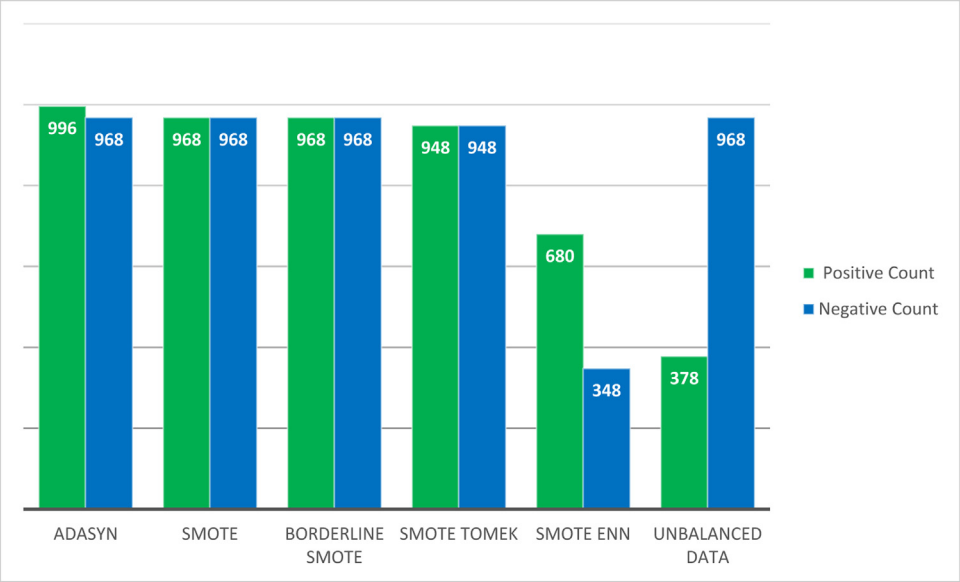

Fig. 4. Count of negative and positive class data samples before and after balancing with multiple techniques.

Table 3
Structure of 1-D CNN.

Model	Network description			
	Layer No.	Role	Activation function	Number of filters/Units
1-D CNN	Layer 1	Layer Conv 1D -1 (Input)	LeakyReLU	19 filters
	Layer 2	Layer Conv 1D-2	LeakyReLU	10 filters
	Layer 3	Layer Conv 1D-3	LeakyReLU	3 filters
	Layer 4	Max Pooling	NA	NA
	Layer 5	Dropout	NA	NA
	Layer 6	Flatten the output	NA	NA
	Layer 7	Dense layer 1	Sigmoidal	1 unit

SMOTE-ENN combines SMOTE and the concept of Edited Nearest Neighbor [32]. The ENN method uses principles from K-neighboring points. The majority class from these neighboring points is compared with the class of the data point. If the classes differ, then there occurs the deletion of the data point and its neighboring points. This technique ensures there is a large separation between the two classes.

3.4. Classifier methodology

After feature engineering, multiple architectures consist of five data balancing techniques and one with the original unbalanced data. Python libraries such as sklearn, Light GBM, and Catboost were deployed to create the machine learning pipelines. The balanced train datasets were used to model ten machine learning classifiers, namely: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) with the polynomial kernel, K-nearest Neighbor (KNN), Naïve Bayes (NB), Extratrees, Adaboost, Xgboost, and Catboost. Hyperparameter tuning was conducted for all the ML classifier pipelines with GridSearchCV. GridSearchCV tests all combinations of hyperparameters and narrows the model parameters to the optimal ones. Further, 5-fold cross-validation was executed. While exploring the deep learning domain for predicting the severity of a COVID-19 patient, we created a custom seven-layer 1-D CNN model with the TensorFlow library. A stochastic-gradient approach was utilized to obtain the optimum solution with Adaptive Moment Estimation (Adam) optimizer. Further, the model training was conducted for 200 epochs, and the binary cross-entropy function evaluated the loss in each epoch. The batch size consisted of 10 data samples. Details about the structure of this neural network are tabulated in Table 3.

4. Results and discussions

In this study, we observed that the performance of various classifiers varies with the type of data balancing technique. This section elaborates on the observed performance changes in the deployed pipelines. Further, Explainable AI techniques are applied to interpret the best-performing classifiers. SHAP, LIME, ELI5, Qlattice, and Anchor are deployed. These XAI techniques are validated by COVID-19 literature and feature the importance of the best-performing tree-based models.

4.1. Performance metrics

We evaluated the classification performance of the models using standard metrics such as precision, recall, f-1 score, accuracy, and Area Under the Receiver Operating Characteristic Curve (AUC) value [33]. When the ICU and Non-ICU cases are predicted correctly, True positives (TP) and True negatives (TN) predictions are obtained. When a Critically ill patient is wrongly classified as a mild case, such predictions are called False Negatives (FN). If a mild case is termed Critically ill, such predictions are categorized under False Positives (FP). When considering severity and risk prediction problems, FN is highly important as such false negative predictions can increase the risk of fatality.

Accuracy: A widely used performance metric is a ratio of number of correct predictions to the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Precision: Depicts a ratio of rightly categorized severe ICU patients and the total number of cases predicted as critically ill.

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

Table 4
Comparison of all deployed data-balancing classification pipelines.

