

Original Paper

Prognostic Assessment of COVID-19 in the Intensive Care Unit by Machine Learning Methods: Model Development and Validation

Pan Pan¹, MD; Yichao Li², BSc; Yongjiu Xiao³, MD; Bingchao Han⁴, MD; Longxiang Su⁵, MD; Mingliang Su², BSc; Yansheng Li², BSc; Siqi Zhang², BSc; Dapeng Jiang², BSc; Xia Chen², BSc; Fuquan Zhou², BSc; Ling Ma³, BSc; Pengtao Bao⁶, MD; Lixin Xie⁶, MD

¹Chinese PLA General Hospital, Medical School Of Chinese PLA, College of Pulmonary and Critical Care Medicine, Beijing, China

²DHC Mediway Technology Co Ltd, Beijing, China

³The 940th Hospital of Jiont Logistics Support Force of Chinese People's Liberation Army, Lanzhou, China

⁴The 980th Hospital of Jiont Logistics Support Force of Chinese People's Liberation Army, Shijiazhuang, China

⁵Peking Union Medical College Hospital, Beijing, China

⁶College of Pulmonary and Critical Care Medicine, Chinese PLA General Hospital, Beijing, China

Corresponding Author:

Lixin Xie, MD

College of Pulmonary and Critical Care Medicine

Chinese PLA General Hospital

No.28 Fuxing Road, Haidian District, Beijing

Beijing

China

Phone: 86 55499128

Email: xielx301@126.com

Abstract

Background: Patients with COVID-19 in the intensive care unit (ICU) have a high mortality rate, and methods to assess patients' prognosis early and administer precise treatment are of great significance.

Objective: The aim of this study was to use machine learning to construct a model for the analysis of risk factors and prediction of mortality among ICU patients with COVID-19.

Methods: In this study, 123 patients with COVID-19 in the ICU of Vulcan Hill Hospital were retrospectively selected from the database, and the data were randomly divided into a training data set (n=98) and test data set (n=25) with a 4:1 ratio. Significance tests, correlation analysis, and factor analysis were used to screen 100 potential risk factors individually. Conventional logistic regression methods and four machine learning algorithms were used to construct the risk prediction model for the prognosis of patients with COVID-19 in the ICU. The performance of these machine learning models was measured by the area under the receiver operating characteristic curve (AUC). Interpretation and evaluation of the risk prediction model were performed using calibration curves, SHapley Additive exPlanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), etc, to ensure its stability and reliability. The outcome was based on the ICU deaths recorded from the database.

Results: Layer-by-layer screening of 100 potential risk factors finally revealed 8 important risk factors that were included in the risk prediction model: lymphocyte percentage, prothrombin time, lactate dehydrogenase, total bilirubin, eosinophil percentage, creatinine, neutrophil percentage, and albumin level. Finally, an eXtreme Gradient Boosting (XGBoost) model established with the 8 important risk factors showed the best recognition ability in the training set of 5-fold cross validation (AUC=0.86) and the verification queue (AUC=0.92). The calibration curve showed that the risk predicted by the model was in good agreement with the actual risk. In addition, using the SHAP and LIME algorithms, feature interpretation and sample prediction interpretation algorithms of the XGBoost black box model were implemented. Additionally, the model was translated into a web-based risk calculator that is freely available for public usage.

Conclusions: The 8-factor XGBoost model predicts risk of death in ICU patients with COVID-19 well; it initially demonstrates stability and can be used effectively to predict COVID-19 prognosis in ICU patients.

(*J Med Internet Res* 2020;22(11):e23128) doi: [10.2196/23128](https://doi.org/10.2196/23128)

KEYWORDS

COVID-19; ICU; machine learning; death prediction model; factor analysis; SHAP; LIME

Introduction

COVID-19 is a new and severe infectious disease that has spread to 34 provinces and cities in China and over 30 countries worldwide [1,2]. After the entire nation of China fought against COVID-19, by early May 2020, the numbers of patients with COVID-19 had greatly decreased in almost all provinces and cities in China. However, in mid-June, a new outbreak of COVID-19 cases occurred in Beijing, the capital of China. Government efforts have now brought the overall spread of COVID-19 under control. It is clear that COVID-19 is an infectious disease that requires ongoing attention from the medical community, governments, and the public to prevent future outbreaks. As of the end of June 2020, more than 10,000,000 COVID-19 cases had been recorded worldwide. Therefore, an evaluation and early warning system for COVID-19 prognosis is urgently needed, especially for critically ill patients.

COVID-19 cases are classified as mild, moderate, severe, or critical [3]. At present, most studies of COVID-19 have focused on risk factor analysis and mortality prediction for mild and moderate cases, which comprise a large proportion of patients with COVID-19 [4-8]. However, 14% to 20% of cases are severe or even critical [1,9], and the mortality rate of these patients is as high as 50% [10]. Few studies have reported risk factor prediction and mortality analysis for severe and critical patients with COVID-19. COVID-19 predictive models are rapidly entering the academic literature. These include predictive models that are mainly used to identify high-risk groups in the general population [11-13], diagnostic models that are used to detect COVID-19 [14-16], and models used to predict mortality, serious disease progression, etc [17-20]. The most common predictors of the diagnosis and prognosis of COVID-19 are age, body temperature, lymphocyte count, and lung imaging characteristics. The estimated C indices of these predictive models are between 0.65 and 0.99. Although the estimated C indices of some models appear to be ideal, all the models are rated as being at high risk of bias, mainly because of the high risk of model overfitting. Moreover, many of the report descriptions are vague. Most reports do not include a description of the study population or the intended use of the model, and very few evaluations of the calibration of model predictions were made [21].

The theoretical core of machine learning analysis is the data mining algorithm. Various data mining algorithms based on different data types and formats can more scientifically represent the characteristics of the data and can better penetrate the data trends and recognized values [22]. On this basis, one of the most important application areas is predictive analysis, which involves identifying features (in machine learning, “features” refers to individual characteristics of the data) from mechanical learning, establishing models through science, and then running new data through the models to predict future data [23]. In this study, we clarify that the established model is used to predict the prognosis of patients with COVID-19 in the intensive care unit (ICU).

The model must be continuously optimized and evaluated. In terms of evaluation, we checked the accuracy and calibration of the model. Moreover, to improve the interpretability of the black box model, we also used SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) to explain the prediction model; therefore, the prediction model not only predicts prognostic outcomes but also gives a reasonable explanation for the prediction, which can greatly enhance users’ trust of the model.

Methods**Study Design and Data Source**

Vulcan Hill Hospital, located in Wuhan, Hubei Province, is a special hospital that was built by the Chinese government to treat patients with COVID-19. Construction on the hospital started on January 24, 2020, and was completed on February 1; the hospital entered use on February 2, and it officially closed on April 15. During this period, a total of 3063 patients with laboratory-confirmed hospitalized cases of COVID-19 were admitted. For this study, data of 3063 patients with COVID-19 admitted to Vulcan Hill Hospital were extracted from hospital medical records and screened for eligibility. The study extracted 100 relevant variables, such as baseline patient information, clinical diagnosis, vital signs, laboratory test results, medical advice, and nursing care, as candidate variables for predictors [24]. We established a study cohort of 123 critically ill patients admitted to the ICU, and 2940 patients who did not enter the ICU were excluded. Considering the problem of predictors and the timing of the outcome measurement, we used the time the patient entered the ICU to calculate the first test value of all candidate predictors upon entering the ICU. The output of our study is the prognostic outcome of these critically ill patients, and this outcome is based on the ICU death record in the electronic medical record. After further checking the admission records, we included data of 123 critically ill patients admitted to the ICU, including 65 (52.8%) who survived and 58 (47.2%) who died. We randomly used 80% of these data as the training set, and the remaining 20% were used as the validation set. For the training set data, we completed the statistics and proper processing of missing values, the identification and processing of noise data, and the standardization of all predictive variables. The validation set was processed in exactly the same way as the training set. Then, we researched and analyzed the feature selection, model training, model evaluation, and model interpretation.

