ORIGINAL RESEARCH

Machine Learning Algorithms are Superior to Conventional Regression Models in Predicting Risk Stratification of COVID-19 Patients

Jiru Ye[1]
Meng Hua[2]
Feng Zhu[2]

[1]Department of Respiratory and Critical Care Medicine, The Third Affiliated Hospital of Soochow University, Changzhou, 213003, China; [2]Department of Respiratory and Critical Care Medicine, Wuxi Fifth People's Hospital, Wuxi, 214000, China

Background: It is very important to determine the risk of patients developing severe or critical COVID-19, but most of the existing risk prediction models are established using conventional regression models. We aim to use machine learning algorithms to develop predictive models and compare predictive performance with logistic regression models.

Methods: The medical record of 161 COVID-19 patients who were diagnosed January–April 2020 were retrospectively analyzed. The patients were divided into two groups: asymptomatic-moderate group (132 cases) and severe or above group (29 cases). The clinical features and laboratory biomarkers of these two groups were compared. Machine learning algorithms and multivariate logistic regression analysis were used to construct two COVID-19 risk stratification prediction models, and the area under the curve (AUC) was used to compare the predictive efficacy of these two models.

Results: A machine learning model was constructed based on seven characteristic variables: high sensitivity C-reactive protein (hs-CRP), procalcitonin (PCT), age, neutrophil count (Neuc), hemoglobin (HGB), percentage of neutrophils (Neur), and platelet distribution width (PDW). The AUC of the model was 0.978 (95% CI: 0.960–0.996), which was significantly higher than that of the logistic regression model (0.827; 95% CI: 0.724–0.930) (P=0.002). Moreover, the machine learning model's sensitivity, specificity, and accuracy were better than those of the logistic regression model.

Conclusion: Machine learning algorithms improve the accuracy of risk stratification in patients with COVID-19. Using detection algorithms derived from these techniques can enhance the identification of critically ill patients.

Keywords: COVID-19, machine learning, prediction model, high sensitivity C-reactive protein, procalcitonin

Background

Novel Coronavirus disease 2019 (COVID-19) is a new respiratory and systemic disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Since its report in December 2019, it has been prevalent worldwide and has brought tremendous global public health challenges.[1–3] SARS-CoV-2 infection can be either asymptomatic or with a variety of symptoms. The clinical manifestations of most COVID-19 patients are fever, fatigue, and dry cough, and a few patients can be critically ill and may have serious pneumonia and multi-organ failure, leading to death eventually.[4–6] Thus, it is very important to identify the risk of patients developing severe or critical COVID-19 so that patients with poor prognosis can

Correspondence: Feng Zhu; Meng Hua
Department of Respiratory and Critical Care Medicine, Wuxi Fifth People's Hospital, Wuxi, 214000, China
Tel +8613914135026; +8615950525583
Email 181515688@qq.com;
myecholemon@qq.com

receive timely intervention to minimize the disease's progression early. For this purpose, some COVID-19 risk prediction models have been developed and studied, but most of them were established using conventional regression models.[7–9]

With the development of science and technology, various machine learning algorithms and artificial intelligence technologies have been widely used in patient tracking, vaccine development, and patient screening due to their better extensibility and faster processing ability.[10,11] However, applying machine learning algorithms and artificial intelligence in identifying disease progression and estimating the risk of death is relatively rare.[12] Based on the general data and laboratory indexes of COVID-19 patients, we developed two multivariate prediction models using machine learning algorithms and multivariate logistic regression analysis to predict the risk stratification of COVID-19 patients and compared the prediction performance of the two models.

Methods

General Information

This was a retrospective cross-sectional study. We analyzed the medical record of 170 patients treated for novel coronavirus infection in the negative pressure ward of Wuxi Fifth People's Hospital between January and April 2020. This study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Ethics Committee for retrospective analysis (No. 2021-001-1). Since the patients' medical information was anonymous, informed consent from the participants was not a requirement. Patient inclusion criteria: ① Novel coronavirus nucleic acid positive detected by Real-time fluorescence RT-PCR. ② In line with the diagnostic criteria of the Diagnosis and Treatment Protocol for Novel Coronavirus Pneumonia (trial version 7).[13] Patient exclusion criteria: Patients under 15 years old were excluded (nine patients). The general data, complications, and routine laboratory test results of all patients were recorded.