Models	ADASYN					SMOTE				
	Accuracy	Precision	Recall	F1-score	AUC score	Accuracy	Precision	Recall	F1-score	AUC score
Logistic Regression	0.82	0.75	0.77	0.76	0.83	0.82	0.75	0.77	0.76	0.83
Decision Tree	0.8	0.73	0.75	0.74	0.78	0.79	0.72	0.74	0.73	0.79
Random Forest	0.85	0.8	0.81	0.8	0.88	0.86	0.81	0.81	0.81	0.88
SVM (Poly)	0.79	0.71	0.7	0.71	0.7	0.78	0.69	0.67	0.68	0.66
Naïve Bayes	0.8	0.73	0.71	0.72	0.81	0.8	0.73	0.7	0.71	0.81
K-Nearest Neighbor	0.78	0.69	0.7	0.7	0.7	0.79	0.71	0.71	0.71	0.71
Adaboost	0.83	0.77	0.79	0.78	0.85	0.86	0.8	0.82	0.81	0.88
Xgboost	0.87	0.82	0.82	0.82	0.88	0.87	0.82	0.82	0.82	0.9
Extratrees	0.81	0.74	0.79	0.76	0.84	0.8	0.73	0.74	0.73	0.84
Light GBM	0.85	0.8	0.8	0.8	0.83	0.86	0.8	0.81	0.8	0.88
Catboost	0.85	0.79	0.82	0.8	0.89	0.86	0.8	0.82	0.81	0.89
1-D CNN	0.78	0.71	0.76	0.73	0.81	0.83	0.77	0.8	0.78	0.87
Models	Borderline SMOTE					SMOTE Tomek				
	Accuracy	Precision	Recall	F1-score	AUC score	Accuracy	Precision	Recall	F1-score	AUC score
Logistic Regression	0.81	0.74	0.78	0.75	0.83	0.82	0.75	0.77	0.76	0.83
Decision Tree	0.81	0.73	0.75	0.74	0.76	0.8	0.73	0.75	0.74	0.77
Random Forest	0.87	0.81	0.83	0.82	0.88	0.86	0.81	0.82	0.81	0.89
SVM (Poly)	0.8	0.71	0.7	0.71	0.68	0.81	0.73	0.7	0.71	0.71
Naïve Bayes	0.8	0.73	0.71	0.72	0.81	0.81	0.73	0.7	0.71	0.81
K-Nearest Neighbor	0.79	0.71	0.72	0.71	0.72	0.81	0.73	0.72	0.73	0.72
Adaboost	0.84	0.78	0.8	0.79	0.84	0.84	0.78	0.8	0.79	0.87
Xgboost	0.87	0.81	0.81	0.81	0.89	0.87	0.82	0.81	0.82	0.87
Extratrees	0.81	0.73	0.71	0.72	0.81	0.8	0.72	0.73	0.72	0.84
Light GBM	0.87	0.81	0.82	0.82	0.85	0.85	0.79	0.8	0.8	0.85
Catboost	0.86	0.83	0.77	0.79	0.89	0.85	0.79	0.81	0.8	0.89
1-D CNN	0.81	0.74	0.77	0.75	0.89	0.83	0.77	0.78	0.78	0.84
Models	SMOTE ENN					Unbalanced data				
	Accuracy	Precision	Recall	F1-score	AUC score	Accuracy	Precision	Recall	F1-score	AUC score
Logistic Regression	0.79	0.73	0.78	0.74	0.82	0.85	0.82	0.72	0.75	0.83
Decision Tree	0.8	0.73	0.7	0.75	0.8	0.81	0.74	0.72	0.73	0.84
Random Forest	0.84	0.78	0.82	0.8	0.87	0.86	0.82	0.76	0.79	0.88
SVM (Poly)	0.79	0.72	0.73	0.72	0.75	0.8	0.73	0.64	0.66	0.65
Naïve Bayes	0.8	0.72	0.7	0.71	0.82	0.81	0.75	0.66	0.68	0.81
K-Nearest Neighbor	0.53	0.58	0.61	0.51	0.61	0.82	0.77	0.69	0.72	0.83
Adaboost	0.81	0.74	0.78	0.75	0.84	0.85	0.82	0.75	0.78	0.88
Xgboost	0.84	0.78	0.81	0.79	0.88	0.86	0.83	0.76	0.79	0.89
Extratrees	0.38	0.61	0.59	0.38	0.84	0.82	0.87	0.62	0.64	0.84
Light GBM	0.83	0.76	0.8	0.78	0.86	0.86	0.82	0.78	0.8	0.87
Catboost	0.85	0.79	0.82	0.8	0.86	0.86	0.67	0.77	0.72	0.89
1-D CNN	0.8	0.74	0.79	0.75	0.86	0.85	0.83	0.73	0.76	0.87

Recall: Also known as Sensitivity, depicts the true positive rate, the ability to categorize critically ill patients as ICU patients. 100% recall has all TP and zero FN predictions.

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

F1 score: This measure depicts the balance between Sensitivity and precision and is represented by the harmonic mean both values. The range of values is between 0 and 1. This metric gives information on the model's ability to precisely classify the data points and insights into the model's robustness.

$$F1 - score = 2\frac{Precision \times Recall}{Precision + Recall} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{12}$$

AUC (Area under the Receiver Operator curve): A probability curve ROC and AUC the measure of distinguishability between the ICU and Non-ICU class. Higher the AUC value, better is the model performance in classifying the cases correctly.

4.2. Evaluation of the models

Early severity prediction for COVID-19-positive patients can avoid cases of mistreatment, unoptimized usage of limited hospital facilities, and medical casualties. This study explores multiple pipelines utilizing combinations of classifiers such as LR, DT, RF, SVM, KNN, NB, Adaboost, Light GBM, Xgboost, Extratrees, Catboost, and 1-D CNN with five different data balancing techniques. It was observed that each pipeline has different algorithms and mathematical modeling to provide a classification. Table 4 tabulates the performance of the classifiers trained with different data. Models trained by ADASYN technique Xgboost was the best-performing model with an accuracy of 87%. Catboost, Light GBM, and RF provided a superior execution compared to the remaining ADASYN train models with an accuracy of 85%. When considering the classifiers trained on SMOTE data, it can be observed that Xgboost had the highest accuracy of 87%. And AUC of 0.9. Catboost, Light GBM, Random Forest, and Adaboost had a similar performance with an accuracy of 87%. Models like Adaboost and 1-D CNN improved performance compared to the corresponding ADASYN classifiers. Besides the SVM and DT classifiers, most models have more performance metrics than ADASYN train models. When trained on Borderline SMOTE data, models including DT, RF, SVM, and light Gradient Boosting Machine (GBM) had the highest performance. Specific classifiers, such as Xgboost, and NB, performed similarly in Borderline SMOTE and traditional SMOTE pipelines. Although similar accuracy for the Catboost model was observed in SMOTE and Border SMOTE, the Catboost model trained on Borderline SMOTE had a lower precision and recall. Comparing the performance obtained by the models trained on SMOTE Tomek performance of SVM, NB, and KNN observed an increase in performance than the abovementioned pipelines. However, the metrics for Random Forest, Extratrees, Light GBM, and Catboost were lowers than the Borderline SMOTE frameworks. There was a drastic drop in performance when considering SMOTE-ENN-trained models. DT had a similar score to SMOTE-Tomek.