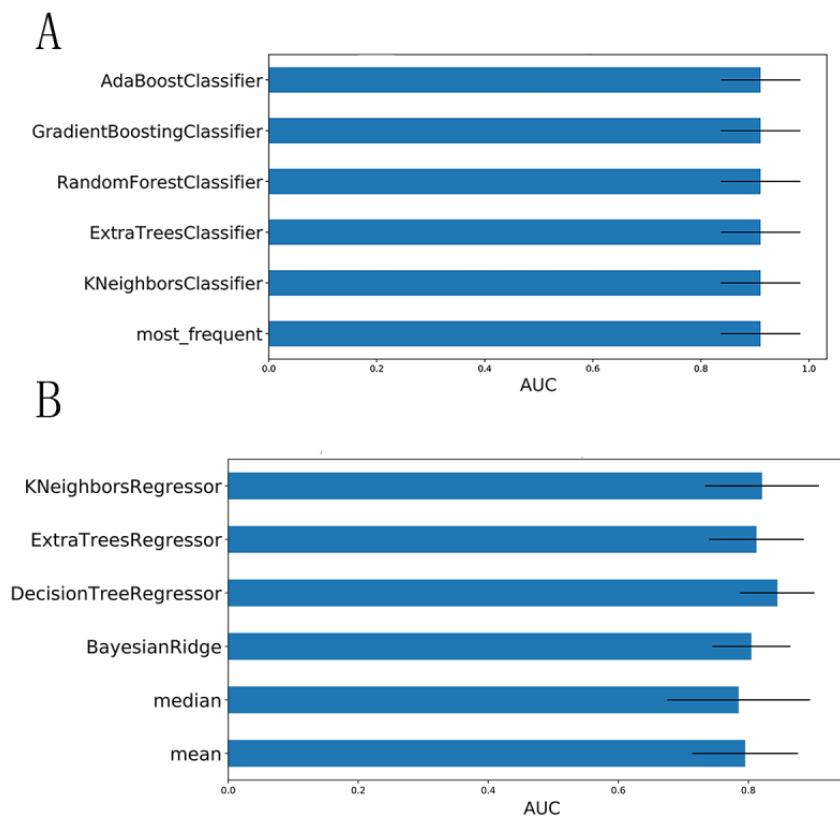
It is worth noting that this study adhered to the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement for reporting, and completion of the model construction and verification was guided by PROBAST (Prediction model Risk Of Bias ASsessment Tool) [25,26]. This study was approved by the ethics committee of Vulcan Hill Hospital; the requirement for informed consent was waived.

Predictor Variables and Data Preprocessing

In this study, a total of 100 candidate predictive features were collected, and the test results were the first measured value after the patient entered the ICU. Among these features, 5 features with a missing ratio greater than 30% were excluded, and the remaining features were filled with missing values using appropriate methods. Because missing data may lead to loss of useful information and even create instability of the model realization, it is more difficult to analyze the model results with missing data; therefore, we carried out a cautious missing value interpolation strategy. We used the Iterative Imputer tool developed by scikit-learn to perform multiple imputations for missing values. The Iterative Imputer uses an algorithm to model each missing value feature as a function of other features. It uses the predicted value of the function as an estimate. In each step, one feature is selected as the output y , and all other features are selected as the input X . Then, a regressor is trained on X and y to predict the missing value of y . The area under the

receiver operating characteristic curve (AUC) values corresponding to each padding method were found to be basically the same. We also applied the K Neighbors Regressor, Decision Tree Regressor, Bayesian Ridge, and Extra Trees Regressor regression algorithms as predictors to complete missing value filling. Moreover, we attempted mean filling and median filling and fed the above six imputation results into the traditional logistic regression model to calculate and compare the areas under the receiving operator characteristic curve (AUC) of their respective prediction results. The results in [Figure 1](#) show that the best filling method is multiple imputation, with Decision Tree Regressor as the regression method; therefore, this method was finally used to fill in the missing data for the continuous features. The missing data are provided in [Multimedia Appendix 1](#). In addition, we drew box plots for continuous features and used IQR criteria to filter and replace outliers. Finally, to obtain more reliable prediction performance, the continuous data were standardized by the z score standardization method.

Figure 1. (A) Comparison of the AUC values obtained after using logistic regression as a carrier and filling in discrete features with multiple missing filling methods. (B) Comparison of the AUC values obtained after continuous feature filling by multiple missing filling methods using logistic regression as a carrier. AUC: area under the receiver operating characteristic curve.



Feature Selection and Statistical Analysis

The categorical variables of the queue data were expressed as n (%). Continuous variables that satisfy normal distribution were expressed as mean (SD); otherwise, medians and quartiles were used. All characteristics were evaluated statistically, and two-sided differences in P values $<.05$ were considered statistically significant. Differences between categorical variables were compared using the chi-square test or Fisher

exact test as needed. The independent sample t test was used to compare continuous variables that satisfy normal distribution, while the Wilcoxon test was used for nonnormally distributed continuous variables. Statistically significant features were selected for further correlation analysis. For redundant features with strong correlation, factor analysis was used to confirm the collinearity of the variables, classify the collinearity as a latent factor, and then calculate the eigenvalues, visualize the gravel map of the eigenvalues, and select the feature root. Values >1

and the first few principal components where the slope decreases were used as principal component factors to eliminate redundant features and to find more efficient, concise, and precise feature combinations, thereby improving the generalization and practical capabilities of the model, as described previously [27]. Note that factor analysis requires the data to be suitable according to the Bartlett sphere test and Kaiser-Meyer-Olkin test.

Derivation and Validation of the Models

A conventional logistic regression method and four popular machine learning classification algorithms, including adaptive boosting (AdaBoost), gradient boosting decision tree (GBDT), eXtreme Gradient Boosting (XGBoost), and CatBoost, were applied in the present study to model the data. The model built by the algorithm uses constant parameter optimization and model evaluation to compare the fitting effects of each model and to select the best model as the risk prediction model. Model optimization is a method that combines grid search and five-fold cross-validation to visualize the AUC values of the model and the standard deviation with the parameters and selects the parameter values corresponding to the best AUC value as the model parameters [28].

A good model explanation must be presented for the black box model. This study is based on the SHAP algorithm, which calculates the marginal contribution of a feature when it is added to the model and then considers whether the factor is different in all factor sequences [29]. The marginal contribution fully explains the influence of all factors included in the model for model prediction and distinguishes the attributes of the factors (risk factors and protective factors).

Finally, validation queue data were used to evaluate the prediction performance of the model and calculate the AUC, threshold, Youden index, 95% CI, SD, and *P* value of the AUC, accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), and negative likelihood ratio (NLR) of the model for the test set. A calibration curve was drawn, the calibration degree of the model was measured, and the degree of consistency between the predicted risk and the actual risk of the model was evaluated. A good calibration degree indicates that the predicted value of the model is closer to the actual probability of the outcome, from the interpretation of the model to the prediction of random samples, as demonstrated previously [30]. The model, code, and parameters are provided in [Multimedia Appendix 2](#).

Interpretation of the Model for Prediction of Random Samples

Advanced machine learning models are usually “black boxes.” When the internal operations of a model are unknown, users do not trust the reliability of the model for making predictions. Although it is known that the accuracy of cross-validation of these models is very high, correlation is still sometimes found between the verification data and the training data due to improper methods, especially when there are few samples. Therefore, cross-validation is no longer the only indicator for evaluating trust. If the rationale by which a model predicts a single sample can be intuitively perceived, users can better trust or distrust single sample prediction. The LIME algorithm was implemented with this concern in mind. This linear model is used to locally approximate a black box model by giving weights to the disturbance input; thus, the observation model gives a basis for interpretation of the sample prediction results [31]. In the present study, we randomly sampled the test set and used the LIME algorithm to fit the predictive behavior of the model to the sample to verify the rationality of the basis of the model for predicting results.

Results

Study Population and Baseline Characteristics

Data from 3063 patients with COVID-19 treated at Vulcan Hill Hospital from February 2 to April 15, 2020, were analyzed retrospectively. A total of 69/3063 deaths occurred (2.3%). The final analytic sample included 123 critically ill patients admitted to the ICU, including 85 critically ill patients (69.1%), 65 surviving patients (52.8%), and 58 patients who died (47.2%). The outcome variable was determined as the prognostic outcome of critically ill patients, and the outcome was based on the death record in the electronic medical record. 100 related variables were established as candidate variables for predictors. [Figure 2](#) shows a flowchart of the overall process of data and feature screening. [Figure 3](#) and [Multimedia Appendix 3](#) list the results comparing all potential risk factors in the study cohort. Overall, the mean age of the patients in the cohort was 69.8 years (SD 11.1), and 79/123 patients (64.2%) were male. Data analysis revealed significant differences between patients in 7 discrete factors, namely ventilator use, critical illness, vasoactive drugs, carbapenem use, antibiotic resistance, anti-gram-positive cocci, and hemodiafiltration; significant differences were found in 46 continuous factors.

Figure 2. Flowchart of the data and feature selection.