Diagnosis and Treatment

The clinical classification of COVID-19 patients was mostly based on symptoms and imaging findings. According to the standard established in the Diagnosis and Treatment Protocol for Novel Coronavirus Pneumonia (trial version 7) published by the General Office of the National Health Commission of China,[13]

COVID-19 patients were classified as mild, moderate, severe, and critical cases. Patients with mild COVID-19 showed mild clinical symptoms without any pneumonia sign on images. Moderate cases showed fever, respiratory symptoms, and imaging findings of pneumonia. Patients who met any of the definitions described below were considered severe cases: respiratory distress (RR ≥30 breaths/min); oxygen saturation ≤93% at rest; arterial oxygen partial pressure/fraction of inspired oxygen ≤300 mmHg. Patients who met any of the criteria described below were considered critical cases: respiratory failure requiring mechanical ventilation; shock; combined with other organ failures that required ICU care. Patients whose pulmonary imaging showed that the lesions significantly progressed over 50% during 24–48 hours were managed as severe cases. For the treatment of COVID-19 patients, antiviral therapy such as interferon-α, lopinavir/ritonavir, and ribavirin can be used. Effective oxygen therapy can be given in time, and traditional Chinese medicine can also be used selectively. For severe and critical cases, in addition to the treatment to relieve the symptoms, extra care should be given to actively prevent and treat complications, treat basic diseases, prevent secondary infection, and provide organ function support in time.

Laboratory Testing

White blood cell (WBC, reference range: $3.50–9.50×10^9$/L), the percentage of neutrophils (Neur, reference range: 40.00–75.00%), neutrophil count (Neuc, reference range: $1.80–6.30×10^9$/L), the percentage of lymphocytes (Lymr, reference range: 20.00–50.00%), lymphocyte count (Lymc, reference range: $1.10–3.20×10^9$/L), the percentage of monocytes (Monr, reference range: 3.00–8.00%), monocyte count (Monc, reference range: $0.10–0.60×10^9$/L), red blood cell count (RBC, reference range: female, $3.80–5.10×10^{12}$/L; male, $4.30–5.80×10^{12}$/L), hemoglobin (HGB, reference range: female, 115–150 g/L; male, 130–175 g/L), hematocrit (HCT, reference range: female, 35.0–45.0%; male, 40.0–50.0%), platelet count (PLT, reference range: $125–350×10^9$/L), red blood cell distribution width (RDW, reference range: 11.50–14.90%), plateletcrit (PTC, reference range: 0.108–0.272 L/L), mean platelet volume (MPV, reference range: 6.00–11.50 fL), platelet distribution width (PDW, reference range: 15.50–18.10 fL), high sensitivity C-reactive protein (hs-CRP, reference range: 0–10 mg/L), procalcitonin (PCT, reference range: 0–0.05 ng/mL). The routine blood test was performed using Sysmex XN9000 hematology analyzer

(Sysmex Corporation, Hyogo, Japan), hs-CRP was determined using the protein analyzer HP-083/4 (Hipro Biotechnology, Shijiazhuang, China), PCT was detected using an Autobio A2000PLUS automatic chemiluminescence analyzer (Sym-Biotechnology, Suzhou, China).[14]

Statistical Analysis

Continuous variables were presented as the mean ± SD or median (Q1-Q3), and categorical variables were presented as frequency (%). Unpaired Student's *t*-test or Mann Whitney nonparametric test was used to comparing continuous variables. Pearson chi-square test and Fisher exact test were used to analyzing categorical variables.

