Influencing Factors of Serious COVID-19 and the Construction of its Risk Prediction Model

Ling Zhang[1], Xinran Li[1], Ziyan Wang[1], Lei Zhao[2], Huixia Gao[2], Conghui Liu[1], Jing Bai[1], TieJun Liu[1], Weibin Chen[1], Wenqiang Li[3], Jingshan Bai[4], Ai-Shuang Fu[1], and Yan-Lei Ge[1]

[1]North China University of Science and Technology Affiliated Hospital
[2]Shijiazhuang Fifth Hospital
[3]Zigong First People's Hospital
[4]Xuanwu Hospital Capital Medical University

February 10, 2025

Abstract

A clinical case-control study was conducted to screen the influencing factors of patients with coronavirus disease 2019 (COVID-19) and to construct a clinical prediction model to provide a reference for the dynamic assessment of the severity of COVID-19 patients. A total of 410 patients with COVID-19 were included in the study, of which 132 were severe or critical cases. The clinical data of patients were collected, and then variables were screened by lasso regression analysis and 10-fold cross-validation. The screened variables were subjected to multifactorial logistic regression analysis to screen out the independent risk factors of patients with severe or critical illnesses, and the independent risk factors were integrated to construct a nomogram. The receiver operating characteristic curve (ROC), calibration curve analysis, and decision curve analysis (DCA) were used to assess the model efficiency. Five variables, including respiratory rate(R), systolic blood pressure (SBP), plasma albumin (ALB), lactate dehydrogenase (LDH), and C-reactive protein (CRP), were finally included to construct a clinical prediction model, with an area under the curve (AUC) of 0.86 (CI: 0.82% to 0.90%).

Influencing Factors of Serious COVID-19 and the Construction of its Risk Prediction Model

Ling Zhang [1], Xinran Li [1], Ziyan Wang [1], Lei Zhao [2,3], Huixia Gao[2,3], Conghui Liu [1], Jing Bai[1], Tiejun Liu [1], Weibin Chen[1], Wenqiang Li [4], Jingshan Bai[5], Aishuang Fu [1*], Yanlei Ge[1,6,7,8,9*]

1North China University of Science and Technology Affiliated Hospital, Tangshan, Hebei, China

2The Fifth Hospital, Shijiazhuang, Hebei Province

3Hebei Key Laboratory of Immune Mechanism of Major Infectious Diseases and New Technology of Diagnosis and Treatment

4The First People's Hospital of Zigong City, Sichuan Province

5Department of Respiratory and Critical Care Medicine, Xuanwu Hospital, Xiongnan, China

6Hebei Biological Cell Function Development and Precision Detection Technology Innovation Centre, Tangshan, Hebei, China

7Tangshan Key Laboratory of Precision Medicine for Medical-Industrial Integration, Tangshan, Hebei, China

8Tangshan Innovative Technology Centre for Biological Cell Function Development and Detection, Tangshan, Hebei, China

9Sun Guogui Innovation Studio, Tangshan, Hebei, China

*Corresponding Author:

-Name: Yanlei Ge

-Address: North China University of Science and Technology Affiliated Hospital,

Tangshan, Hebei, China

-E-mail: 495732196@qq.com

-Name: Aishuang Fu

-Address: North China University of Science and Technology Affiliated Hospital,

Tangshan, Hebei, China

-E-mail: maxfas@163.com

SUMMARY

A clinica l case-control study was conducted to screen the influencing factors of patients with coronavirus disease 2019 (COVID-19) and to construct a clinical prediction model to provide a reference for the dynamic assessment of the se**verity of COVID-19 patients.** A total of 410 patients with COVID-19 were included in the study, of which 132 were severe or critical cases. The clinical data of patients were collected, and then variables were screened by lasso regression analysis and 10-fold cross-validation. The screened variables were subjected to multifactorial logistic regression analysis to screen out the independent risk factors of patients with severe or critical illnesses, and the independent risk factors were integrated to construct a nomogram. The receiver operating characteristic curve (ROC), calibration curve analysis, and decision curve analysis (DCA) were used to assess the model **efficiency** . Five variables, including respiratory rate(R), systolic blood pressure (SBP), plasma albumin (ALB), lactate dehydrogenase (LDH), and C-reactive protein (CRP), were finally included to construct a clinical prediction model, with an area under the curve (AUC) of 0.86 (CI: 0.82% to 0.90%).