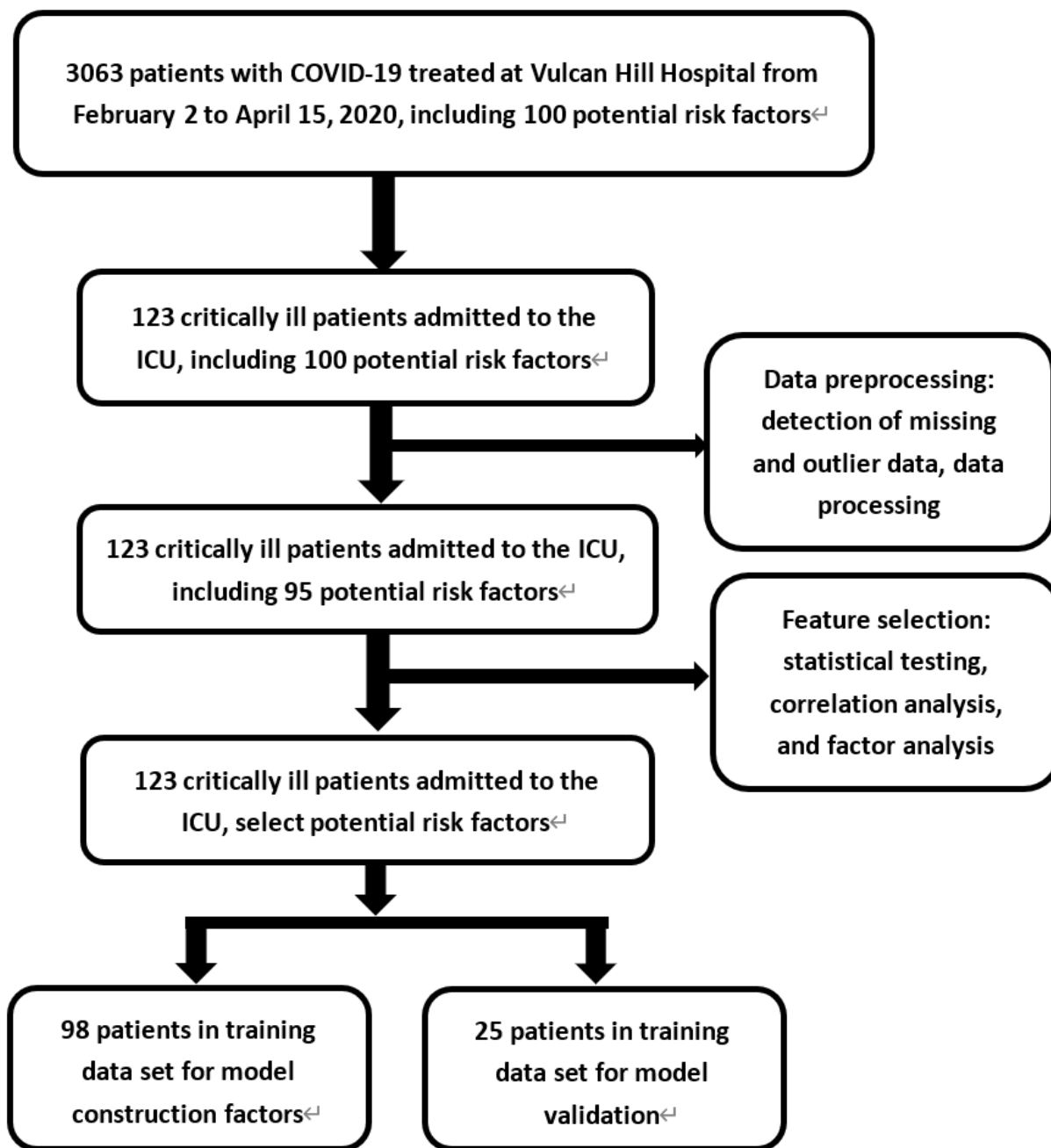
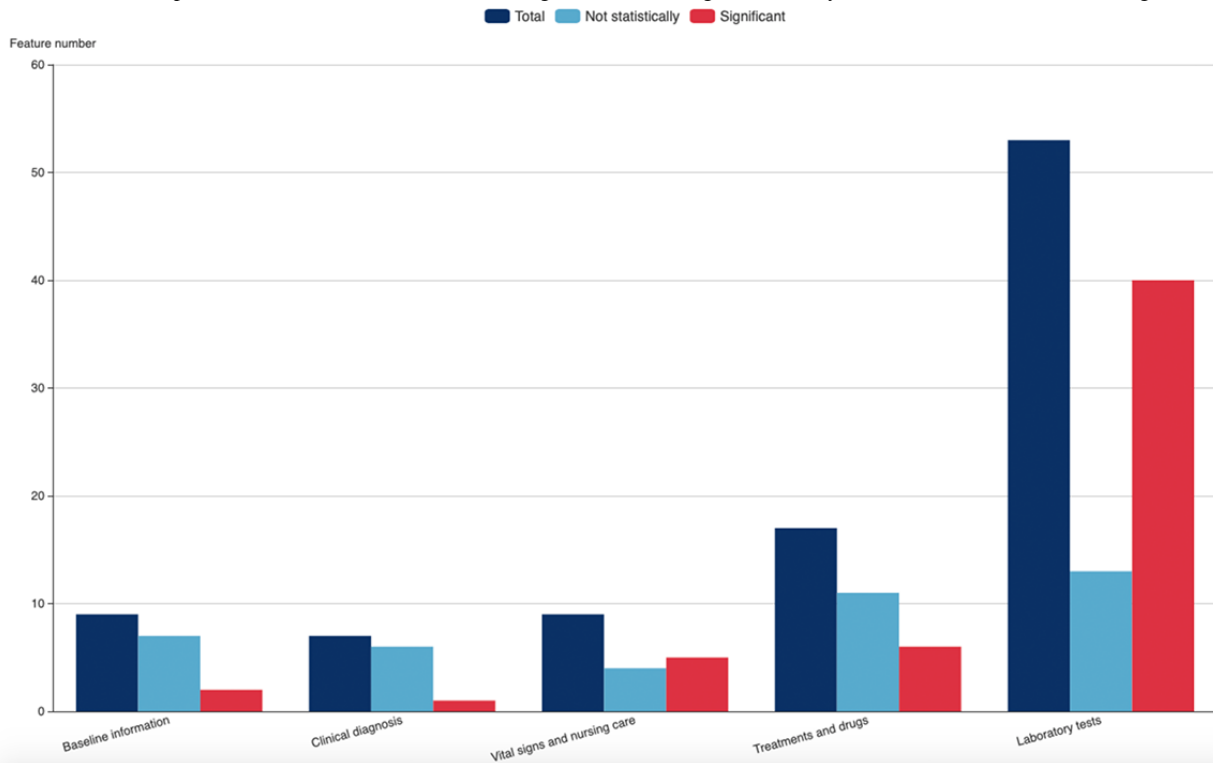


Figure 3. Statistical analysis of the screening results of all potential risk factors in the study cohort. The figure shows the screening of candidate variables in the five dimensions of patient baseline information, clinical diagnosis and vital signs, laboratory tests, medical advice, and nursing care.



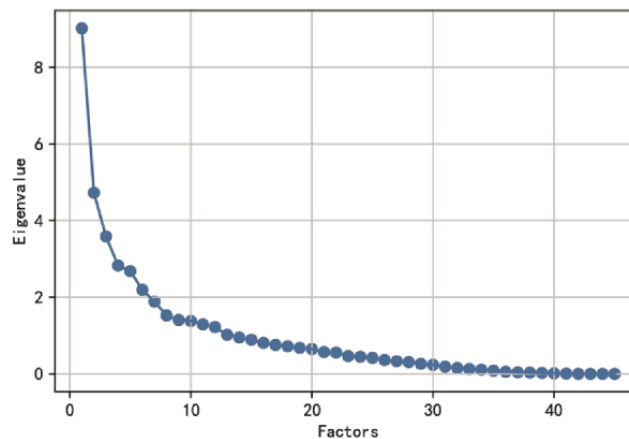
Predictor Selection

By selecting the abovementioned statistically significant factors for correlation analysis, the correlation coefficient matrix heat map (Multimedia Appendix 4) of the features shows that the top five features that were negatively correlated with the outcomes are prothrombin time percentage activity, blood oxygen saturation, lymphocyte percentage (LYM%), albumin level (ALB), and percentage of basophils (BASO%); the top five characteristics that were positively correlated with outcomes are lactate dehydrogenase, alpha-hydroxybutyrate dehydrogenase, C-reactive protein, neutrophil percentage (NEUT%), and original thrombin time. In addition, strong correlations were found between many features. For example, the correlation coefficient between the prothrombin time (PT)

and international standardized ratio reached 0.999; therefore, it was necessary to reduce redundant features.

Factor analysis and visualization of the characteristic root gravel map and load matrix (Figure 4 and Multimedia Appendix 5) revealed that the eight principal component factors were the most predictive; for example, the correlation between the characteristic prothrombin time and the second main factor reached 0.97. Considering the convenience and practicability of using the prediction model, clinical experience and actual comparisons were combined to finally select eight features to represent the eight principal component factors, namely LYM%, PT, lactate dehydrogenase (LDH), total bilirubin (T-Bil), eosinophil percentage (EOS%), creatinine (Cr), NEUT%, and ALB. The Kaiser-Meyer-Olkin test gave a value of 0.5714 and Bartlett's test of sphericity showed a significance level of $P < .001$, indicating that the factor analysis is effective.

Figure 4. Distribution diagram of the correlations between the feature value and the number of features; when the feature value is >1 and the slope change becomes slow, the number of features is 8.