Use general data and laboratory indicators for predicting risk stratification (see Tables 1 and 2) to train a machine learning model (eXtreme Gradient Boosting, XGBoost).[15] The optimized model hyperparameters are set as follows: booster = gbtree, objective = binary: logistic, eta = 0.3, gamma = 5, max_depth = 6, min_child_weight = 1, subsample = 1, colsample_bytree = 1. The XGBoost model had been proven to provide the most advanced results for various medical applications and had won numerous awards in machine learning algorithms. To evaluate the importance of features developed by the model, three importance scores

Table 1 Comparison of the General Data Between the Two Groups

Classification	Asymptomatic-Moderate Group	Severe and Above Group	*P*-value
N	132	29	
Age	43.8 ± 15.0	59.6 ± 14.8	<0.001
Age group			<0.001
<60	114 (86.4%)	17 (58.6%)	
≥60	18 (13.6%)	12 (41.4%)	

Gender
  Female
  Male
	
	61 (46.2%)
	71 (53.8%)
		
		10 (34.5%)
		19 (65.5%)
			0.249

Hypertension	31 (23.5%)	12 (41.4%)	0.049
Diabetes	14 (10.6%)	10 (34.5%)	0.001
Coronary heart disease	3 (2.3%)	2 (6.9%)	0.194
Cerebrovascular diseases	0 (0.0%)	2 (6.9%)	0.002
Tumor	1 (0.8%)	2 (6.9%)	0.084
HBV	1 (0.8%)	0 (0.0%)	0.638
Chronic renal disease	1 (0.8%)	1 (3.4%)	0.239
Chronic liver disease/ cirrhosis	3 (2.3%)	0 (0.0%)	0.412
Alcoholism	4 (3.0%)	1 (3.4%)	0.895
Smoking	5 (3.8%)	2 (6.9%)	0.639

Note: Results in the table: Mean ± SD/N (%).
Abbreviation: HBV, Hepatitis B virus.

Table 2 Comparison of Laboratory Indexes Between Asymptomatic-Moderate Group and Severe or Above Group

Classification	Asymptomatic-Moderate Group	Severe and Above Group	*P*-value
N	132	29	
WBC (10^9/L)	4.95 ± 1.55	5.50 ± 2.72	0.143
Neur (%)	59.84 ± 11.08	70.46 ± 16.12	<0.001
Lymr (%)	29.38 ± 10.28	21.15 ± 12.56	<0.001
Monr (%)	9.67 ± 3.26	7.51 ± 3.67	0.002
Neuc (10^9/L)	3.00 ± 1.28	4.12 ± 2.84	0.121
Lymc (10^9/L)	1.42 ± 0.62	0.97 ± 0.53	<0.001
Monc (10^9/L)	0.47 ± 0.19	0.37 ± 0.18	0.009
RBC (10^{12}/L)	4.73 ± 0.68	4.27 ± 0.57	<0.001
HGB (g/L)	139.71 ± 18.39	130.10 ± 21.25	0.015
HCT (%)	41.21 ± 4.79	37.72 ± 5.53	<0.001
PLT (10^9/L)	179.65 ± 55.38	173.93 ± 56.50	0.617
RDW (%)	13.18 ± 1.91	13.55 ± 2.79	0.526
MPV (fL)	11.06 ± 1.23	10.89 ± 1.17	0.514
PDW (fL)	14.98 ± 3.17	15.34 ± 2.08	0.559
PTC (L/L)	0.20 ± 0.05	0.18 ± 0.06	0.180
Hs-CRP (mg/L)	7.50 (2.30–23.90)	71.77 (25.90–108.10)	<0.001
PCT (ng/mL)	0.12 (0.02–0.20)	0.20 (0.17–0.34)	<0.001

Note: Results in the table: Mean ± SD/Median (Q1-Q3)/N (%).
Abbreviations: WBC, white blood cell; Neur, percentage of neutrophils; Lymr, percentage of lymphocytes; Monr, percentage of monocytes; Neuc, neutrophil count; Lymc, lymphocyte count; Monc, monocyte count; RBC, red blood cell count; HGB, hemoglobin; HCT, hematocrit; PLT, platelet count; RDW, red blood cell distribution width; MPV, mean platelet volume; PDW, platelet distribution width; PTC, plateletcrit; Hs-CRP, high sensitivity C-reactive protein; PCT, procalcitonin.

of Gain, Cover, and Frequency are calculated, where Gain is the most relevant attribute that explains the relative importance of each feature.

The indicators selected by the machine learning algorithm are used as the model parameters (independent variables), the patient risk stratification is used as the dependent variable, the multivariate Logistic regression method is used to establish a prediction model of all independent variables, and the best model parameters (including intercept, regression coefficients of each independent variable).