KeyWords: COIVD-19 , SARS-CoV-2 , preticting model , nomogram

Introduction

COVID-19 is a group of acute respiratory infectious diseases found in December 2019 in Wuhan, Hubei Province, China [1]. A virus isolated from it is known as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) [2]. The SARS-CoV-2 virus, like other RNA viruses, exhibits a high degree of variability. Several variant strains have emerged: Alpha, Beta, Gamma, Delta, and Omicron. The Omicron strain is more accessible to transmit than previous variants, has a more extraordinary ability to evade the human immune response, and is characterized by reduced natural virulence [3]. Despite the relatively low virulence of Omicron variants, some populations, such as the elderly with comorbidities, are at high risk of serious infection and death.[4]. The highly infectious nature of SARS-CoV-2 means that it can trigger seasonal epidemic spikes, while virus mutations are unpredictable. The lack of effective monitoring mechanisms or appropriate responses may lead to the emergence of new pandemic patterns. Early and rapid identification of patients with severe COVID-19 during a COVID-19 crisis outbreak and dynamic monitoring of disease changes are essential both for treatment and assessment of prognosis.

The last two years have also seen an influx of prediction models for COVID-19 patients as the virus has mutated. In a cohort study conducted in England [5],the researchers developed a risk prediction algorithm called QCOVID4. The algorithm is based on demographic data and is designed to assess the risk of COVID-19 patients being hospitalized or dying from the disease during the Omicron. The model has a Harrell C statistic 0.970 (0.962 to 0.979). Baker[6] et al. presented a machine learning model with predictive outcomes including acute respiratory failure, intensive care unit admission, or ventilator use. The model exceeded similar predictors with an area under the curve of 0.83 without laboratory data predicting patient risk. The

study's primary outcome by Weng [7] et al. was survival or death at discharge. Unlike Weng, in the study of Moon [8] et al., the prediction time for mortality was defined as 30 and 60 days, which showed the survival time of the patients while predicting their deaths. All of the above models demonstrate good predictive efficacy, but they are all prognostic predictive models, which focus more on the prognosis of patients than diagnostic predictive models. The diagnostic prediction model predicts the current state of the patient. For example, our team's research on machine learning and imaging histology for community-acquired pneumonia [9] uses clinical data and imaging features to identify the current severity of a patient's condition and provide timely treatment to prevent disease progression. Similarly, in a study of diagnostic, predictive models for COVID-19, Zhu [10] et al. used a deep learning approach to analyze CT images to differentiate between severe and mild cases. The use of artificial intelligence in medicine is increasing yearly, and there is a proliferation of studies combining AI for clinical predictive model construction. However, when using AI, we may have difficulty understanding some of the features of the output, the AI may lack common sense in considering background knowledge and qualitative inspection when constructing the model, and differences in inspection equipment in different regions cannot be avoided [11].

In clinical practice, considering the highly contagious nature of SARS-CoV-2, it is easy to concentrate on the outbreak, resulting in a shortage of medical resources. An ideal prediction model should have the following characteristics: low cost, easy to detect, with high specificity and sensitivity, and able to dynamically detect and reflect the severity of the disease promptly.

Therefore, this study aimed to collect clinical data and construct a clinical prediction model using simple and easy-to-access indicators to achieve dynamic monitoring of disease severity in COVID-19 patients.