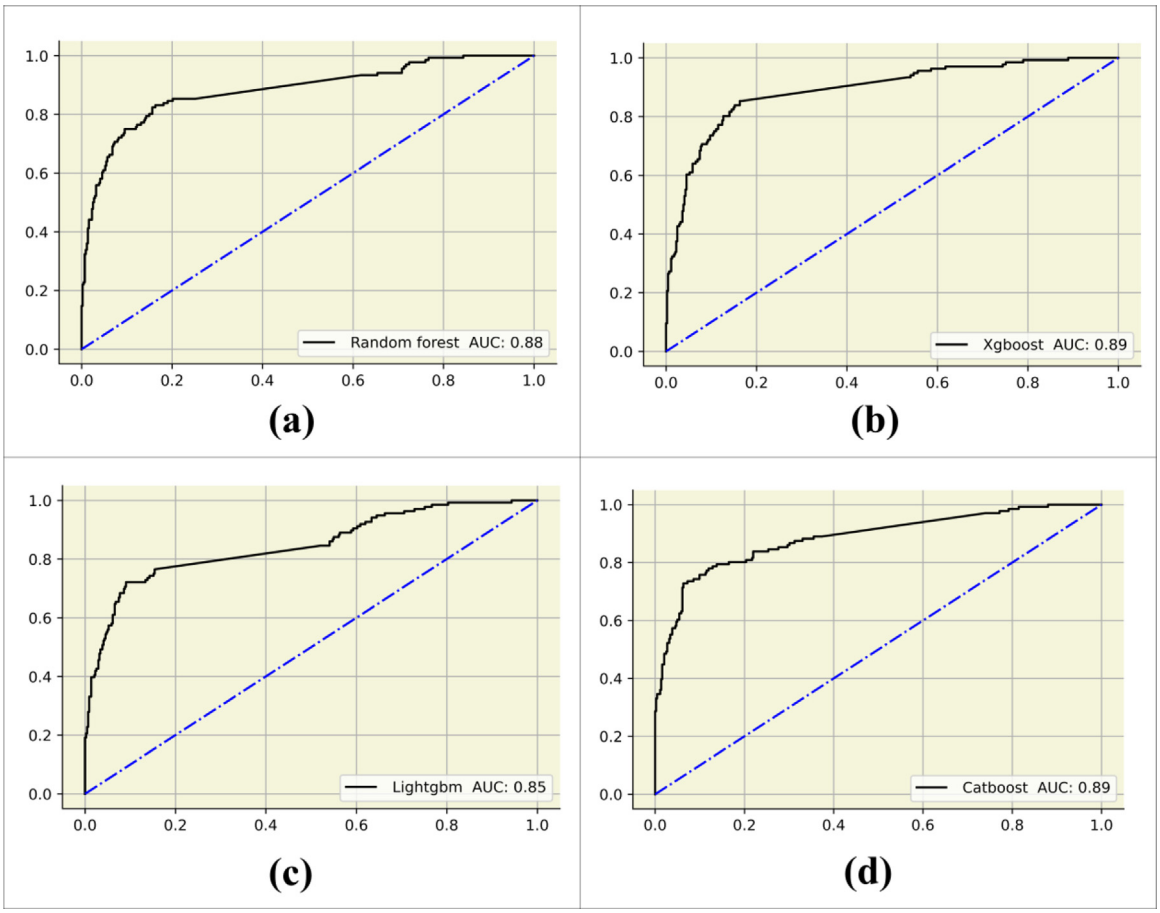

Fig. 5. AUC curves for the four high performing tree-based models trained on Borderline SMOTE. The AUC plots are as follows: (a) Random Forest, (b) Xgboost, (c) Light GBM and (d) Catboost.

However, most SMOTE-ENN pipeline models performed worse than the abovementioned data-balancing techniques. It was observed that models like LR and 1-D CNN gave the best outcomes when trained on unbalanced data. When compared to SMOTE-ENN, the performance of most models increases. However, due to the data imbalance, such models cannot provide higher recalls, as observed in Borderline SMOTE pipeline classifiers. In this study, recall metrics are given higher importance as these can correspond to the fatalities that could occur if a Severe COVID-19 patient is not assigned an ICU space. During performance evaluation, the highest recall was 83% for the Random Forest model trained on Borderline SMOTE data. Tree-based models such as Catboost and Light GBM consistently performed throughout various pipelines. Fig. 5 represents the AUC curves for the best-performing tree-based classifiers trained on Borderline SMOTE.

4.3. Explainable Artificial Intelligence

On model evaluation, it was observed that the tree-based models trained by Borderline SMOTE were more reliable than other deployed models. In this research, we aim to create classification pipelines and provide interpretability for the predictions made [34]. For explaining various classifiers, we utilized various XAI tools such as Shapley Additive Explanations, Local Interpretable Model-Agonistic Explanations, Quantum Lattice, ELI5, and Anchor. These XAI techniques are then validated by four tree-based feature importance graphs along with medical literature.

Shapley Additive Explanations (SHAP): SHAP is a widely used game theoretic classifier explanation tool. This mathematical model was proposed by Lundberg et al. [35]. Shapley values are evaluated for each feature based on their contribution to a prediction. Fig. 6 depicts the SHAP Beeswarm plot obtained for the Random Forest classifier. The beeswarm plot provides an information-dense global interpretation summary for the random forest classifier. Each dot on the plot represents one data point. The feature names are placed along the y-axis in descending importance. The SHAP values are present along the x-axis. The color gradient indicates the feature values. The redder the dot on the plot, the higher its feature value. In this plot, it can be observed that the Respiratory Rate mean has the most significant contribution to the output. Higher values of features such as Respiratory Rate mean, Calcium mean, and Systolic Blood pressure value had a positive correlation with the SHAP values. This positive relation indicates that a higher value of these features could increase the patient's probability of being classified as severe (in need of an ICU). These findings coincide with the research conducted regarding COVID-19 severity [11].

Local Interpretable Model-Agonistic Explanations (LIME): LIME is another classifier explanations tool. This technique can provide interpretations for individual predictions. This method was proposed by Ribeiro [36]. This algorithm evaluates the predictions obtained by modifying the inputs. The local predictions provided by the Light GBM model trained on Borderline SMOTE data was explained by LIME. This tool provides a plot of individual feature contribution to a prediction. Fig. 7 illustrates two plots. Plot 7(a) represents an explanation for a patient being predicted as a severe case with a high probability. It can be observed that most important features like Respiratory Rate, Systolic blood pressure, The Venous PH, calcium mean and patient temperature has contributed towards increasing the probability of the patient being predicted severe. Plot 7(b) indicates a visualization explaining a prediction made for a mild COVID-19 case. It is observed that apart from respiratory rate and calcium mean, most features have contributed towards producing a result indicating the patient has a mild case of

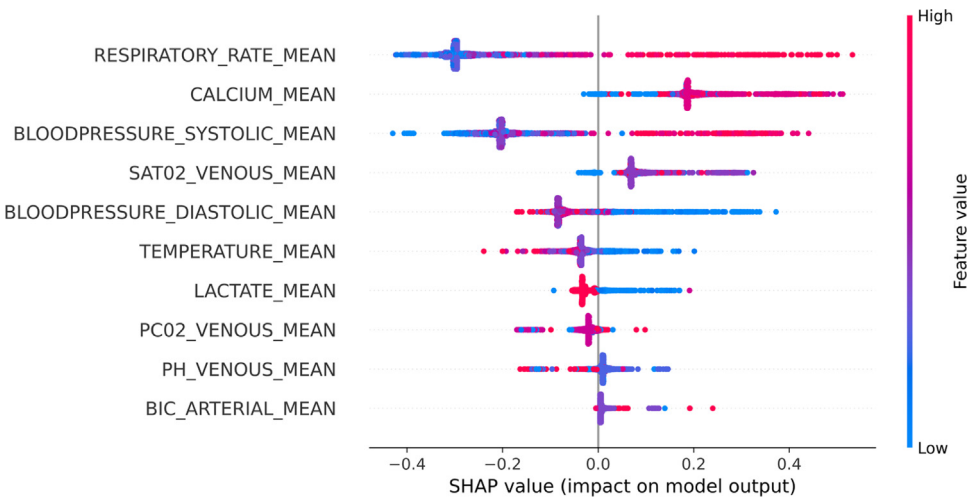

Fig. 6. SHAP plot for explaining the random forest classifier.