The XGBoost model will predict each case and generate the predicted probability (P) whether the patient is diagnosed as severe or above; use different cutoffs for the predicted probability to determine the stratification of the patient, and for each cutoff, the corresponding sensitivity and specificity and draw a receiver operating characteristic (ROC) curve, calculate the area under the curve (AUC) and 95% confidence interval (CI). For the Logistic regression model, the ROC curve is drawn by calculating each case's Logit (P), and the AUC and 95% CI are calculated. Use the DeLong test to compare whether the AUC of the two models is significantly different. R software was used for all statistical analyses, version 3.4.3 (http://www.R-project.org). $P<0.05$ was considered that the difference was statistically significant.

Results

Our study eventually included 161 COVID-19 patients, including 90 males and 71 females, with an average age of 46.7 ± 16.1 years (range: 15–91 years). Based on the symptoms and imaging findings at hospital admission, the patients were classified as 12 asymptomatic cases, 32 mild cases, 88 moderate cases, 27 severe cases, and 2 critical cases of COVID-19 (15 patients with moderate COVID-19 at admission progressed to severe cases two days after admission and thus they were classified as severe cases). Based on the clinical significance of treatment, we divided the patients into two groups: an asymptomatic-moderate group (132 cases) and a severe or above group (29 cases). We compared the general data between the two groups (Table 1) and found that compared with the asymptomatic-moderate group, the patients in severe group were older ($P<0.001$), and the number of patients with hypertension, diabetes, and cerebrovascular disease was remarkably higher (all $P<0.05$). We noticed that the number of patients who had tumor history was also higher in severe group, but the increase was not statistically significant ($P=0.084$).

Comparing the laboratory indexes between the two groups, we found that the percentage of neutrophils, hs-CRP, and PCT were markedly higher in severe group than in asymptomatic-moderate group ($P<0.001$), but the percentage of lymphocytes, percentage of monocytes, lymphocyte count, monocyte count, red blood cell count, hemoglobin and hematocrit were all remarkably lower in severe group than in asymptomatic-moderate group (Table 2, $P<0.05$).

Machine Learning Algorithms

The machine learning algorithms were applied to predict the risk stratification using the general data and laboratory indexes of COVID-19 patients. The goal of classification was to identify critically ill COVID-19 patients. We sorted various factors according to their importance to the risk prediction (Table 3 and Figure 1) and found that hs-CRP, PCT, and age were the top three risk factors, followed by four routine hematological indexes: Neuc, HGB, Neur, and PDW. Among these factors, hs-CRP and PCT were more important in COVID-19 risk prediction than the four routine hematological indexes.

Next, we established another prediction model using multivariate logistic regression analysis with the above seven parameters as independent variables, and whether the patient was diagnosed as severe or above case as the dependent variable:

Logit (P) = −7.05139 +2.31599 × procalcitonin + 0.00264 × high sensitivity C-reactive protein + 0.06364 × age + 0.14735 × neutrophil count + 0.02677 × percentage of neutrophils − 0.00751 × hemoglobin + 0.02564 × platelet distribution width. The P value in the above formula was the probability that the patient was diagnosed as severe or above case.

Comparing the ROC curve of machine learning model and logistic regression model in predicting clinical classification of COVID-19 (Figure 2), we found that the AUC

Table 3 The Importance of Variables

Feature	Gain	Cover	Frequency
hs-CRP	0.40746360	0.47782767	0.31578947
PCT	0.24389755	0.19765398	0.26315789
Age	0.19192806	0.20253600	0.21052632
Neuc	0.04459240	0.03011219	0.05263158
HGB	0.04254448	0.04640159	0.05263158
Neur	0.04108720	0.01675670	0.05263158
PDW	0.02848671	0.02871187	0.05263158