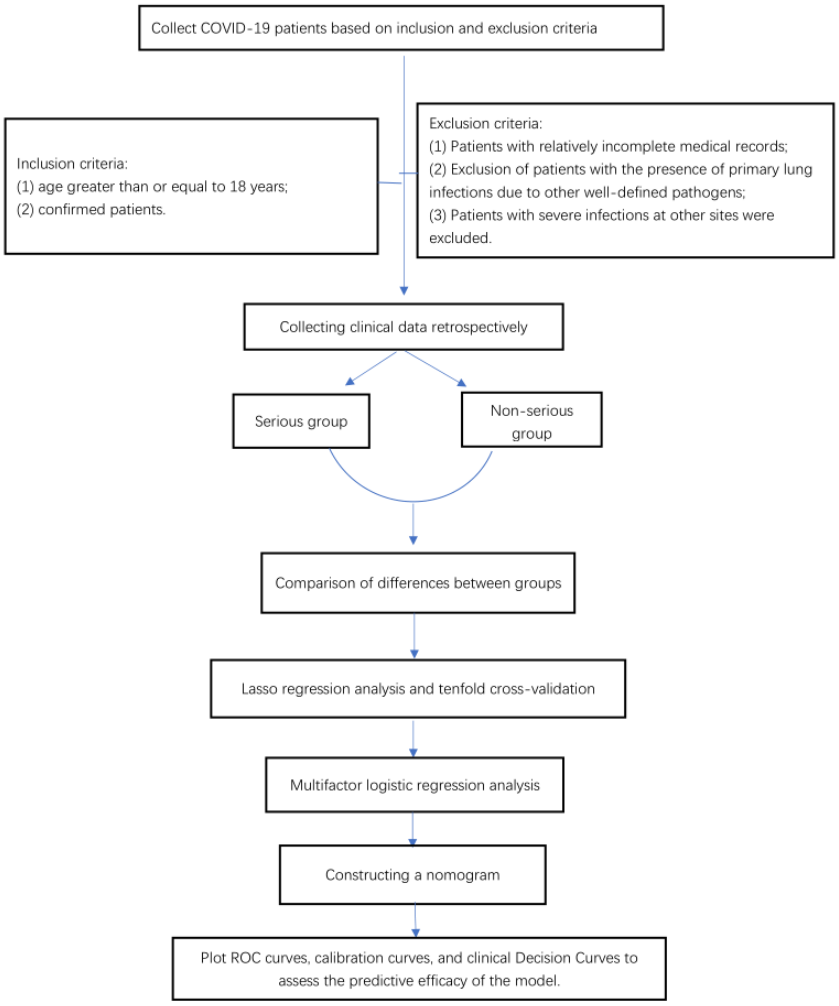

Figure 1 shows the overall flow of this study.

Materials and Methods

Study population

The Institutional Review Committee of the Affiliated Hospital of North China University of Science and Technology has approved this study. The research followed the 1964 Declaration of Helsinki principles and its later amendments or similar ethical standards. Due to the retrospective nature of our study, the Institutional Review Committee of the Affiliated Hospital of North China University of Science and Technology waived informed consent. Measures were implemented to ensure the confidentiality and anonymity of the data collected by appropriate privacy protection protocols.

Patients with novel coronavirus infections who were hospitalized at the Affiliated Hospital of North China University of Science and Technology from 2023.01 to 2023.05 were retrospectively selected for the study. Those diagnosed as severe or critical during hospitalization were chosen as the case group (severe group). Those non-serious during hospitalization (mild or medium cases) were selected as the control group (non-serious group).

According to the Chinese diagnostic and treatment guidelines-Diagnostic and Treatment Guidelines for COVID-19(Trial 10th edition):

Diagnostic Criteria: (1) There are clinical manifestations associated with COVID-19. (2) Have one or more of the following pathogenic and serological findings: Positive nucleic acid test for SARS-CoV-2; Positive test for SARS-CoV-2 antigen; Positive isolation and culture of SARS-CoV-2; SARS-CoV-2-specific IgG antibody levels are 4-fold or more elevated in the recovery phase than in the acute phase.

Clinical Classification:(1) Mild cases: Upper respiratory tract infection is the primary manifestation, with dry throat, sore throat, cough, fever, etc. (2) Medium cases: Persistent high fever for rate (R) <30 breaths/minute and oxygen saturation COVID-19 is seen on imaging. (3) Severe cases: Adults with any of the following that cannot be explained by causes other than COVID-19: Shortness of breath with RR[?] 30 beats/minute;Oxygen saturation [?]93% on air inhalation at rest;Arterial partial pressure of oxygen (PaO2)/oxygen concentration (FiO2) [?] 300 mmHg (1 mmHg = 0.133 kPa), and PaO2 / FiO2 should be corrected for high altitude (more than 1,000 meters above sea level) according to the following formula: PaO2/FiO2×[760/atmospheric pressure (mmHg)];Progressive worsening of clinical symptoms;Lung imaging shows considerable progression of meet one of the following conditions:Respiratory failure, and requires mechanical ventilation;go into shock;Combined with other organ failure requiring ICU monitoring and treatment.

Inclusion criteria: (1) age greater than or equal to 18 years old; (2)confirmed patients who met the diagnostic criteria.

Exclusion criteria: (1) patients with relatively incomplete medical records; (2) patients with the presence of primary lung infections caused by other well-defined pathogens were excluded; (3) patients with severe infections at other sites were excluded.

The above clinical diagnosis and clinical classification are combined with the clinical judgment of senior physicians.