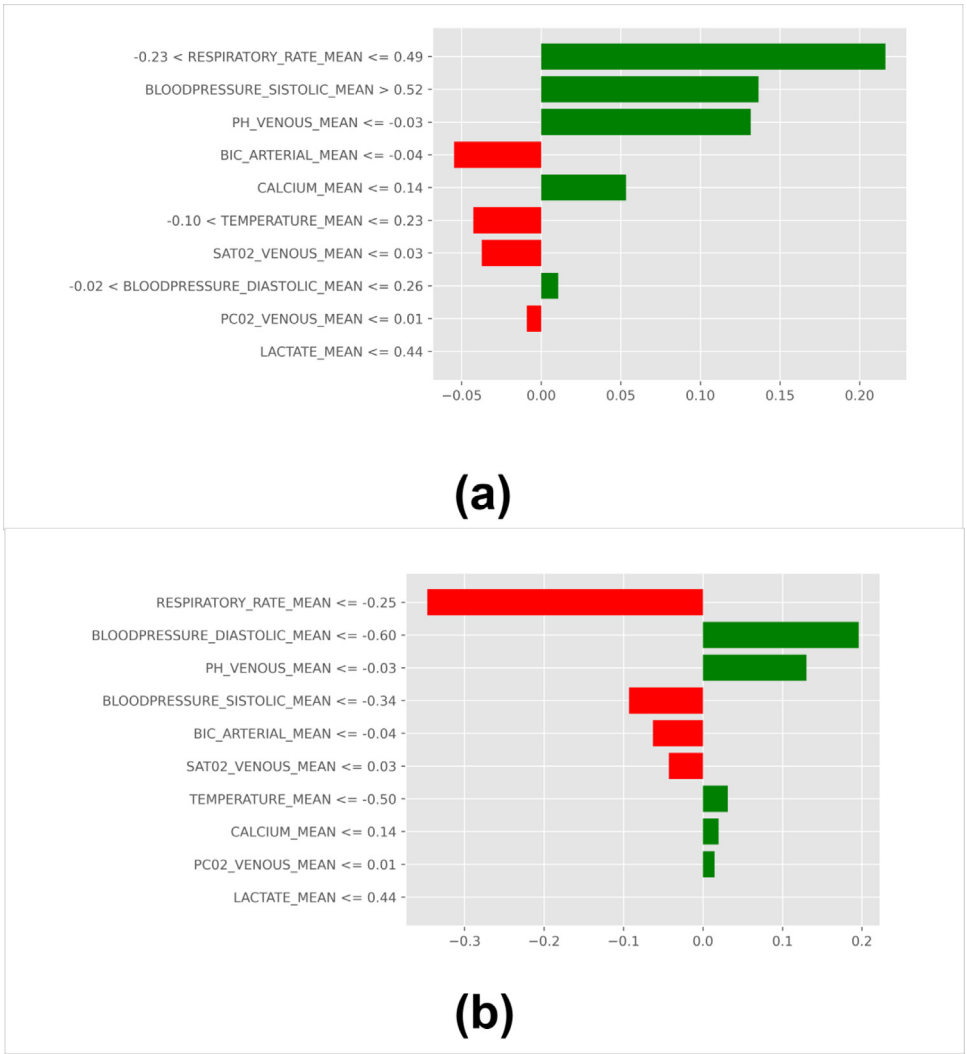

Fig. 7. Local Interpretable Model-Agonist for Light GBM.

COVID-19. These findings obtained by the LIME plot aligns with the literature [9,12].

Explain Like I'm 5 (ELI5 tool): This is a python package that not only assists in explaining predictions but also in debugging the machine

Weight	Feature
0.4228	RESPIRATORY_RATE_MEAN
0.1626	BLOODPRESSURE_DIASTOLIC_MEAN
0.1454	BLOODPRESSURE_SYSTOLIC_MEAN
0.1339	TEMPERATURE_MEAN
0.0478	CALCIUM_MEAN
0.0328	LACTATE_MEAN
0.0229	SAT02_VENOUS_MEAN
0.0149	PH_VENOUS_MEAN
0.0085	BIC_ARTERIAL_MEAN
0.0083	PC02_VENOUS_MEAN

(a)

y=0 (probability **0.913**, score **-2.352**) top features

Contribution?	Feature	Value
+1.496	TEMPERATURE_MEAN	2.282
+0.793	CALCIUM_MEAN	0.137
+0.443	BLOODPRESSURE_DIASTOLIC_MEAN	0.780
+0.230	PC02_VENOUS_MEAN	0.008
+0.195	SAT02_VENOUS_MEAN	0.029
+0.136	RESPIRATORY_RATE_MEAN	0.035
+0.095	BIC_ARTERIAL_MEAN	-0.040
+0.059	PH_VENOUS_MEAN	-0.027
-0.159	BLOODPRESSURE_SYSTOLIC_MEAN	0.529
-0.340	<BIAS>	1.000
-0.598	LACTATE_MEAN	-2.303

(b)

Fig. 8. ELI5 explanation plots: (a) Global explanation, (b) Local explanation.

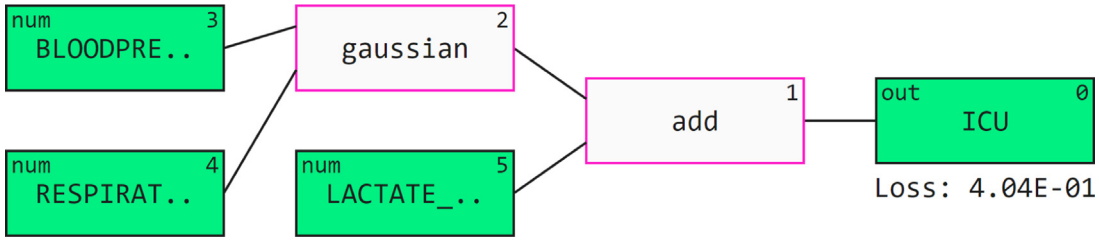

Fig. 9. QGraph plot.

learning classifier. This tool supports tree-based classifiers. In this study, ELI5 was used to explain the Xgboost model. ELI5 can provide both local and global explanations [34]. Fig. 8 provides the tabular interpretation for the prediction made by Xgboost. In plot 8(a) a list of the features is depicted, the highest contributing feature is present on the topmost row. Respiratory rates mean of the patient was given the highest weight and was selected as the root-node while constructing a tree-based model. The weights of all features are depicted from highest to lowest corresponding to their contribution in the prediction. Plot 8(b) represents a single prediction providing a feature-wise deconstructed view into the rationale behind a prediction. In this case the prediction was 0, depicting the patient was a mild case and did not require an ICU. These values of features can be validated by research conducted on COVID-19 patient severity [10,11].