Abbreviations: hs-CRP, high sensitivity C-reactive protein; PCT, procalcitonin; Neuc, neutrophil count; HGB, hemoglobin; Neur, percentage of neutrophil; PDW, platelet distribution width.
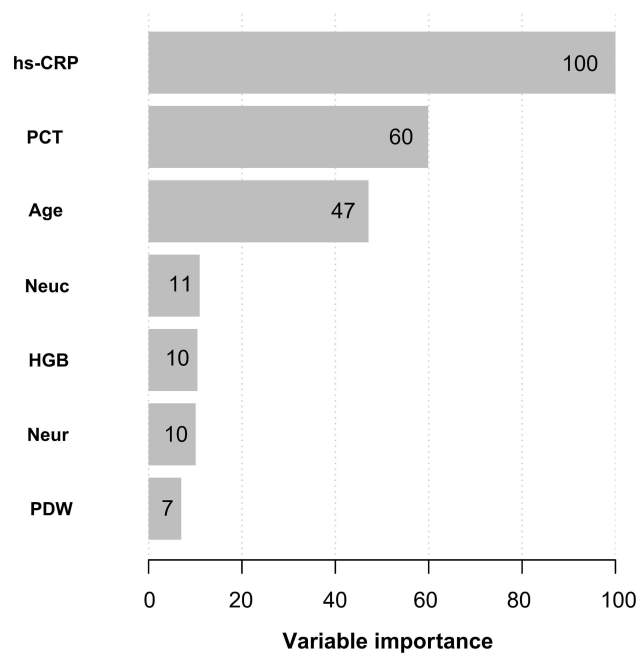
Figure 1 Importance of the predictor variables in the XGBoost model, scaled to a maximum of 100.
Abbreviations: hs-CRP, high sensitivity C-reactive protein; PCT, procalcitonin; Neuc, neutrophil count; HGB, hemoglobin; Neur, percentage of neutrophil; PDW, platelet distribution width.

value was significantly different between the two models (0.978 vs 0.827, P=0.002), and the sensitivity, specificity, and accuracy of machine learning model were all the better than those of logistic regression model (Table 4).
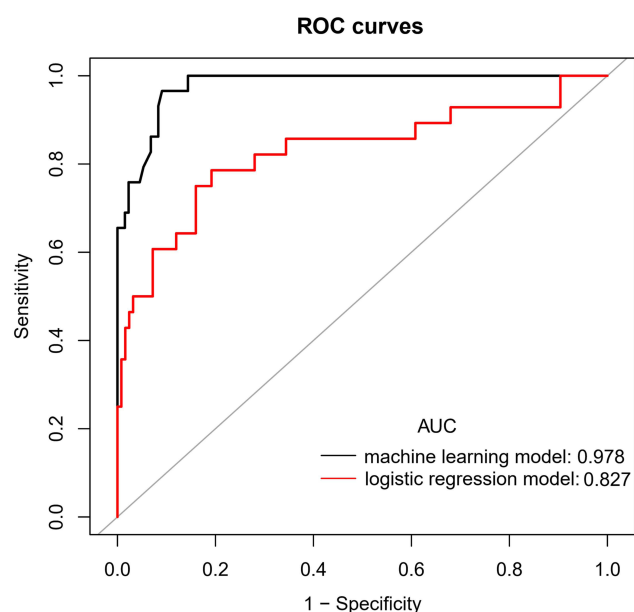

Figure 2 Receiver operating characteristic curve for estimating the prediction efficacy of the logistic regression model and machine learning model.

Discussion

Our research showed that the machine learning model constructed with seven characteristic variables, including hs-CRP, PCT, age, Neuc, HGB, Neur, and PDW, has good COVID-19 risk prediction ability, which can be helpful for physicians to predict the progress of disease and intervene the disease in time.

According to the importance of characteristic parameters of XGBoost model, hs-CRP and PCT are important parameters to predict severe COVID-19 patients. The expression level of hs-CRP is usually low, but it increases rapidly and significantly during acute inflammation. Therefore, hs-CRP is a sensitive biomarker of inflammation, infection, and tissue damage.[16] Previous reports have also suggested that hs-CRP is an important biomarker for poor prognosis in COVID-19 patients, revealed an enduring status of inflammation,[17,18] which may deeply interact with the inflammatory storm, leading to lung injury and pulmonary edema in COVID-19 patients.[19,20] PCT is a glycoprotein with no hormonal activity. It is a precursor of calcitonin[21,22] that can be used as a biomarker to assess the severity of sepsis and the prognosis of patients with sepsis.[23] It can be used to guide antibiotic treatment. Some studies have shown that the level of PCT is positively correlated with the progression of COVID-19.[24,25] It has been reported that angiotensin converting enzyme 2 (ACE2) receptor is expressed in vascular endothelial cells. SARS-CoV-2 can bind to ACE2 receptors and invade host cells, which leads to endothelial dysfunction, increases the possibility of cytokine storm, and produces a series of immune responses,[3,26–28] resulting in adverse clinical outcomes and death. Since the immune function of severe COVID-19 patients is low, these patients are more prone to infection,[8] which leads to the increase of inflammatory markers. Wang et al[29] also suggested that 81.7% of the deaths in COVID-19 patients were associated with a bacterial infection. Although there is a correlation between the above inflammatory markers and the severity of COVID-19 patients, these inflammatory markers' role in the pathogenesis of COVID-19 is not fully understood and needs further verification and in-depth study.