Research method

Patients eligible for the study were included through inclusion and exclusion criteria, and clinical data on patients were collected:(1) General clinical information: gender, age, smoking history, alcohol consumption history, vital signs, and clinical classification;(2) Disease history: combining disease categories <2 and [?]2. category 1 (cerebrovascular disease, coronary heart disease, diabetes, hypertension and hyperlipidemia), category 2 (chronic lung disease: tuberculosis, bronchiectasis, COPD, etc.), category 3 (others: liver and kidney disease, benign haematological disorders, surgical traumas, etc.), and category 4 Immunocompromised (malignant neoplasms, immunosuppressant drug use);(3)Laboratory tests: Blood routine: white blood cell count (WBC), neutrophil count (NEU), lymphocyte count (LYM), platelet count (PLT), haemoglobin (HBG); Blood biochemistry: Liver function: aspartate aminotransferase (AST), alanine aminotransferase (ALT), ALB, renal function: serum creatinine (sCr), blood urea nitrogen (BUN), myocardial enzymes: myoglobin (MYO), creatine kinase (CK), creatine kinase isoenzyme (CK-MB), LDH; Inflammatory markers: procalcitonin (PCT), CRP, erythrocyte sedimentation rate (ESR), interleukin-6 (IL-6); Coagulation series: Fibrinogen (FIB), D-dimer (D-Dimer) (first fasting blood test results after diagnosis);The patients with COVID-19 were then divided into severe and non-severe groups;Comparison of data between groups was performed to analyze the clinical characteristics of patients;The variables were screened using lasso regression analysis and 10-fold cross-validation. The screened variables were subject nomograms constructed by integrating the independent risk factors and the predictive efficacy of ed to multifactorial logistic regression analysis to screen

the independent risk factors of patients with severe or critical illnesses. The model was evaluated by plotting the ROC curve. Based on the maximum value of the Jordon index, the best critical value was selected, the confusion matrix of the prediction model was plotted, the sensitivity and specificity were calculated, and the Kappa value was calculated to evaluate the consistency of the model and plotted using calibration curve analysis combined with the Hosmer-Lemeshow test to assess the model's accuracy and by DCA to evaluate its clinical utility.

Statistical analysis

Statistical analyses were performed using SPSS software (26.0) and R software (version 4.2.3). Multiple interpolations using the R language 'Mice' package to deal with missing data. Measurement data were tested for normality and expressed as 'mean +- standard deviation' for normal or approximately normal distribution and as 'median (interquartile spacing)' for markedly skewed distribution. Qualitative variables are presented as frequencies and percentages (%). Compositional comparisons between groups for count data were performed using the chi-square test; comparisons of means between groups for measurement data were performed using the t-test for independent samples; and comparisons of medians were performed using the nonparametric rank-sum test. A two-sided test was used, and a difference of $P<0.05$ was considered statistically significant. Lasso regression analysis and 10-fold cross-validation to filter variables using the R package 'glmnet.' Multi-factor logistic regression analysis and construction of regression models and line graphs using the R language 'rms' package. ROC curves are plotted using the R package 'pROC,' calibration curves are plotted using the rms package, and DCA is plotted using the 'rmda' package.

Results

Patient characteristics

A total of 410 eligible patients were enrolled, including 132 critically ill patients and 278 non-critically ill patients, including 219 males and 191 females, with a median age of 68 years. The results showed differences in age, Coronary heart disease, Hypertension, Comorbidities, Disorders of Consciousness, temperature(T), R, SBP, WBC, NEU, LMY, HGB, ALT, AST, ALB, BUN, MYO, CK-MB, LDH, FIB, D-dimer, ESR, CRP, PCT, and IL-6 between the two groups ($p < 0.05$). In contrast, gender, Smoking, Drinking, Diabetes, Pulse, Diastolic blood pressure(DBP), Platelets(P), sCr, and CK were not ($p > 0.05$). Intergroup comparisons of patients with COVID-19 are shown in Table l.

Table 1 Comparison of the data of the two groups of COVID-19 patients

Variables	Total (n = 410)	0 (n = 278)	1 (n = 132)	P
Age, years	68 (59, 76)	66(57, 73)	72 (65, 79)	0.000
Gender, n (%)				0.597
female	191 (46.59)	132 (47.48)	59 (44.70)	
male	219 (53.41)	146 (52.52)	73 (55.30)	
Smoking, n(%)				0.968
NO	335 (81.71)	227 (81.65)	108 (81.82)	
YES	75 (18.29)	51 (18.35)	24 (18.18)	
Drinking, n(%)				0.259
NO	356 (86.83)	245 (88.13)	111 (84.09)	
YES	54 (13.17)	33 (11.87)	21 (15.91)	
Coronary heart disease, n(%)				0.014
NO	305 (74.39)	217 (78.06)	88 (66.67)	
YES	105 (25.61)	61 (21.94)	44 (33.33)	
Diabetes, n(%)				0.055
NO	310 (75.61)	218 (78.42)	92 (69.70)	
YES	100 (24.39)	60 (21.58)	40 (30.30)	
Hypertension, n(%)				0.015