Quantum Lattice (QLattice): This tool is inspired by the Richard Feynman's path. It creates a new model to accurately fit the data. This technique provides a good performance even when the amount of data is less. The method was first proposed by Abzu [37]. This supervised machine learning approach analyses the data and accordingly creates multiple models, multiple iterations take place, and the best fit model is selected. In this study it was observed that the systolic value of the patients' blood pressure along with respiratory rate mean and lactate mean were significant features that were used to create a COVID-19 severity prediction classifier. On selection of the best model, a Qgraph is created to provide the rationale behind a prediction. Fig. 9 depicts the QGraph obtained. Further, a simplified equation for the model is obtained that provides insights into the mapping of inputs to outputs.

$$logreg(-0.84 * LACTATEMEAN - 3.76e^{-1.89(BLOODPRESSURESYSTOLICMEAN+0.415)^2-23.5(RESPIRATORYRATEMEAN+0.608)^2} + 1.67) \quad (13)$$

Anchors: These are local conditions of features that assist provide high precision predictions. Ribeiro [38] proposed that black-box classifiers can be explained by such decision rules. Anchors adopt a concept of coverage wherein its found to how many other possibly unseen instances can these conditions be applied to. Table 5 illustrates the anchor conditions created for five negative and five positive data samples classified by the best performing model, Random Forest. The dataset being previously normalized does not provide the exact values of the features. However, conditions were created based on the patterns observed in the dataset. The top five represents the negative predictions, wherein such patients were not categorized as severe. Considering the first negative prediction, this prediction is explained by two feature conditions. The respiratory rate mean was less than or equal to −0.23 and the Blood pressure systolic mean was less than or equal to −0.12. These two anchors were able to explain the negative classification with a precision of 83% and a coverage of 44%. The bottom five rows of the table represent the positive predictions. Observing the sixth row, the positive prediction is explained by two conditions. The respiratory mean was higher was more than 0.33 and systolic blood pressure of the patient is more than −0.12. These two anchors can explain the positive prediction (wherein the patient is in severe condition and required intensive care) with a precision of 97% and coverage of 15%. As observed previously, anomalies in non-invasive parameters such as respiratory rate, systolic and diastolic blood pressure have a significant contribution in increasing the severity of a COVID-19 patient.

Feature Importance:

The technique of feature importance entails the estimation of scores for all input features for a given classifier. These scores provide an understanding of the significance and relevance of the features in a particular prediction. A higher importance score corresponds to a more significant effect of the feature on the prediction model. In this study, we have used feature importance to validate the findings obtained by XAI tools. Fig. 10 illustrates the feature importance plots obtained by four best-performing tree-based ML classifiers. Plot 10(a) depicts the random Forest feature importance, wherein respiratory rate, blood pressure, and temperature are given higher scores. Plot 10(b) provided a diagrammatic representation of the feature importance of the Xgboost model. In this plot, respiratory rate and lactate levels carried higher importance. The graph in 10(c) illustrates the importance of the Light GBM feature importance. Further, plot 10(d) depicts the Catboost feature importance. Similar to 10(a), both 10(c) and (d) assigned

Table 5
Anchor conditions with precision and coverage for 5 negative and 5 positive patients.

Prediction	Anchors		Precision	Coverage
	First parameter condition	Second parameter condition		
Negative	RESPIRATORY_RATE_MEAN ≤ −0.23	BLOODPRESSURE_SYSTOLIC_MEAN ≤ −0.12	0.83	0.44
Negative	RESPIRATORY_RATE_MEAN ≤ −0.23	BLOODPRESSURE_SYSTOLIC_MEAN ≤ −0.12	0.85	0.44
Negative	RESPIRATORY_RATE_MEAN ≤ −0.28	BLOODPRESSURE_SYSTOLIC_MEAN ≤ −0.12	0.86	0.13
Negative	RESPIRATORY_RATE_MEAN ≤ −0.28	BLOODPRESSURE_SYSTOLIC_MEAN ≤ −0.12	0.87	0.13
Negative	RESPIRATORY_RATE_MEAN ≤ −0.23	BLOODPRESSURE_SYSTOLIC_MEAN ≤ −0.12	0.84	0.44
Positive	RESPIRATORY_RATE_MEAN > 0.33	BLOODPRESSURE_SYSTOLIC_MEAN > −0.12	0.97	0.15
Positive	BLOODPRESSURE_SYSTOLIC_MEAN ≤ 0.41	BLOODPRESSURE_DIASTOLIC_MEAN ≤ −0.51	0.85	0.08
Positive	RESPIRATORY_RATE_MEAN ≤ −0.28	BLOODPRESSURE_SYSTOLIC_MEAN ≤ 0.41	0.82	0.17
Positive	RESPIRATORY_RATE_MEAN ≤ −0.28	BLOODPRESSURE_DIASTOLIC_MEAN > 0.26	0.83	0.18
Positive	RESPIRATORY_RATE_MEAN > 0.33	BLOODPRESSURE_SYSTOLIC_MEAN > −0.12	0.98	0.15

Table 6
Model evaluation with the seven most important features.

Model	Accuracy	Precision	Recall	AUC
Random Forest	0.8	0.8	0.8	0.9
Xgboost	0.8	0.81	0.81	0.89
Light GBM	0.81	0.81	0.81	0.9
Catboost	0.82	0.83	0.83	0.9

the highest feature importance to respiratory rate, temperature, and blood pressure indicating their significant contribution to the COVID-19 severity prediction.