Machine Learning Algorithm Comparison and Best Model

Comparing the AUCs of the logistic regression and four machine learning algorithms for 5-fold cross-validation on the training set (Figure 5), it can be found that the AUC values of each algorithm are similar; however, the AUC value of the XGBoost algorithm is higher. The XGBoost algorithm reflects a good learning curve on the training set, effectively preventing overfitting. In terms of prediction performance, the results of the logistic regression and four machine learning algorithms on

the test set show AUCs of 0.92 for XGBoost, 0.9133 for CatBoost, 0.9133 for AdaBoost, 0.85 for GBDT, and 0.84 for LR; the best prediction performance was observed with XGBoost. In addition, the AUC, threshold, Youden index, 95% CI, SD, and P value of the AUC, accuracy, sensitivity, specificity, PPV, NPV, PLR, and NLR values of each model in the test data set are listed in Table 1. In summary, the results of the test data set show that the XGBoost model demonstrates the best performance based on eight salient features. The Youden index value of this model is 0.6667.

Figure 5. ROC curves showing the fitting performance (A) and prediction performance (B) of the LR, CatBoost, GBDT, XGBoost, and AdaBoost prediction models based on the eight important features in the training data set and the test data set. (C) The learning curve on the training set showing the learning process of the XGBoost model. The red line is the fitting effect of the model on the overall training set, and the green line is the fitting effect of the model on the training set with 5-fold cross-validation. The two curves finally merge near 0.85, indicating that the model is well fitted for training. AdaBoost: adaptive boosting; GBDT: gradient boosting decision tree; LR: logistic regression; ROC: receiver operating characteristic; XGBoost: eXtreme Gradient Boosting.

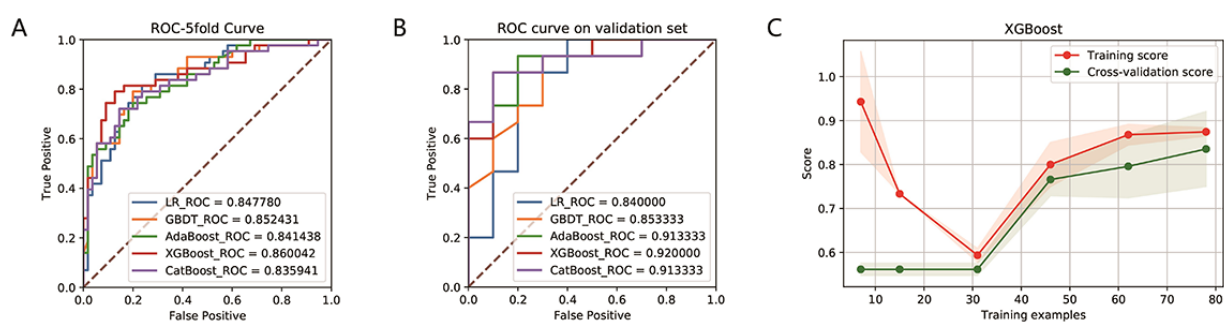


Table 1. Summary of prediction results of multiple models on the test set.

Value	Models				
	Logistic regression	AdaBoost ^a	GBDT ^b	XGBoost ^c	CatBoost
AUC ^d	0.84	0.9133	0.85	0.92	0.9133
Threshold	0.3962	0.4283	0.4583	0.4478	0.5063
Youden index	0.6667	0.7333	0.6333	0.7667	0.7667
95% CI of the AUC	0.6556-1.0	0.8024-1.0	0.6997-1.0	0.8142-1.0	0.7997-1.0
SD of the AUC	0.094	0.0566	0.0784	0.054	0.058
P value of the AUC	.003	<.001	.002	<.001	<.001
Accuracy	0.76	0.76	0.76	0.84	0.84
Specificity	0.8	0.9	0.8	0.9	0.8
Sensitivity	0.7333	0.6667	0.7333	0.8	0.8667
Positive predictive value	0.8462	0.9091	0.8462	0.9231	0.8667
Negative predictive value	0.6667	0.6429	0.6667	0.75	0.8
Positive likelihood ratio	3.6667	6.6667	3.6667	8	4.3333
Negative likelihood ratio	0.3333	0.3704	0.3333	0.2222	0.1667

^aAdaBoost: adaptive boosting.

^bGBDT: gradient boosting decision tree.

^cXGBoost: eXtreme Gradient Boosting.

^dAUC: area under the receiver operating characteristic curve.

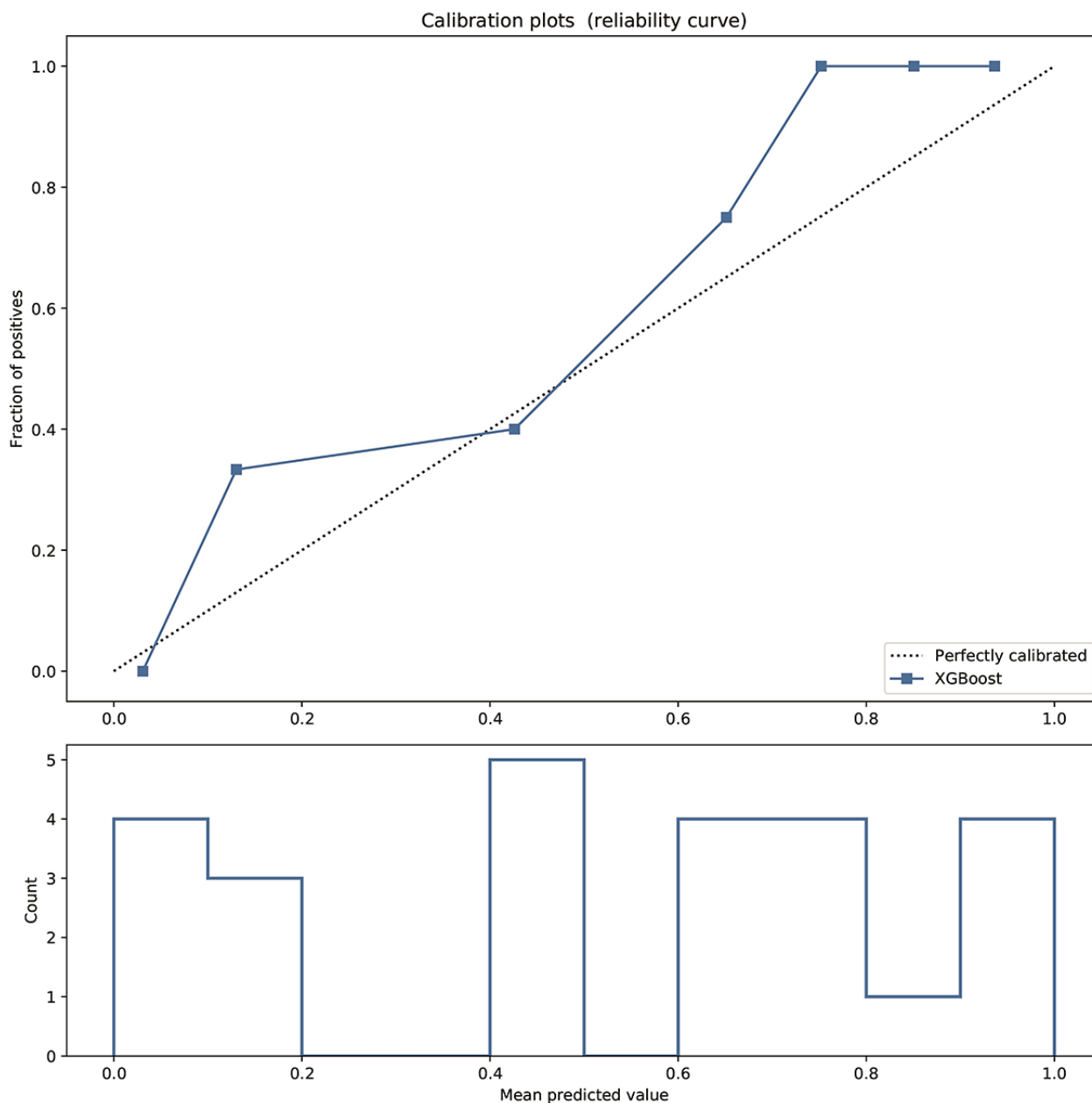
Model Validation and Predictor Parameters

The prediction behavior of XGBoost in the test set was visualized. The calibration curve (Figure 6) shows that the

predicted risk of the XGBoost model is in good agreement with the actual risk. The predicted value of the model is close to the actual probability of the outcome. The details of the optimal

model parameters constructed by the XGBoost algorithm can be viewed in [Multimedia Appendix 2](#).

Figure 6. Calibration curve reflecting the degree of consistency between the predicted risk and the actual risk of the XGBoost model. The predicted curve of the model fits well with the diagonal, indicating that the predicted value of the model is basically close to the actual probability of the outcome.