NO	231 (56.34)	168 (60.43)	63 (47.73)	
YES	179 (43.66)	110 (39.57)	69 (52.27)	
Comorbidities, n(%)				0.044?¿?
2	219 (53.41)	158 (56.83)	61 (46.21)	
¿2	191 (46.59)	120 (43.17)	71 (53.79)	
Disorders of Consciousness, n(%)				0.000
NO	399 (97.32)	277 (99.64)	122 (92.42)	
YES	11 (2.68)	1 (0.36)	10 (7.58)	
T	36.7 (36.4, 37.2)	36.6 (36.4, 37.1)	36.7 (36.5, 37.7)	0.036
P	86 ± 16	85 ± 13	88 ± 20	0.075
R	21 ± 2	20 ± 1	22 ± 3	0.000
SBP, mmHg	131 ± 20	128 ± 18	137 ± 22	0.000
DBP, mmHg	79 ± 10	79 ± 10	78 ± 11	0.251
WBC(x 10^9)	5.90 (4.32, 8.30)	5.40 (4.20, 7.50)	7.50 (5.40, 10.38)	0.000
NEU(x 10^9)	4.13 (2.70, 6.35)	3.68 (2.50, 5.18)	5.53 (3.83, 8.59)	0.000
LYM(x 10^9)	1.00 (0.66, 1.51)	1.13 (0.81, 1.63)	0.73 (0.49, 1.16)	0.000
PLT(x 10^9)	186.00(130.00, 246.75)	192.50(137.00, 254.50)	182.50(119.75, 242.25)	0.057
HGB(g/L)	127.00(115.00, 139.00)	129.00(118.00, 139.75)	122.50(110.75, 135.00)	0.007
ALT(U/L)	23.00 (15.00, 35.00)	22.00(14.00,33.00)	25.50(16.00,40.25)	0.034
AST(U/L)	25.00 (19.00, 32.00)	23.50(18.00,30.00)	27.00(22.00,40.25)	0.000
ALB(g/L)	38.60 (34.90, 41.30)	39.70(36.92,42.40)	34.85(31.23,38.62)	0.000
Scr(umol/L)	68.00 (56.00, 84.75)	68.00(57.00,84.00)	65.50(55.00,87.75)	0.684
BUN(mmol/L)	5.52 (4.20, 6.95)	5.33 (4.08, 6.73)	5.84 (4.74, 7.78)	0.000
MYO(ug/L)	21.00 (13.00, 40.00)	20.00(12.00,32.00)	31.50(16.00,53.25)	0.000
CK(U/L)	62.00 (42.25, 95.00)	63.00(46.00,90.75)	56.50(33.75, 117.50)	0.203
CK-MB(U/L)	11.00 (9.00, 15.00)	11.00 (8.00, 14.00)	12.00 (9.00, 17.25)	0.000
LDH(U/L)	246.50(201.00, 311.75)	227.00(188.25, 268.00)	318.00(247.25, 417.00)	0.000
FIB(g/L)	4.91 (4.05, 5.83)	4.62 (3.91, 5.40)	5.50 (4.44, 6.49)	0.000
D-dimer(ng/ml)	531.00(292.00, 1168.50)	427.50(236.25, 838.50)	841.50(475.50, 1720.75)	0.000
ESR(mm/h)	40.00 (24.00, 63.00)	35.00(20.25,57.00)	56.00(35.75,79.00)	0.000
CRP(mg/ L)	29.95 (11.15, 58.80)	21.70 (7.25, 39.55)	50.90(28.18,95.15)	0.000
PCT(mg/mL)	0.08 (0.04, 0.10)	0.07 (0.04, 0.10)	0.10 (0.05, 0.24)	0.000
IL-6(pg/m)	16.87 (5.00, 41.90)	11.70 (5.00, 30.35)	31.73 (9.59, 58.83)	0.000

LASSO regression and tenfold cross-validation

LASSO regression (Least Absolute Shrinkage and Selection Operator Regression) is a regularisation technique for linear regression models proposed by Robert Tibshirani in 1996. LASSO regression through the addition of L1 paradigm penalty term in the loss function, both to achieve the model parameters of the Shrinkage but also to achieve the variable selection, the regression coefficients of some unnecessary variables can be compressed to zero and then eliminated from the model, in the high-dimensional data analysis, can effectively reduce the data dimensions, to solve the multicollinearity. On the other hand, cross-validation is used to fit a model using the training set by dividing the dataset into a training set and a validation set and evaluating the model performance on the validation set. The λ that gives the best performance on the validation set is selected. In this study, 6 variables were screened from the 31 variables collected based on the non-zero coefficients calculated from lasso regression analysis (Figure 2). The 6 variables include disorders of consciousness, R, SBP, ALB, LDH, and CRP.