4.4. Further discussion

The COVID-19 pandemic posed a grave threat to the healthcare infrastructure. There was a shortage of medical supplies, equipment, and ICU space. Multiple fatalities occurred from the mistreatment and suboptimal usage of medical infrastructure. An early severity prediction system could assist in managing and optimizing medical supplies. In this study, we have developed and presented a comparative study of the functioning of various frameworks and their effect on classifier prediction. We deployed multiple combinations of five balancing tools and twelve models to predict the severity of the size of 1925 COVID-19-positive patients in Brazil. The best-performing pipeline consisted of Random Forest trained on Borderline SMOTE with the highest recall of 83%. Various Explainable Artificial Intelligence tools, SHAP, LIME, Qlattice, ELI5, and Anchor, explained all the high-performing tree-based models. The results of these tools were validated by the feature importance graphs obtained from the tree-based classifiers. It was observed that most XAI tools align with research suggesting that anomalies in respiratory rate, blood pressure, temperature, lactate, and calcium levels could increase the risk of the patient's condition worsening. Various studies agree with the findings obtained. Respiratory failures often cause COVID-19 fatalities; World Health Organization (WHO) advised special medical attention to patients experiencing shortness of breath [39]. Difficulty breathing is reflected in the patient's vitals, such as respiratory rate, Oxygen saturation, and partial pressure of oxygen and carbon dioxide in the blood. Through this research, we observed the impact of individual features on the severity of the patient.

We observed that the top four best-performing models were trained by borderline SMOTE: Random Forest, Xgboost, Light GBM, and Catboost. The seven most important features were selected to train these classifiers. On further model training and evaluation, we obtained the results tabulated in Table 6. It was understood that our original models tend to perform higher while providing a holistic prediction with multiple parameters.

Multiple studies have been conducted on this dataset to predict the COVID-19 severity risk. Aktar et al. [14] deployed models on 545 data samples. Models such as KNN, SVM, RF, and DT were used, and the highest accuracy of 93% for KNN. 94% was the highest F-1 Score achieved by the RF model. In contrast, the KNN and ANN models gave comparatively lower F-1 scores at 74% and 79%. They obtained several significant features like Respiratory Rate, blood lactate, Oxygen saturation, and blood pressure. Further, Parbate et al. proposed a study that primarily focused on using the RF classifier and Xgboost techniques to improve the overall model performance. The data needed to be balanced. Furthermore, they selected 31 features, including Gender, Potassium, and Urea values. RF achieved high Recall and F-1 scores of 97% and 91%, respectively. However, Xgboost gave higher accuracy, precision, and AUC scores approximating 97%, 97%, 96%, and 96%, respectively, compared to the RF classifier [40].

Chung et al. proposed a neural network architecture to predict the severity with 96% accuracy. They selected 178 clinical features and built a 'COVID-Net Clinical ICU' model. They trained the model by Adam optimizer deploying binary cross for 1000 epochs. The data needed to be balanced. However, they concluded, Heart rate, sodium, and potassium levels were among a few features that significantly contributed to the severity of the patients. Further, they observed that 50 of the selected 178 produced a negative importance score [41]. The rest of the research is described in Table 7.

Various clinical markers which could predict a patient's COVID-19 prognosis were studied in this research. These critical markers can be monitored to protect the patient from severe COVID-19. Appropriate treatments and medicines can be provided to the patients if these markers fluctuate. Further, five XAI methods have been used to interpret the results obtained from the best models. XAI visualizations can be useful to the medical professionals to decide the course of the treatment. Data imbalance is very common in medical machine learning. However, the results obtained by the models are not reliable when there is an imbalance between the two classes. Hence, five data balancing techniques have been used and compared. The decision support system developed in this study can be useful to predict a COVID-19 patient's prognosis in advance. Valuable lives can be saved and judicial use of medical resources such as ventilators and ICU beds can be made.

Providing care to a critically ill patient is highly time-sensitive. The significant factors mentioned can be tested for in-patients. It is cheaper to conduct blood and vital body tests than imaging scans. Clinics in remote areas without abundant modern equipment can record vital signs. With the assistance of a decision-making tool, prognosis and severity prediction for COVID-19 can be made accessible.

5. Conclusion

COVID-19 caused a shortage of hospital resources and health practitioners. The lack of available ICU accommodation adversely impacted the lives of all critically ill patients. Treatment of a severe case is time sensitive. With the hospitals overburdened with the continuous inflow of patients, providing optimal care to each individual becomes vital. We created a decision-making model wherein the severity of the patients can be assessed early on using machine learning. The effect of five different balancing techniques and using twelve models