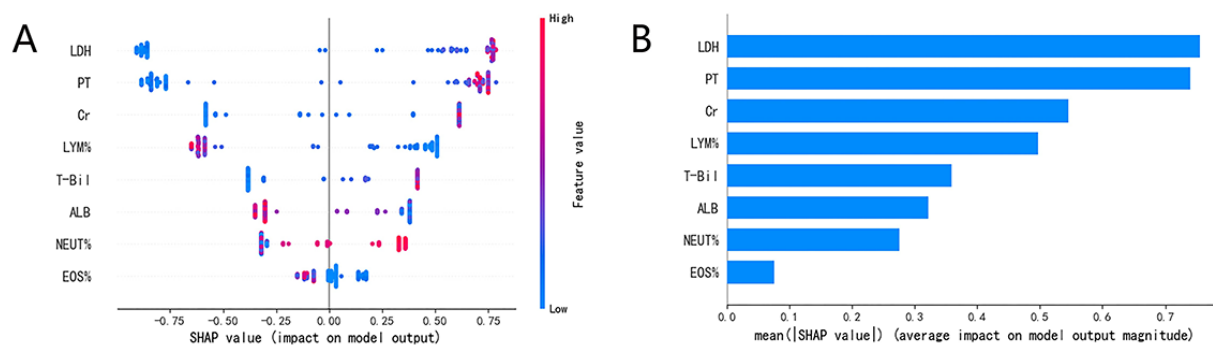


Interpretation and Evaluation of the Machine Learning Model

Based on the SHAP algorithm, the feature ranking interpretation of the XGBoost model (Figure 7) shows that LDH, PT, Cr, LYM%, NEUT%, EOS%, T-Bil, and ALB were the

characteristics of the XGBoost model with the greatest impact in predicting outcomes. Overall, the characteristics of LDH, PT, Cr, T-Bil, and NEUT% correlated positively with the outcomes and are risk factors; meanwhile, LYM%, EOS%, and ALB correlated negatively with the outcomes and are protective factors.

Figure 7. The XGBoost model based on the SHAP algorithm. (A) The attributes of the features in the black box model. Each line represents a feature, and the abscissa is the SHAP value, which represents the degree of influence on the outcome. Each dot represents a sample. The redder the color, the greater the value of the feature, and the bluer the color, the lower the value. (B) Ranking of feature importance indicated by SHAP. ALB: albumin level; Cr: creatinine; EOS%: eosinophil percentage; LDH: lactate dehydrogenase; LYM%: lymphocyte percentage; NEUT%: neutrophil percentage; PT: prothrombin time; SHAP: SHapley Additive exPlanations; T-Bil: total bilirubin.



Interpretation of sample prediction results requires random drawing of samples to make model predictions and observe the model through the LIME algorithm. The four prediction scenarios are shown in Figure 8A. In addition, the prediction results of the XGBoost model on all samples on the training set and the test set were counted, and the distribution of the four cases of the aggregated prediction results was visualized. Figure 8B shows that the XGBoost model has inappropriate prediction

behavior in that its judgment of the false positive prediction results found through LIME is inaccurate; however, this situation is very rare, which indicates that the performance of the XGBoost prediction model is stable and reliable and that the interpretation of random sample prediction is basically reasonable. This is sufficient to confirm the practicability of the XGBoost model and will help increase physicians' trust in the prediction model and help them make good auxiliary decisions.

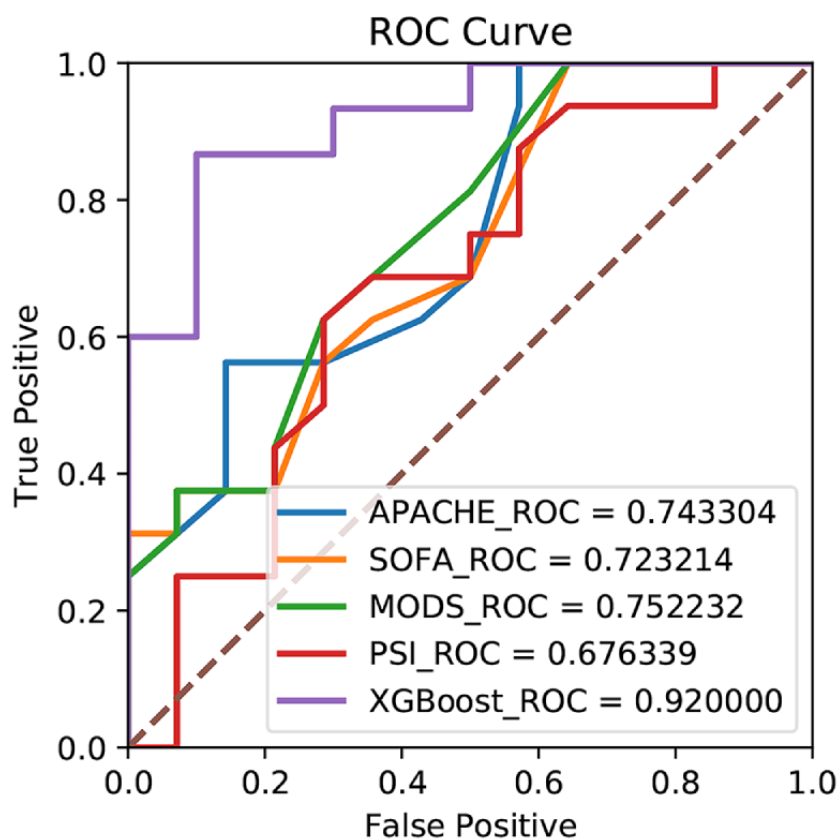
Figure 8. Interpretation of sample prediction results by randomly drawing samples to make model predictions and observing the model through the Local Interpretable Model-Agnostic Explanations (LIME) algorithm. (A) The four different prediction behaviors of the model (true negative, true positive, false negative, and false positive); (B) the ratios of the four prediction behaviors of the model on the training set and the test set. A:LB: albumin level; CR: creatinine; EOS%: eosinophil percentage; LDH: lactate dehydrogenase; LYM%: lymphocyte percentage; NEUT%: neutrophil percentage; PT: prothrombin time; T-Bil: total bilirubin.



Additionally, comparison of the results of our machine learning model and the Acute Physiologic Assessment and Chronic Health Evaluation II (APACHE II), Sequential Organ Failure Assessment (SOFA), Multiple Organ Dysfunction Score

(MODS), and Pneumonia Severity Index (PSI) scores indicated that the AUC of the XGBoost model was higher than those of the other four scores (Figure 9).

Figure 9. Comparison between the risk prediction model of the present study and the ROC curves of various critical scores. APACHE: Acute Physiologic Assessment and Chronic Health Evaluation; MODS: Multiple Organ Dysfunction Score; PSI: Pneumonia Severity Index; ROC: receiver operating characteristic; SOFA: Sequential Organ Failure Assessment; XGBoost: eXtreme Gradient Boosting.



Discussion

Principal Findings

The results of the present study show that the XGBoost method is a more reliable and more accurate method for predicting outcomes for critically ill patients with COVID-19 in the ICU than conventional logistic regression and scoring. Especially, the eigenvalues were reduced using the XGBoost model from 100 parameters to 8. Correlation analysis and characteristic analysis showed that the LDH, PT, Cr, T-Bil, LYM%, ALB, NEUT%, and EOS% indicators had strong correlations with the prognosis of severe and critical patients with COVID-19 in the ICU. Physicians should be wary of poor prognosis when encountering such patients. After full verification by SHAP, LIME, etc, the model was found to be accurate and stable. A web-based calculator based on the risk model is available on the internet [32].

According to the XGBoost algorithm model used in our study, LDH, PT, Cr, T-Bil, and NEUT% correlated positively with patients' outcomes, indicating that these values are risk factors; meanwhile, LYM%, ALB, and EOS% correlated negatively with patients' outcomes, indicating that these values are

protective factors. In addition to identifying risk and protective factors, the results suggested that the XGBoost algorithm model achieves a good prediction effect, with an AUC of 0.92, sensitivity of 0.8, and specificity of 0.9. Based on the same data, logistic regression analysis showed an AUC of 0.84, with a sensitivity of 0.7 and specificity of 0.83. These results indicate that the predictive effect of machine learning is more accurate and sensitive than that of regression analysis. In contrast to our study, other researchers tended to apply Cox regression and logistic regression to analyze risk factors. Wang et al [19] found that the risk factors for in-hospital mortality from COVID-19 were lymphopenia and LDH, as analyzed by multivariable Cox proportional hazard regression models. Chen and colleagues [20] studied 1859 patients with confirmed COVID-19 from seven centers in Wuhan, China, of whom 1651 recovered and 208 died. Multivariable Cox regression analyses indicated that increased hazards of in-hospital death were associated with one indicator, log₁₀ serum creatinine (sCr) per μ mol/L increase [33]. In another study that analyzed 167 confirmed patients with severe COVID-19, the LDH concentration was higher and the albumin concentration was lower in these patients, with significant differences [34]. Recently, Liang et al [22] used Least Absolute Shrinkage and Selection Operator (LASSO) and

logistic regression to construct a predictive risk score (COVID-GRAM), in which the AUC in the development cohort was 0.88 (95% CI 0.85-0.91) and the AUC in the validation cohort was 0.88 (95% CI 0.84-0.93) for predicting patients' risk of developing critical illness. The results in that study coincide with our results to some extent, and the machine learning algorithm can identify potential indicators better than the conventional algorithm. Finally, Abdulaal et al [20] used a neural network model to predict the prognosis of patients with COVID-19, with an AUC of 90.12%. This study only used eight indicators and achieved good predictive value; thus, it is more convenient and efficient.