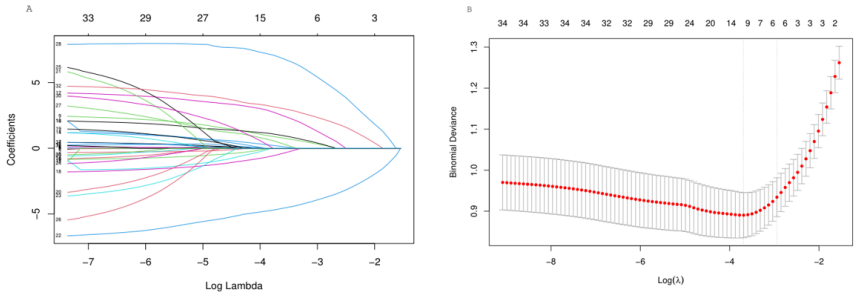

Figure 2 Lasso regression analysis and tenfold cross-validation

A Fig Lasso regression coefficient path plots, vertical coordinates represent coefficients, and horizontal coordinates represent log(λ). B Fig 10-fold cross-validation curves for lasso regression, vertical coordinates represent likelihood deviations, horizontal coordinates represent log(λ), the left dashed line represents λ with the minor deviation (lambda-min), and the right dashed line represents one standard error to the right of the smallest λ (lambda- 1-SE) (chosen for this study).

Screening for independent risk factors by multifactorial logistic regression analysis

The multifactorial logistic regression analysis was performed to determine whether the COVID-19 patients were serious (yes=1, no=0) as the dependent variable and the factors with significance in the above Lasso regression analysis and ten-fold cross-validation as the independent variables. The results showed that R, SBP, ALB, LDH, and CRP were independent risk factors for serious COVID-19 patients (P<0.05). This means that the above factors can increase the risk of serious COVID-19 patients and can be used as predictors of whether a COVID-19 patient is serious. In contrast, Disorders of Consciousness were not an independent risk factor for serious COVID-19 patients. (P>0.05) (table 2).

Table 2 Results of the multifactorial analysis of patients with serious COVID-19

Variables	β	SE	Z	P	OR (95%CI)
Intercept	-4.69	2.37	-1.98	0.048	0.01 (0.00 ~ 0.96)
Disorders of consciousness					
0					1.00 (Reference)
1	2.00	1.15	1.74	0.082	7.40 (0.77 ~ 70.81)
R	0.25	0.09	2.81	0.005	1.28 (1.08 ~ 1.53)
SB	0.02	0.01	2.46	0.014	1.02 (1.01 ~ 1.03)
ALB	-0.16	0.03	-5.01	<.001	0.85 (0.80 ~ 0.91)
LDH	0.01	0.00	4.60	<.001	1.01 (1.01 ~ 1.01)
CRP	0.01	0.00	2.95	0.003	1.01 (1.01 ~ 1.02)

Construction of a risk prediction nomogram for patients with serious COVID-19

Logistic regression models were constructed for the predictors identified by lasso regression analysis and multifactorial analysis (R, SBP, ALB, LDH, and CRP) and plotted nomogram used to predict the risk of a COVID-19 patient being severe or critical cases (Figure 3). According to a patient's vital signs and laboratory indicators, the corresponding Points of R, SBP, ALB, LDH, and CRP were calculated on the points of the nomogram. Then, the total score was obtained by summing the points of the five indicators. The prediction probability of the patient's severe current condition is received by the Total Points line at the bottom of the nomogram and the corresponding Serious Risk line.

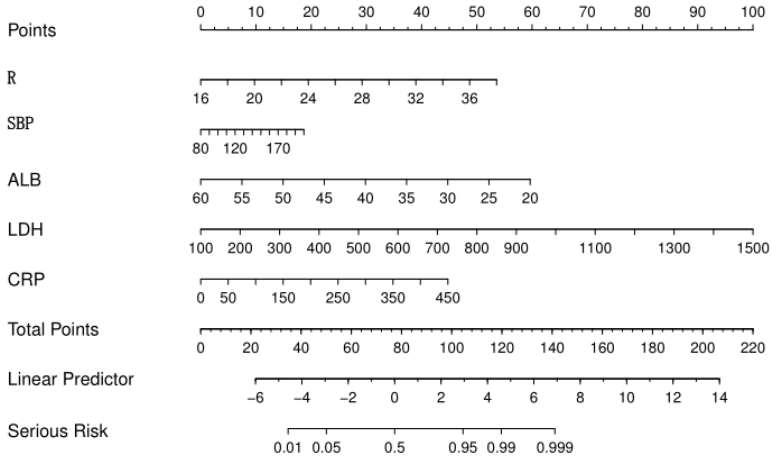

Figure 3 nomogram of predicted risk of severe or critical cases in COVID-19 patients