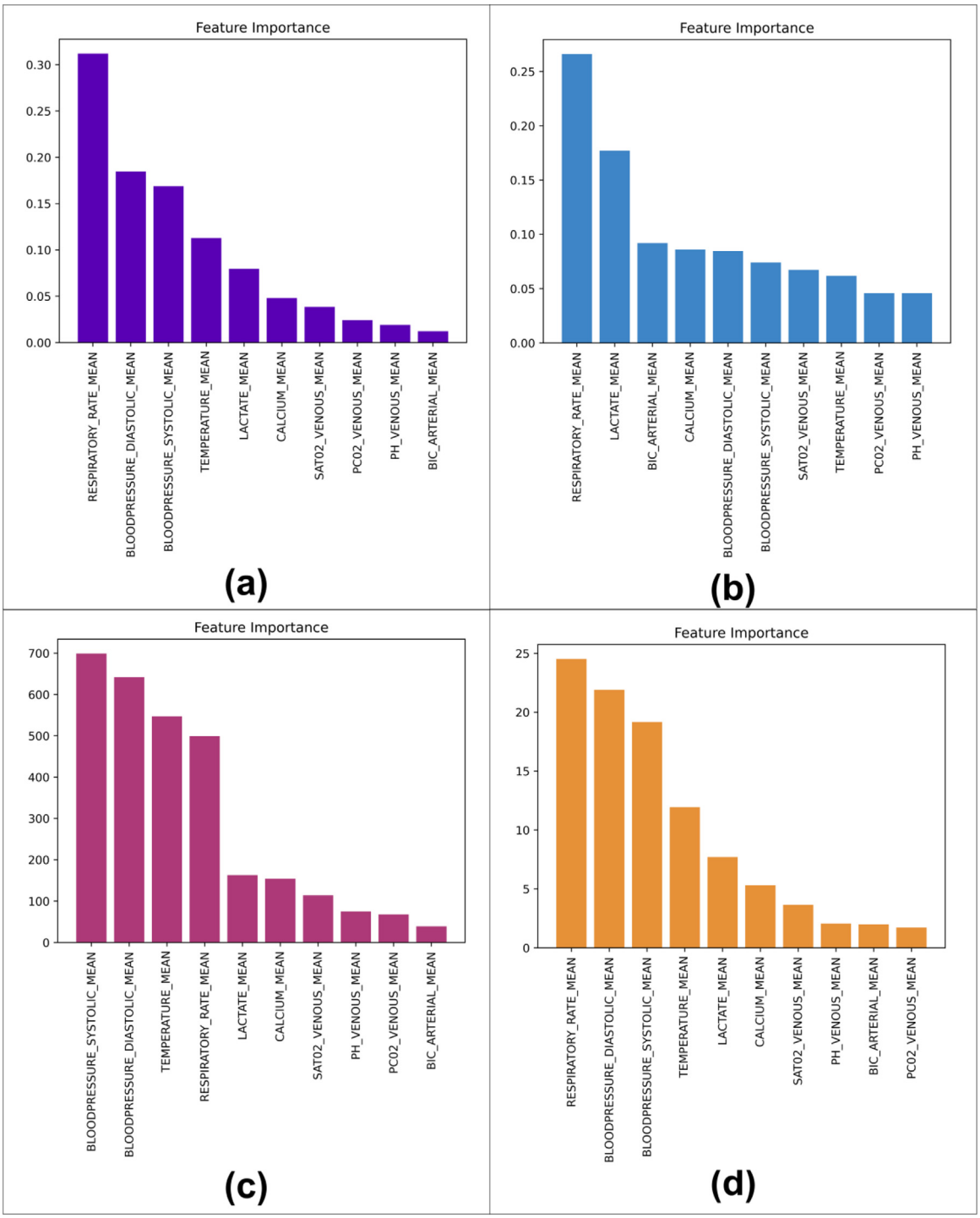

Fig. 10. Tree-based feature importance, plot (a) Random forest, (b) Xgboost, (c) Light GBM and (d) Catboost.

Table 7
Comparing this study with previously proposed models.

Reference	Samples considered	Data balancing algorithm	Models used	Best accuracy score	Best recall score	Best F1-score	AUC
[42]	1110	–	LR, SVM, ANN	95%	–	–	95%
[43]	1925	–	RF, DT, Gradient Boosting, ANN	86.96%	–	–	–
Proposed study	1925	ADASYN, SMOTE, Borderline-SMOTE, SMOTE-Tomek, SMOTE-ENN	LR, DT, RF, SVM, NB, KNN, Xgboost, Adaboost, Light GBM, Extratrees, Catboost, 1-D CNN	87%	83%	82%	89%

gave an in-depth view of performance-improving factors. Initially, we analyzed the parameters and their significance. Later, we deployed the ML models to accurately predict the probability of a patient's admission to the ICU. We obtained valuable insights into the rationale behind the best-performing models with explainable AI tools such as SHAP, LIME, ELI5, Qlattice, and Anchor. The XAI tools were validated with feature importance and previously proposed research. We deduced that anomalies in features such as Respiratory Rate, Blood pressure, Body temperature, calcium, and lactate levels positively contribute to the patient severity.

Such automated systems could be integrated with medical triaging architectures. The advancements in XAI could assist in bridging the gap between the medical and Artificial Intelligence domains. Nevertheless, multiple testing, advanced architectures, and higher quality datasets followed by validation from healthcare professionals can help improve the accuracy, interpretation, and performance in the future.

CRediT authorship contribution statement

Varada Vivek Khanna: Software and intial draft preparation. **Krishnaraj Chadaga:** Conceptualization, Review & editing. **Niranjana Sampathila:** Supervision, Formal analysis. **Srikanth Prabhu:** Visualization, Validation. **Rajagopala Chadaga P.:** Review and editing, resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Funding

The authors did not receive funding from any sources.

References

[1] T. Singhal, A review of coronavirus disease-2019 (COVID-19), Indian J. Pediatr. 87 (4) (2020) 281–286.

[2] S.C. Cheng, Y.C. Chang, Y.L.F. Chiang, Y.C. Chien, M. Cheng, C.H. Yang ..., Y.N. Hsu, First case of coronavirus disease 2019 (COVID-19) pneumonia in Taiwan, J. Formos. Med. Assoc. 119 (3) (2020) 747–751.

[3] M.A. Andrews, B. Areekal, K.R. Rajesh, J. Krishnan, R. Suryakala, B. Krishnan ., P.V. Santhosh, First confirmed case of COVID-19 infection in India: A case report, Indian J. Med. Res. 151 (5) (2020) 490.

[4] M.A. Andrews, B. Areekal, K.R. Rajesh, J. Krishnan, R. Suryakala, B. Krishnan ., P.V. Santhosh, First confirmed case of COVID-19 infection in India: A case report, Indian J. Med. Res. 151 (5) (2020) 490.

[5] M. Döhla, C. Boesecke, B. Schulte, C. Diegmann, E. Sib, E. Richter ., H. Streeck, Rapid point-of-care testing for SARS-CoV-2 in a community screening setting shows low sensitivity, Public Health 182 (2020) 170–172.

[6] S. Helleringer, B.L. Queiroz, Commentary: Measuring excess mortality due to the COVID-19 pandemic: progress and persistent challenges, Int. J. Epidemiol. 51 (1) (2022) 85–87.

[7] S. Beyaz, A brief history of artificial intelligence and robotic surgery in orthopedics & traumatology and future expectations, Jt. Dis. Relat. Surg. 31 (3) (2020) 653.

[8] E. Egorov, C. Pieters, H. Korach-Rechtman, J. Shklover, A. Schroeder, Robotics, microfluidics, nanotechnology and AI in the synthesis and evaluation of liposomes and polymeric drug delivery systems, Drug Deliv. Transl. Res. 11 (2) (2021) 345–352.

[9] A. Pourbagheri-Sigaroodi, D. Bashash, F. Fateh, H. Abolghasemi, Laboratory findings in COVID-19 diagnosis and prognosis, Clin. Chim. Acta 510 (2020) 475–482.

[10] C. Fernández-de Las-Peñas, D. Palacios-Ceña, V. Gómez-Mayordomo, L.L. Florencio, M.L. Cuadrado, G. Plaza-Manzano, M. Navarro-Santana, Prevalence of post-COVID-19 symptoms in hospitalized and non-hospitalized COVID-19 survivors: A systematic review and meta-analysis, Eur. J. Intern. Med. 92 (2021) 55–70.

[11] A. Natarajan, H.W. Su, C. Heneghan, L. Blunt, C. O'Connor, L. Niehaus, Measurement of respiratory rate using wearable devices and applications to COVID-19 detection, NPJ Digit. Med. 4 (1) (2021) 1–10.

[12] V.V. Khanna, K. Chadaga, N. Sampathila, S. Prabhu, R. Chadaga, S. Umakanth, Diagnosing COVID-19 using artificial intelligence: a comprehensive review, Netw. Model. Anal. Health Inform. Bioinform. 11 (1) (2022) 1–23.

[13] Sirio-libanes COVID-19—Clinical data to assess diagnosis. Kaggle, 2022, Available online: https://www.kaggle.com/datasets/S\'{\i}rio-Libanes/covid19 assessed on 15 July 2022.

[14] S. Aktar, M.M. Ahamad, M. Rashed-Al-Mahfuz, A.K.M. Azad, S. Uddin, A.H.M. Kamal ., M.A. Moni, Machine learning approach to predicting COVID-19 disease severity based on clinical blood test data: statistical analysis and model development, JMIR Med. Inform. 9 (4) (2021) e25884.