Elevated LDH was identified as a significant risk predictive marker of COVID-19. Li et al [23] revealed that relatively high levels of LDH play a crucial role in predicting mortality of patients with COVID-19 when using an interpretable mortality prediction model. LDH is an enzyme that is involved in energy production through the conversion of lactate to pyruvate; it is present in almost all body cell types (n=156), with the highest levels in the heart, liver, lungs, muscles, kidneys, and blood cells. LDH is released from cells upon damage to their cytoplasmic membrane, and it is not only a metabolic marker but also an immune surveillance prognostic biomarker [35]. LDH increases the production of lactate, which leads to enhancement of immune-suppressive cells and inhibition of cytolytic cells. These changes weaken the immune response mounted against viral infection, which results in more severe disease in patients with COVID-19 who have elevated LDH [36]. PT is another typical indicator associated with patient prognosis. Through pathological examination of patients who died of COVID-19, researchers found that the virus can lead to disorders of the coagulation system, resulting in a hypercoagulable state and microthrombosis [37]. Moreover, viral infections may induce even more severe complications, such as acute respiratory distress syndrome and multi-organ dysfunction syndrome, which are two conditions frequently associated with hypercoagulation and disseminated intravascular coagulation [38]. These processes and conditions help to explain why PT was prolonged in patients with severe and critical COVID-19. Approximately 14.4% of patients with COVID-19 have elevated sCr levels, and kidney disease has been associated with in-hospital death of patients with COVID-19 [39]. SARS-CoV2 has been suggested to modulate the renin-angiotensin-aldosterone system (RAAS). Evidence of activation of the RAAS in patients with COVID-19 who have acute kidney injury, leading to increased sCr, has been reported [40]. Several studies have also reported that liver damage occurred in severe cases of COVID-19 infection at rates ranging from 58% to 78% [41,42]. COVID-19 uses angiotensin converting enzyme 2 (ACE2) as the binding site to enter host cells in the lungs, kidneys, and heart. A previous study [43] showed that both liver and bile duct cells express ACE2; this

may result in elevated T-Bil levels, accompanied by slightly decreased ALB levels. Hematologic and immunologic impairment showed significantly different profiles between survival and mortality of patients with COVID-19 with different disease severities. The results of our study suggest that increased NEUT% and decreased LYM% are risk factors for patient prognosis. Interestingly, a decrease in EOS% was also a risk factor, and we were surprised to find that the results of two studies [44,45] were consistent with our results. In those studies, it was found that eosinophils decreased at the early stage and were associated with disease severity and clinical outcomes. Impaired immune cell function leads to low lymphocyte levels and immune system dysfunction, causing patients with severe COVID-19 to be more sensitive to bacterial infection [46]. The decline in eosinophils may be due to the patients' response to the stress of acute SARS-CoV-2 infection. However, whether COVID-19 has a direct effect on eosinophils remains unknown. In one study, it was found that the *Clostridioides difficile* transferase toxin induces pathogenic host inflammation via a toll-like receptor 2-dependent pathway, resulting in suppression of the protective host eosinophilic response [47]. Additionally, eosinophils can be reduced after an innate immune challenge [48].

Limitations

This study has several limitations. First, data from only one center were used, and the sample size was small, which may indicate bias. However, Vulcan Hill Hospital is a large medical center that focused on the treatment of COVID-19. The patients are representative of all patients with COVID-19, providing a reliable basis for the treatment of critical patients. Second, the treatment of patients in the ICU is not necessarily the initial treatment because the patients were transferred from different hospitals and different medical treatment units; this may have affected the baseline characteristics of the patients. In the future, based on what we learned in this study, we will attempt to correct defects of the model and the machine learning approach. Also, we will collect more data to conduct external tests on the prediction model and further improve the generalized prediction ability of the model for multicenter data.

Conclusions

Machine learning has a good predictive effect on the mortality of critically ill patients with COVID-19 in the ICU. The XGBoost model has higher diagnostic performance than conventional statistical methods and can be used to select and simplify the core indicators for mortality prediction, such as LDH, PT, Cr, T-Bil, LYM%, ALB, and the white blood cell parameters NEUT% and BASO%. Machine learning may be a valuable prognostic indicator for early warning of critically ill patients; this warning plays a significant role in the allocation of medical resources, triage of patients, formulation of treatment decisions, and evaluation of progressive COVID-19.

Acknowledgments

The China National Key Research Program (Grant number 2018ZX09201013) and The Emergency Scientific Research Project for COVID-19 (20yjky007) supported this study. Dr Pan Pan, Yichao Li, Yongjiu Xiao, Bingchao Han, Longxiang Su contributed equally to this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summary table of the number and proportion of missing values.

[\[DOC File , 68 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Mathematical algorithm.

[\[DOCX File , 13 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Comparative analysis of all potential risk factors in the study cohort.

[\[DOC File , 200 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Correlation coefficient matrix heat maps.

[\[DOCX File , 551 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

Load matrix diagram showing the correlations between the characteristics and the main factors.

[\[PNG File , 482 KB-Multimedia Appendix 5\]](#)

References

1. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 2020 Apr 30;382(18):1708-1720 [[FREE Full text](#)] [doi: [10.1056/NEJMoa2002032](https://doi.org/10.1056/NEJMoa2002032)] [Medline: [32109013](https://pubmed.ncbi.nlm.nih.gov/32109013/)]
2. Zhen-Dong Y, Gao-Jun Z, Run-Ming J, Zhi-Sheng L, Zong-Qi D, Xiong X, et al. Clinical and transmission dynamics characteristics of 406 children with coronavirus disease 2019 in China: A review. *J Infect* 2020 Aug;81(2):e11-e15 [[FREE Full text](#)] [doi: [10.1016/j.jinf.2020.04.030](https://doi.org/10.1016/j.jinf.2020.04.030)] [Medline: [32360500](https://pubmed.ncbi.nlm.nih.gov/32360500/)]
3. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 2020 Feb 24;1239-1242. [doi: [10.1001/jama.2020.2648](https://doi.org/10.1001/jama.2020.2648)] [Medline: [32091533](https://pubmed.ncbi.nlm.nih.gov/32091533/)]
4. Wander PL, Orlov M, Merel SE, Enquobahrie DA. Risk factors for severe COVID-19 illness in healthcare workers: Too many unknowns. *Infect Control Hosp Epidemiol* 2020 Apr 27;1-2 [[FREE Full text](#)] [doi: [10.1017/ice.2020.178](https://doi.org/10.1017/ice.2020.178)] [Medline: [32336303](https://pubmed.ncbi.nlm.nih.gov/32336303/)]
5. Cen Y, Chen X, Shen Y, Zhang X, Lei Y, Xu C, et al. Risk factors for disease progression in patients with mild to moderate coronavirus disease 2019-a multi-centre observational study. *Clin Microbiol Infect* 2020 Jun 09;1242-1247 [[FREE Full text](#)] [doi: [10.1016/j.cmi.2020.05.041](https://doi.org/10.1016/j.cmi.2020.05.041)] [Medline: [32526275](https://pubmed.ncbi.nlm.nih.gov/32526275/)]
6. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020 Mar 28;395(10229):1054-1062 [[FREE Full text](#)] [doi: [10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)] [Medline: [32171076](https://pubmed.ncbi.nlm.nih.gov/32171076/)]
7. Liu D, Wang Y, Wang J, Liu J, Yue Y, Liu W, et al. Characteristics and Outcomes of a Sample of Patients With COVID-19 Identified Through Social Media in Wuhan, China: Observational Study. *J Med Internet Res* 2020 Aug 13;22(8):e20108 [[FREE Full text](#)] [doi: [10.2196/20108](https://doi.org/10.2196/20108)] [Medline: [32716901](https://pubmed.ncbi.nlm.nih.gov/32716901/)]
8. Li J, Chen Z, Nie Y, Ma Y, Guo Q, Dai X. Identification of Symptoms Prognostic of COVID-19 Severity: Multivariate Data Analysis of a Case Series in Henan Province. *J Med Internet Res* 2020 Jun 30;22(6):e19636 [[FREE Full text](#)] [doi: [10.2196/19636](https://doi.org/10.2196/19636)] [Medline: [32544071](https://pubmed.ncbi.nlm.nih.gov/32544071/)]
9. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med* 2020 Apr;26(4):506-510 [[FREE Full text](#)] [doi: [10.1038/s41591-020-0822-7](https://doi.org/10.1038/s41591-020-0822-7)] [Medline: [32284616](https://pubmed.ncbi.nlm.nih.gov/32284616/)]
10. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, the Northwell COVID-19 Research Consortium, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* 2020 May 26;323(20):2052-2059 [[FREE Full text](#)] [doi: [10.1001/jama.2020.6775](https://doi.org/10.1001/jama.2020.6775)] [Medline: [32320003](https://pubmed.ncbi.nlm.nih.gov/32320003/)]