Evaluation of the predictive effectiveness of nomogram

1)ROC curve analysis of a predictive model for serious COVID-19 ROC curves were plotted by selecting a myriad of Cutoff values and calculating the sensitivity and specificity of the model predictions at the corresponding critical values. It can be used to evaluate the discriminatory power of the model, and the results showed that the AUC value of the column-line plot for discriminating between serious COVID-19 and non-serious COVID-19 was 0.86 (95% CI: 0.82 ˜ 0.90), which demonstrated the excellent discriminatory power of the column-line plot model (Figure 4).

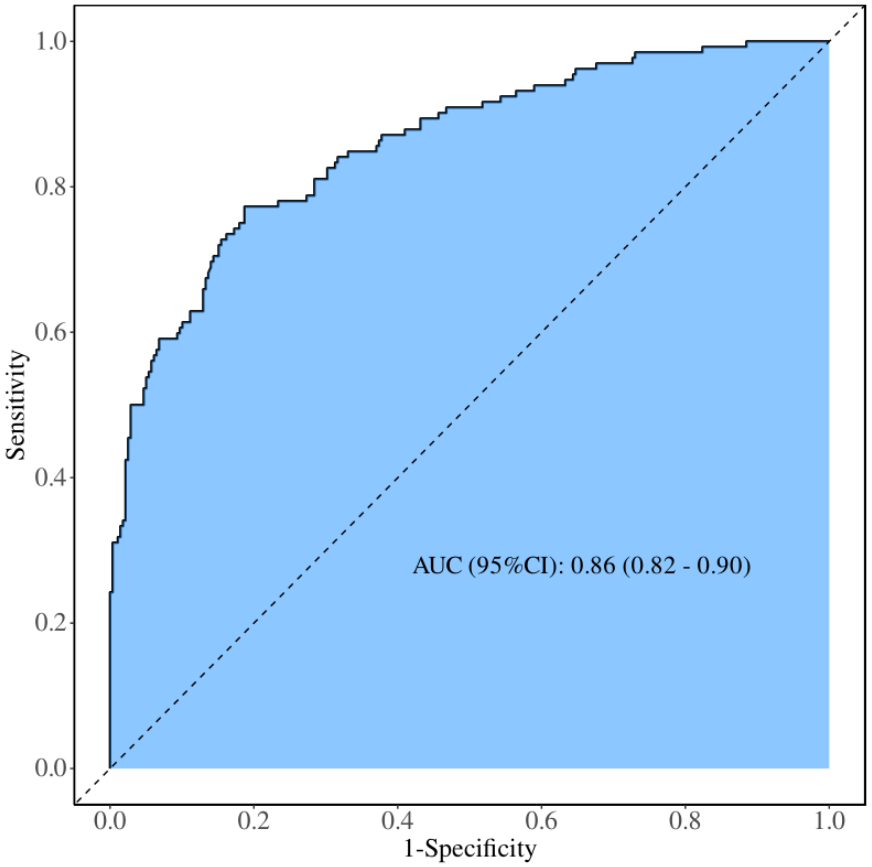

Figure 4 ROC curve analysis of the predictive model for serious COVID-19

2)Valuation of the effectiveness of the nomogram model for predicting serious COVID-19

Based on the ROC curve, we obtained the Youden index, and based on the maximum value of the Youden index, we determined that the critical value for the prediction of the nomogram model was total score = 70.7 points, i.e., when the patient's total score of the nomogram model was [?] 70.7 points, the clinical doctors could predict that the patient with COVID-19 was serious. With this as the baseline, the change in the condition could be dynamically assessed in the hospitalization process. The predicted patient outcomes are shown in Table 3 and Figure 5 when we use the criterion. The results show that the nomogram model has high sensitivity and specificity (sensitivity 77.3%, specificity 81.3%) and good agreement between the predicted situation and the actual occurrence (Kappa value 0.541). This indicates that the nomogram model can distinguish serious and non-serious COVID-19 patients and is worthy of clinical promotion and use.