Age is also an important determinant in XGBoost model. SARS-CoV-2 infection may deteriorate the chronic inflammation in elderly patients, such as hypertension, diabetes, and cardiovascular disease, leading to death. Therefore, old age is also a risk factor for severe COVID-19 patients.[30,31]

Table 4 Comparison of the Diagnostic Efficiency Between the Machine Learning Model and the Logistic Regression Model

Model	AUC (95% CI)	Cutoff	Specificity	Sensitivity	Accuracy	Positive-LR	Negative-LR
Machine learning model	0.978 (0.960–0.996)	0.1743	0.909	0.966	0.919	10.621	0.038
Logistic regression model	0.827 (0.724–0.930)	−1.3998	0.808	0.786	0.804	4.092	0.265

Abbreviations: Positive LR, positive likelihood ratio; Negative LR, negative likelihood ratio.

Also, routine blood indexes: Neuc, HGB, Neur, and PDW are also COVID-19 risk prediction factors in the model, and the key roles of these markers have been confirmed in other reports.[32,33] However, the importance of hs-CRP and PCT in COVID-19 risk prediction is much higher than that of routine blood indexes.

It has been reported that machine learning algorithms can maximize clinical parameters and improve the accuracy of diagnosis.[34] Our results showed that XGBoost model has a good prediction performance, the specificity, sensitivity, and accuracy are all above 90%, and the diagnostic efficiency is better than the logistic regression method established using the same parameters, indicating that machine learning algorithms are more accurate and sensitive than conventional logistic regression analysis, which is consistent with the results of Pan et al.[35] Fernandez et al[36] also pointed out that XGBoost method is more reliable, especially when the sample size is limited. Moreover, the seven key parameters of the risk prediction model we established in this study can be obtained at admission. Therefore, early detection of these parameters can help identify patients with severe COVID-19 to receive timely intervention and appropriate intensive care to minimize disease progression.

There are some limitations to our study. Firstly, this is a retrospective, single-center study, which may lead to biased conclusions. Secondly, since we did not conduct external verification on the model, it is necessary to establish a prospective study cohort to further verify the model's accuracy. Finally, the data used to build the COVID-19 risk prediction model is completely from China, which may not apply to other regions.

Conclusions

In this study, seven characteristic variables, namely hs-CRP, PCT, age, Neuc, HGB, Neur, and PDW, were used to construct XGBoost model and logistic regression model. The machine learning method improved the accuracy of risk stratification for patients with COVID-19 and could effectively assess the severity of patients with COVID-19. The diagnostic

efficiency was better than logistic regression method based on the same parameters. It is helpful for clinicians to identify critical patients in the early stage, but further verification is needed to make our findings applied in clinical practice.

Data Sharing Statement

All data generated or analyzed during this study are available from the corresponding author upon reasonable request.

Ethics Approval and Informed Consent

This study was approved by the Institutional Ethics Committee of Wuxi Fifth People's Hospital for retrospective analysis (No. 2021-001-1).

Acknowledgments

Author Contributions

Feng Zhu and Meng Hua are co-corresponding authors. All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Disclosure

The authors report no conflicts of interest in this work.