11. Lu J, Hu S, Fan R, Liu Z, Yin X, Wang Q, et al. ACP Risk Grade: A Simple Mortality Index for Patients with Confirmed or Suspected Severe Acute Respiratory Syndrome Coronavirus 2 Disease (COVID-19) During the Early Stage of Outbreak in Wuhan, China. SSRN Journal Preprint posted online on February 28, 2020. [doi: [10.2139/ssrn.3543603](https://doi.org/10.2139/ssrn.3543603)]
12. Barda N, Riesel D, Akriv A, Levy J, Finkel U, Yona G, et al. Developing a COVID-19 mortality risk prediction model when individual-level data are not available. *Nat Commun* 2020 Sep 07;11(1):4439 [FREE Full text] [doi: [10.1038/s41467-020-18297-9](https://doi.org/10.1038/s41467-020-18297-9)] [Medline: [32895375](https://pubmed.ncbi.nlm.nih.gov/32895375/)]
13. Ewan CR, Bean D, Stammers M, Wang W, Zhang H, Searle T, et al. Evaluation and Improvement of the National Early Warning Score (NEWS2) for COVID-19: a multi-hospital study. medRxiv Preprint posted online on September 30, 2020. [doi: [10.1101/2020.04.24.20078006](https://doi.org/10.1101/2020.04.24.20078006)]
14. Cheng JW, Cao Y, Xu Z, Tan Z, Zhang X, Deng L, et al. Development and Evaluation of an AI System for COVID-19 Diagnosis. medRxiv Preprint posted online on June 02, 2020. [doi: [10.1101/2020.03.20.20039834](https://doi.org/10.1101/2020.03.20.20039834)]
15. Zhiyi WJ, Li Z, Hou R, Zhou L, Ye H, Chen Y, et al. Development and Validation of a Diagnostic Nomogram to Predict COVID-19 Pneumonia. medRxiv Preprint posted online on April 06, 2020. [doi: [10.1101/2020.04.03.20052068](https://doi.org/10.1101/2020.04.03.20052068)]
16. Imran A, Posokhova I, Qureshi HN, Masood U, Riaz MS, Ali K, et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Inform Med Unlocked* 2020;20:100378 [FREE Full text] [doi: [10.1016/j.imu.2020.100378](https://doi.org/10.1016/j.imu.2020.100378)] [Medline: [32839734](https://pubmed.ncbi.nlm.nih.gov/32839734/)]
17. Xie J, Hungerford D, Chen H, Abrams ST, Li S, Wang G, et al. Development and External Validation of a Prognostic Multivariable Model on Admission for Hospitalized Patients with COVID-19. SSRN Journal Preprint posted online on April 06, 2020. [doi: [10.2139/ssrn.3562456](https://doi.org/10.2139/ssrn.3562456)]
18. Bai X, Fang C, Zhou Y, Bai S, Liu Z, Xia L, et al. Predicting COVID-19 Malignant Progression with AI Techniques. SSRN Journal Preprint posted online on March 31, 2020. [doi: [10.2139/ssrn.3557984](https://doi.org/10.2139/ssrn.3557984)]
19. Ying ZZ, Guo Y, Geng S, Gao S, Ye S, Hu Y, et al. A New Predictor of Disease Severity in Patients with COVID-19 in Wuhan, China. Research Square Preprint posted online on May 21, 2020. [doi: [10.21203/rs.3.rs-29566/v1](https://doi.org/10.21203/rs.3.rs-29566/v1)]
20. Abdulaal A, Patel A, Charani E, Denny S, Mughal N, Moore L. Prognostic Modeling of COVID-19 Using Artificial Intelligence in the United Kingdom: Model Development and Validation. *J Med Internet Res* 2020 Aug 25;22(8):e20259 [FREE Full text] [doi: [10.2196/20259](https://doi.org/10.2196/20259)] [Medline: [32735549](https://pubmed.ncbi.nlm.nih.gov/32735549/)]
21. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020 Apr 07;369:m1328. [doi: [10.1136/bmj.m1328](https://doi.org/10.1136/bmj.m1328)] [Medline: [32265220](https://pubmed.ncbi.nlm.nih.gov/32265220/)]
22. Zampieri FG, Salluh JIF, Azevedo LCP, Kahn JM, Damiani LP, Borges LP, ORCHESTRA Study Investigators. ICU staffing feature phenotypes and their relationship with patients' outcomes: an unsupervised machine learning analysis. *Intensive Care Med* 2019 Nov;45(11):1599-1607. [doi: [10.1007/s00134-019-05790-z](https://doi.org/10.1007/s00134-019-05790-z)] [Medline: [31595349](https://pubmed.ncbi.nlm.nih.gov/31595349/)]
23. Zack JE, Garrison T, Trovillion E, Clinkscale D, Coopersmith CM, Fraser VJ, et al. Effect of an education program aimed at reducing the occurrence of ventilator-associated pneumonia. *Crit Care Med* 2002 Nov;30(11):2407-2412. [doi: [10.1097/00003246-200211000-00001](https://doi.org/10.1097/00003246-200211000-00001)] [Medline: [12441746](https://pubmed.ncbi.nlm.nih.gov/12441746/)]
24. Lijiao ZJ, Liao M, Hua R, Huang P, Zhang M, Zhang Y, et al. Risk assessment of progression to severe conditions for patients with COVID-19 pneumonia: a single-center retrospective study. medRxiv Preprint posted online on March 30, 2020. [doi: [10.1101/2020.03.25.20043166](https://doi.org/10.1101/2020.03.25.20043166)]
25. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015 Jan 06;162(1):W1-73. [doi: [10.7326/M14-0698](https://doi.org/10.7326/M14-0698)] [Medline: [25560730](https://pubmed.ncbi.nlm.nih.gov/25560730/)]
26. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med* 2019 Jan 01;170(1):W1-W33. [doi: [10.7326/M18-1377](https://doi.org/10.7326/M18-1377)] [Medline: [30596876](https://pubmed.ncbi.nlm.nih.gov/30596876/)]
27. de Oliveira Kaizer UA, Alexandre NMC, Rodrigues RCM, Cornélio ME, de Melo Lima MH, São-João TM. Measurement properties and factor analysis of the Diabetic Foot Ulcer Scale-short form (DFS-SF). *Int Wound J* 2020 Jun;17(3):670-682. [doi: [10.1111/iwj.13310](https://doi.org/10.1111/iwj.13310)] [Medline: [32065735](https://pubmed.ncbi.nlm.nih.gov/32065735/)]
28. Jung Y. Multiple predicting K-fold cross-validation for model selection. *J Nonparametr Stat* 2017 Nov 21;30(1):197-215. [doi: [10.1080/10485252.2017.1404598](https://doi.org/10.1080/10485252.2017.1404598)]
29. Rodríguez-Pérez R, Bajorath J. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *J Comput Aided Mol Des* 2020 Oct 02;34(10):1013-1026 [FREE Full text] [doi: [10.1007/s10822-020-00314-0](https://doi.org/10.1007/s10822-020-00314-0)] [Medline: [32361862](https://pubmed.ncbi.nlm.nih.gov/32361862/)]
30. Meloni A, Maggio A, Positano V, Leto F, Angelini A, Putti MC, et al. CMR for myocardial iron overload quantification: calibration curve from the MIOT Network. *Eur Radiol* 2020 Jun;30(6):3217-3225. [doi: [10.1007/s00330-020-06668-1](https://doi.org/10.1007/s00330-020-06668-1)] [Medline: [32052169](https://pubmed.ncbi.nlm.nih.gov/32052169/)]
31. Palatnik de Sousa I, Maria Bernardes Rebuzzi Vellasco M, Costa da Silva E. Local Interpretable Model-Agnostic Explanations for Classification of Lymph Node Metastases. *Sensors (Basel)* 2019 Jul 05;19(13):2969 [FREE Full text] [doi: [10.3390/s19132969](https://doi.org/10.3390/s19132969)] [Medline: [31284419](https://pubmed.ncbi.nlm.nih.gov/31284419/)]