[15] M.A. Deif, A.A. Solyman, M.H. Alsharif, P. Uthansakul, Automated triage system for intensive care admissions during the COVID-19 pandemic using hybrid Xgboost-AHP approach, Sensors 21 (19) (2021) 6379.

[16] Y. Statsenko, F. Al Zahmi, T. Habuza, K. Neidl-Van Gorkom, N. Zaki, Prediction of COVID-19 severity using laboratory findings on admission: informative values, thresholds, ML model performance, BMJ Open 11 (2) (2021) e044500.

[17] Y. Xiong, Y. Ma, L. Ruan, D. Li, C. Lu, L. Huang, Comparing different machine learning techniques for predicting COVID-19 severity, Infect. Dis. Poverty 11 (1) (2022) 1–9.

[18] V. Schöning, E. Liakoni, C. Baumgartner, A.K. Exadaktylos, W.E. Hautz, A. Atkinson, F. Hammann, Developing and validating a prognostic COVID-19 severity assessment (COSA) score and machine learning models for patient triage at a tertiary hospital, J. Transl. Med. 19 (1) (2021) 1–11.

[19] E. Karlafti, A. Anagnostis, E. Kotzakioulafi, M.C. Vittoraki, A. Eufraimidou, K. Kasarjyan ., T. Didangelos, Does COVID-19 clinical status associate with outcome severity? An unsupervised machine learning approach for knowledge extraction, J. Pers. Med. 11 (12) (2021) 1380.

[20] S.S. Aljameel, I.U. Khan, N. Aslam, M. Aljabri, E.S. Alsulmi, Machine learning-based model to predict the disease severity and outcome in COVID-19 patients, Sci. Program. 2021 (2021).

[21] T.D. Bennett, R.A. Moffitt, J.G. Hajagos, B. Amor, A. Anand, M.M. Bissell ., N3C Consortium, The national COVID cohort collaborative: clinical characterization and early severity prediction, MedRxiv (2021).

[22] K. Zhou, Y. Sun, L. Li, Z. Zang, J. Wang, J. Li ., T. Guo, Eleven routine clinical features predict COVID-19 severity uncovered by machine learning of longitudinal measurements, Comput. Struct. Biotechnol. J. 19 (2021) 3640–3649.

[23] T. Rahman, F.A. Al-Ishaq, F.S. Al-Mohannadi, R.S. Mubarak, M.H. Al-Hitmi, K.R. Islam ., M.E. Chowdhury, Mortality prediction utilizing blood biomarkers to predict the severity of COVID-19 using machine learning technique, Diagnostics 11 (9) (2021) 1582.

[24] C. Hu, Z. Liu, Y. Jiang, O. Shi, X. Zhang, K. Xu ., X. Chen, Early prediction of mortality risk among patients with severe COVID-19, using machine learning, Int. J. Epidemiol. 49 (6) (2020) 1918–1929.

[25] J. Kang, T. Chen, H. Luo, Y. Luo, G. Du, M. Jiming-Yang, Machine learning predictive model for severe COVID-19, Infect. Genet. Evol. 90 (2021) 104737.

[26] D. Assaf, Y.A. Gutman, Y. Neuman, G. Segal, S. Amit, S. Gefen-Halevi ., A. Tirosh, Utilization of machine-learning models to accurately predict the risk for critical COVID-19, Intern. Emerg. Med. 15 (8) (2020) 1435–1443.

[27] L. Yan, H.T. Zhang, Y. Xiao, M. Wang, C. Sun, J. Liang ., X. Tang, Prediction of survival for severe Covid-19 patients with three clinical features: development of a machine learning-based prognostic model with clinical data in wuhan, MedRxiv (2020).

[28] H. Zhou, X. Wang, R. Zhu, Feature selection based on mutual information with correlation coefficient, Appl. Intell. 52 (5) (2022) 5457–5474.

[29] P. Good, Robustness of pearson correlation, Interstat 15 (5) (2009) 1–6.

[30] H. He, Y. Bai, E.A. Garcia, S. Li, ADASYN: Adaptive synthetic sampling approach for imbalanced learning, in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), IEEE, 2008, pp. 1322–1328.

[31] Q. Ning, X. Zhao, Z. Ma, A novel method for identification of glutarylation sites combining borderline-SMOTE with tomek links technique in imbalanced data, IEEE/ACM Trans. Comput. Biol. Bioinform. 19 (5) (2021) 2632–2641.

[32] M. Lin, X. Zhu, T. Hua, X. Tang, G. Tu, X. Chen, Detection of ionospheric scintillation based on xgboost model improved by smote-enn technique, Remote Sens. 13 (13) (2021) 2577.

[33] A. Botchkarev, Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology, 2018, arXiv preprint arXiv:1809.03006.

[34] V.V. Khanna, K. Chadaga, N. Sampathila, S. Prabhu, V. Bhandage, G.K. Hegde, A distinctive explainable machine learning framework for detection of polycystic ovary syndrome, Appl. Syst. Innov. 6 (2) (2023) 32.

[35] S.M. Lundberg, S.I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. (30) (2017).
[36] M.T. Ribeiro, S. Singh, C. Guestrin, Model-agnostic interpretability of machine learning, 2016, arXiv preprint arXiv:1606.05386.
[37] K.R. Broløs, M.V. Machado, C. Cave, J. Kasak, V. Stentoft-Hansen, V.G. Batanero ., C. Wilstrup, An approach to symbolic regression using feyn, 2021, arXiv preprint arXiv:2104.05417.
[38] M.T. Ribeiro, S. Singh, C. Guestrin, Anchors: High-precision model-agnostic explanations, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, 2018, No. 1.
[39] https://www.who.int/health-topics/coronavirus#tab=tab_1 accessed on 24 July 2022.
[40] N.D. Parbate, A.P. Palande, S.K. Wagh, A. Kamble, S. Isave, ICU admission prediction using machine learning for Covid-19 patients, in: 2021 International Conference on Communication Information and Computing Technology, ICCICT, IEEE, 2021, pp. 1–6.
[41] A. Chung, M. Famouri, A. Hryniowski, A. Wong, COVID-Net Clinical ICU: Enhanced prediction of ICU admission for COVID-19 patients via explainability and trust quantification, 2021, arXiv preprint arXiv:2109.06711.
[42] N. Alhawas, S. Kartal, Machine learning-based predictors for ICU admission of COVID-19 patients, in: International Summit Smart City 360°, Springer, Cham, 2022, pp. 602–616.
[43] S. Rihana, C.B. Rjeily, Artificial intelligence framework for COVID19 patients monitoring, in: 2021 Sixth International Conference on Advances in Biomedical Engineering, ICABME, IEEE, 2021, pp. 213–218.