References

1. Fan Z, Chen L, Li J, et al. Clinical features of COVID-19-related liver functional abnormality. *Clin Gastroenterol Hepatol*. 2020;18 (7):1561–1566. doi:10.1016/j.cgh.2020.04.002

2. Wu J, Song S, Cao HC, Li LJ. Liver diseases in COVID-19: etiology, treatment and prognosis. *World J Gastroenterol*. 2020;26 (19):2286–2293. doi:10.3748/wjg.v26.i19.2286

3. Martins-Filho PR, Tavares CSS, Santos VS. Factors associated with mortality in patients with COVID-19. A quantitative evidence synthesis of clinical and laboratory data. *Eur J Intern Med*. 2020;76:97–99. doi:10.1016/j.ejim.2020.04.043

4. Liu Q, Song NC, Zheng ZK, Li JS, Li SK. Laboratory findings and a combined multifactorial approach to predict death in critically ill patients with COVID-19: a Retrospective Study. *Epidemiol Infect*. 2020;148:e129. doi:10.1017/S0950268820001442

5. Wang F, Hou H, Luo Y, et al. The laboratory tests and host immunity of COVID-19 patients with different severity of illness. *JCI Insight*. 2020;5(10):e137799. doi:10.1172/jci.insight.137799

6. Vultaggio A, Vivarelli E, Virgili G, et al. Prompt predicting of early clinical deterioration of moderate-to-severe COVID-19 patients: usefulness of a combined score using IL-6 in a Preliminary Study. *J Allergy Clin Immunol Pract*. 2020;8(8):2575–2581.e2. doi:10.1016/j.jaip.2020.06.013

7. Wang F, Hou H, Wang T, et al. Establishing a model for predicting the outcome of COVID-19 based on combination of laboratory tests. *Travel Med Infect Dis*. 2020;36:101782. doi:10.1016/j. tmaid.2020.101782

8. Zeng Z, Ma Y, Zeng H, et al. Simple nomogram based on initial laboratory data for predicting the probability of ICU transfer of COVID-19 patients: Multicenter Retrospective Study. *J Med Virol*. 2021;93(1):434–440. doi:10.1002/jmv.26244

9. Liu YP, Li GM, He J, et al. Combined use of the neutrophil-to-lymphocyte ratio and CRP to predict 7-day disease severity in 84 hospitalized patients with COVID-19 pneumonia: a Retrospective Cohort Study. *Ann Transl Med*. 2020;8(10):635. doi:10.21037/atm-20-2372

10. Albahri AS, Hamid RA, Alwan JK, et al. Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (COVID-19): a systematic review. *J Med Syst*. 2020;44(7):122. doi:10.1007/s10916-020-01582-x

11. Aggarwal L, Goswami P, Sachdeva S. Multi-criterion intelligent decision support system for COVID-19. *Appl Soft Comput*. 2021;101:107056. doi:10.1016/j.asoc.2020.107056

12. Guan X, Zhang B, Fu M, et al. Clinical and inflammatory features based machine learning model for fatal risk prediction of hospitalized COVID-19 patients: results from a Retrospective Cohort Study. *Ann Med*. 2021;53(1):257–266. doi:10.1080/07853890.2020.1868564

13. National Health Commission. Diagnosis and treatment protocol for novel coronavirus pneumonia (trial version 7). *Chin Med J (Engl)*. 2020;133(9):1087–1095. doi:10.1097/CM9.0000000000000819

14. Ye J, Zhang X, Zhu F, Tang YJE. Application of a prediction model with laboratory indexes in the risk stratification of patients with COVID-19. *Medicine*. 2021;21(3):182.

15. Livne M, Boldsen JK, Mikkelsen IK, Fiebach JB, Sobesky J, Mouridsen K. Boosted tree model reforms multimodal magnetic resonance imaging infarct prediction in acute stroke. *Stroke*. 2018;49(4):912–918. doi:10.1161/STROKEAHA.117.019440

16. Wang K, Zuo P, Liu Y, et al. Clinical and laboratory predictors of in-hospital mortality in patients with coronavirus disease-2019: a Cohort Study in Wuhan, China. *Clin Infect Dis*. 2020;71 (16):2079–2088. doi:10.1093/cid/ciaa538

17. Toniati P, Piva S, Cattalini M, et al. Tocilizumab for the treatment of severe COVID-19 pneumonia with hyperinflammatory syndrome and acute respiratory failure: a Single Center Study of 100 patients in Brescia, Italy. *Autoimmun Rev*. 2020;19(7):102568. doi:10.1016/j. autrev.2020.102568