32. Calculation tool for predicting the death probability of COVID-19 patients when entering the ICU. URL: <http://114.251.235.51:1226/index> [accessed 2020-10-27]
33. Chen L, Yu J, He W, Chen L, Yuan G, Dong F, et al. Risk factors for death in 1859 subjects with COVID-19. *Leukemia* 2020 Aug;34(8):2173-2183 [FREE Full text] [doi: [10.1038/s41375-020-0911-0](https://doi.org/10.1038/s41375-020-0911-0)] [Medline: [32546725](https://pubmed.ncbi.nlm.nih.gov/32546725/)]
34. Wei Y, Wang R, Zhang D, Tu Y, Chen C, Ji S, et al. Risk factors for severe COVID-19: Evidence from 167 hospitalized patients in Anhui, China. *J Infect* 2020 Jul;81(1):e89-e92 [FREE Full text] [doi: [10.1016/j.jinf.2020.04.010](https://doi.org/10.1016/j.jinf.2020.04.010)] [Medline: [32305487](https://pubmed.ncbi.nlm.nih.gov/32305487/)]
35. Poggiali E, Zaino D, Immovilli P, Rovero L, Losi G, Dacrema A, et al. Lactate dehydrogenase and C-reactive protein as predictors of respiratory failure in CoVID-19 patients. *Clin Chim Acta* 2020 Oct;509:135-138 [FREE Full text] [doi: [10.1016/j.cca.2020.06.012](https://doi.org/10.1016/j.cca.2020.06.012)] [Medline: [32531257](https://pubmed.ncbi.nlm.nih.gov/32531257/)]
36. Ding J, Karp JE, Emadi A. Elevated lactate dehydrogenase (LDH) can be a marker of immune suppression in cancer: Interplay between hematologic and solid neoplastic clones and their microenvironments. *Cancer Biomark* 2017 Jul 04;19(4):353-363. [doi: [10.3233/CBM-160336](https://doi.org/10.3233/CBM-160336)] [Medline: [28582845](https://pubmed.ncbi.nlm.nih.gov/28582845/)]
37. Zhang Y, Xiao M, Zhang S, Xia P, Cao W, Jiang W, et al. Coagulopathy and Antiphospholipid Antibodies in Patients with Covid-19. *N Engl J Med* 2020 Apr 23;382(17):e38 [FREE Full text] [doi: [10.1056/NEJMc2007575](https://doi.org/10.1056/NEJMc2007575)] [Medline: [32268022](https://pubmed.ncbi.nlm.nih.gov/32268022/)]
38. Chang JC. Acute Respiratory Distress Syndrome as an Organ Phenotype of Vascular Microthrombotic Disease: Based on Hemostatic Theory and Endothelial Molecular Pathogenesis. *Clin Appl Thromb Hemost* 2019;25:1076029619887437. [doi: [10.1177/1076029619887437](https://doi.org/10.1177/1076029619887437)] [Medline: [31775524](https://pubmed.ncbi.nlm.nih.gov/31775524/)]
39. Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int* 2020 May;97(5):829-838 [FREE Full text] [doi: [10.1016/j.kint.2020.03.005](https://doi.org/10.1016/j.kint.2020.03.005)] [Medline: [32247631](https://pubmed.ncbi.nlm.nih.gov/32247631/)]
40. Dudoignon E, Moreno N, Deniau B, Coutrot M, Longer R, Amiot Q, et al. Activation of the renin-angiotensin-aldosterone system is associated with Acute Kidney Injury in COVID-19. *Anaesth Crit Care Pain Med* 2020 Aug;39(4):453-455 [FREE Full text] [doi: [10.1016/j.accpm.2020.06.006](https://doi.org/10.1016/j.accpm.2020.06.006)] [Medline: [32565254](https://pubmed.ncbi.nlm.nih.gov/32565254/)]
41. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)] [Medline: [31986264](https://pubmed.ncbi.nlm.nih.gov/31986264/)]
42. Pan L, Mu M, Yang P, Sun Y, Wang R, Yan J, et al. Clinical Characteristics of COVID-19 Patients With Digestive Symptoms in Hubei, China: A Descriptive, Cross-Sectional, Multicenter Study. *Am J Gastroenterol* 2020 May;115(5):766-773 [FREE Full text] [doi: [10.14309/ajg.0000000000000620](https://doi.org/10.14309/ajg.0000000000000620)] [Medline: [32287140](https://pubmed.ncbi.nlm.nih.gov/32287140/)]
43. Parohan M, Yaghoubi S, Seraji A. Liver injury is associated with severe coronavirus disease 2019 (COVID-19) infection: A systematic review and meta-analysis of retrospective studies. *Hepatol Res* 2020 Aug;50(8):924-935 [FREE Full text] [doi: [10.1111/hepr.13510](https://doi.org/10.1111/hepr.13510)] [Medline: [32386449](https://pubmed.ncbi.nlm.nih.gov/32386449/)]
44. Lu G, Wang J. Dynamic changes in routine blood parameters of a severe COVID-19 case. *Clin Chim Acta* 2020 Sep;508:98-102 [FREE Full text] [doi: [10.1016/j.cca.2020.04.034](https://doi.org/10.1016/j.cca.2020.04.034)] [Medline: [32405079](https://pubmed.ncbi.nlm.nih.gov/32405079/)]
45. Chen R, Sang L, Jiang M, Yang Z, Jia N, Fu W, Medical Treatment Expert Group for COVID-19. Longitudinal hematologic and immunologic variations associated with the progression of COVID-19 patients in China. *J Allergy Clin Immunol* 2020 Jul;146(1):89-100 [FREE Full text] [doi: [10.1016/j.jaci.2020.05.003](https://doi.org/10.1016/j.jaci.2020.05.003)] [Medline: [32407836](https://pubmed.ncbi.nlm.nih.gov/32407836/)]
46. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7)] [Medline: [32007143](https://pubmed.ncbi.nlm.nih.gov/32007143/)]
47. Cowardin CA, Buonomo EL, Saleh MM, Wilson MG, Burgess SL, Kuehne SA, et al. The binary toxin CDT enhances *Clostridium difficile* virulence by suppressing protective colonic eosinophilia. *Nat Microbiol* 2016 Jul 11;1(8):16108 [FREE Full text] [doi: [10.1038/nmicrobiol.2016.108](https://doi.org/10.1038/nmicrobiol.2016.108)] [Medline: [27573114](https://pubmed.ncbi.nlm.nih.gov/27573114/)]
48. Hassani M, Leijte G, Bruse N, Kox M, Pickkers P, Vriskoop N, et al. Differentiation and activation of eosinophils in the human bone marrow during experimental human endotoxemia. *J Leukoc Biol* 2020 Nov;108(5):1665-1671. [doi: [10.1002/JLB.1AB1219-493R](https://doi.org/10.1002/JLB.1AB1219-493R)] [Medline: [31922294](https://pubmed.ncbi.nlm.nih.gov/31922294/)]

Abbreviations

- ACE2:** angiotensin converting enzyme 2
- AdaBoost:** adaptive boosting
- ALB:** albumin level
- APACHE II:** Acute Physiologic Assessment and Chronic Health Evaluation II
- AUC:** area under the receiver operating characteristic curve
- BASO%:** basophil percentage
- Cr:** creatinine
- EOS%:** eosinophil percentage
- GBDT:** gradient boosting decision tree

ICU: intensive care unit
LASSO: Least Absolute Shrinkage and Selection Operator
LDH: lactate dehydrogenase
LYM%: lymphocyte percentage
MODS: Multiple Organ Dysfunction Score
NEUT%: neutrophil percentage
NLR: negative likelihood ratio
NPV: negative predictive value
PLR: positive likelihood ratio
PPV: positive predictive value
PROBAST: Prediction model Risk Of Bias ASsessment Tool
PSI: Pneumonia Severity Index
PT: prothrombin time
RAAS: renin-angiotensin-aldosterone system
sCr: serum creatinine
SHAP: SHapley Additive exPlanations
SOFA: Sequential Organ Failure Assessment
T-Bil: total bilirubin
TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
XGBoost: eXtreme Gradient Boosting

Edited by G Eysenbach, R Kukafka; submitted 02.08.20; peer-reviewed by A Benis, J Li; comments to author 26.08.20; revised version received 06.09.20; accepted 08.10.20; published 11.11.20

Please cite as:

*Pan P, Li Y, Xiao Y, Han B, Su L, Su M, Li Y, Zhang S, Jiang D, Chen X, Zhou F, Ma L, Bao P, Xie L
Prognostic Assessment of COVID-19 in the Intensive Care Unit by Machine Learning Methods: Model Development and Validation
J Med Internet Res 2020;22(11):e23128
URL: <https://www.jmir.org/2020/11/e23128>
doi: [10.2196/23128](https://doi.org/10.2196/23128)
PMID: [33035175](https://pubmed.ncbi.nlm.nih.gov/33035175/)*

©Pan Pan, Yichao Li, Yongjiu Xiao, Bingchao Han, Longxiang Su, Mingliang Su, Yansheng Li, Siqi Zhang, Dapeng Jiang, Xia Chen, Fuquan Zhou, Ling Ma, Pengtao Bao, Lixin Xie. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 11.11.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.