18. Liu Y, Yang Y, Zhang C, et al. Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. *Sci China Life Sci*. 2020;63(3):364–374. doi:10.1007/ s11427-020-1643-8

19. Barton LM, Duval EJ, Stroberg E, Ghosh S, Mukhopadhyay S. COVID-19 autopsies, Oklahoma, USA. *Am J Clin Pathol*. 2020;153 (6):725–733. doi:10.1093/ajcp/aqaa062

20. Kapur R, Kim M, Shanmugabhavananthan S, Liu J, Li Y, Semple JW. C-reactive protein enhances murine antibody-mediated transfusion-related acute lung injury. *Blood*. 2015;126 (25):2747–2751. doi:10.1182/blood-2015-09-672592

21. Saeed K, Dale AP, Leung E, et al. Procalcitonin levels predict infectious complications and response to treatment in patients undergoing cytoreductive surgery for peritoneal malignancy. *Eur J Surg Oncol*. 2016;42(2):234–243. doi:10.1016/j.ejso.2015.10.004

22. Choi JJ, McCarthy MW. Novel applications for serum procalcitonin testing in clinical practice. *Expert Rev Mol Diagn*. 2018;18(1):27–34. doi:10.1080/14737159.2018.1407244

23. Sager R, Kutz A, Mueller B, Schuetz P. Procalcitonin-guided diagnosis and antibiotic stewardship revisited. *BMC Med*. 2017;15(1):15. doi:10.1186/s12916-017-0795-7

24. Hu R, Han C, Pei S, Yin M, Chen X. Procalcitonin levels in COVID-19 patients. *Int J Antimicrob Agents*. 2020;56(2):106051. doi:10.1016/j.ijantimicag.2020.106051

25. Liu F, Li L, Xu M, et al. Prognostic value of interleukin-6, C-reactive protein, and procalcitonin in patients with COVID-19. *J Clin Virol*. 2020;127:104370. doi:10.1016/j.jcv.2020.104370

26. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. *JAMA*. 2020;324(8):782. doi:10.1001/jama.2020.12839

27. Gheblawi M, Wang K, Viveiros A, et al. Angiotensin-converting enzyme 2: SARS-CoV-2 receptor and regulator of the renin-angiotensin system: celebrating the 20th anniversary of the discovery of ACE2. *Circ Res*. 2020;126(10):1456–1474. doi:10.1161/CIRCRESAHA.120.317015

28. Bavishi C, Maddox TM, Messerli FH. Coronavirus disease 2019 (COVID-19) infection and renin angiotensin system blockers. *JAMA Cardiol*. 2020;5(7):745–747. doi:10.1001/ jamacardio.2020.1282

29. Wang L, He W, Yu X, et al. Coronavirus disease 2019 in elderly patients: characteristics and prognostic factors based on 4-week follow-up. *J Infect*. 2020;80(6):639–645. doi:10.1016/j. jinf.2020.03.019

30. Ou M, Zhu J, Ji P, et al. Risk factors of severe cases with COVID-19: a meta-analysis. *Epidemiol Infect*. 2020;148:e175. doi:10.1017/ S095026882000179X

31. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ*. 2020;369:m1328. doi:10.1136/bmj.m1328

32. Zhang B, Zhang J, Chen H, et al. Novel coronavirus disease 2019 (COVID-19): relationship between chest CT scores and laboratory parameters. *Eur J Nucl Med Mol Imaging*. 2020;47(9):2083–2089. doi:10.1007/s00259-020-04854-3

33. Ghahramani S, Tabrizi R, Lankarani KB, et al. Laboratory features of severe vs. non-severe COVID-19 patients in Asian populations: a systematic review and meta-analysis. *Eur J Med Res*. 2020;25 (1):30. doi:10.1186/s40001-020-00432-3

34. Assaf D, Gutman Y, Neuman Y, et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Intern Emerg Med*. 2020;15(8):1435–1443. doi:10.1007/s11739-020-02475-0

35. Pan P, Li Y, Xiao Y, et al. 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. doi:10.2196/23128

36. Fernandez DM, Cernadas E, Barro S, Amorim DJ. Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res*. 2014;15:3133–3181.