Table 3 Evaluation of the effectiveness of the nomogram model in predicting serious COVID-19

True results	Projected results Non-serious(n)	sensitivity (%) Serious(n)	specificity (%)	Youden index	*Kappa value*	
Non-serious(n)	224	30	77.3	80.6	0.586	0.552
Serious(n)	54	102				

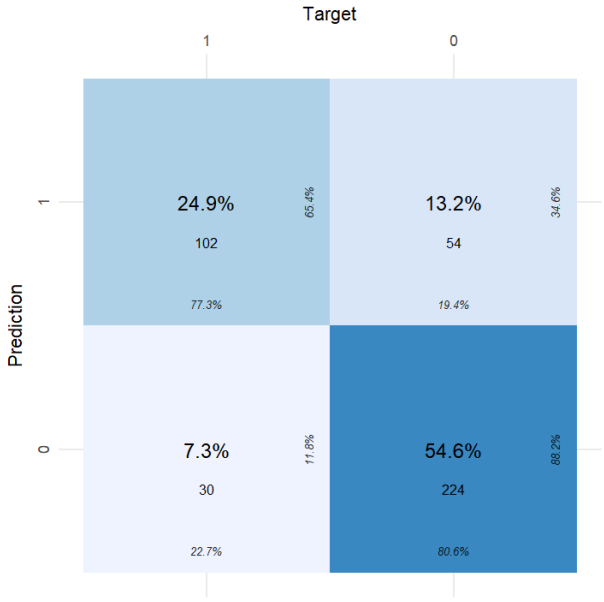

Figure 5 Confusion matrix predicted by the model

3)Calibration curve analysis

To more accurately evaluate the prediction ability of the nomogram in the risk of serious illness of COVID-19 patients and avoid overestimation of the accuracy of the nomogram, the Bootstrap self-sampling method was used to conduct 1000 repeated samples on the nomogram to reduce over-fitting bias. The calibration curve is then plotted (Figure 6). In the figure, the X-axis represents the predicted probability of serious COVID-19 patients; the Y-axis represents the actual observed probability of serious COVID-19 patients; the black dashed line represents the ideal curve, the orange line represents the prediction curve that has not been calibrated, and the blue line represents the calibrated prediction curve. The calibrated curve in the figure is very close to the ideal curve, and the Hosmer-Lemeshow test results show P=0.958, all of which indicate that the model consistently predicts the risk of serious COVID-19 patients and the actual risk.

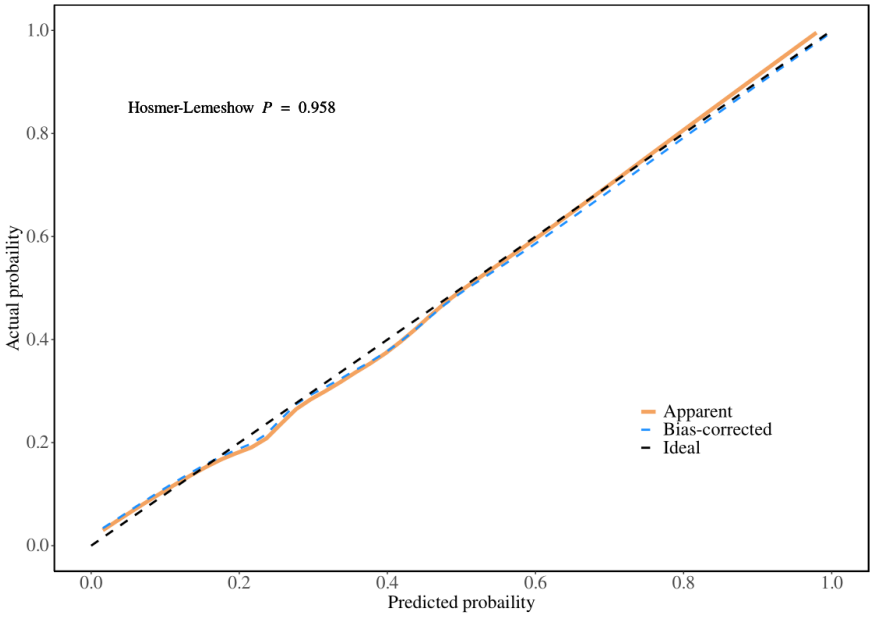

Figure 6 Calibration curve analysis of serious COVID-19 prediction model

4)Clinical decision curve analysis

The DCA curves graphically show the value of the clinical application of the nomogram model. The vertical coordinate represents the net benefit, the horizontal axis represents the threshold probability, the grey line assumes all COVID-19 patients are serious, the black line assumes that all COVID-19 patients are non-serious, and the red line represents the model. The curves generated indicate that when the threshold probability is approximately 1.5% to 95%, this model to identify patients with serious COVID-19 generates a greater net benefit than if all COVID-19 patients are treated as serious or all are not treated as serious (Figure 7).

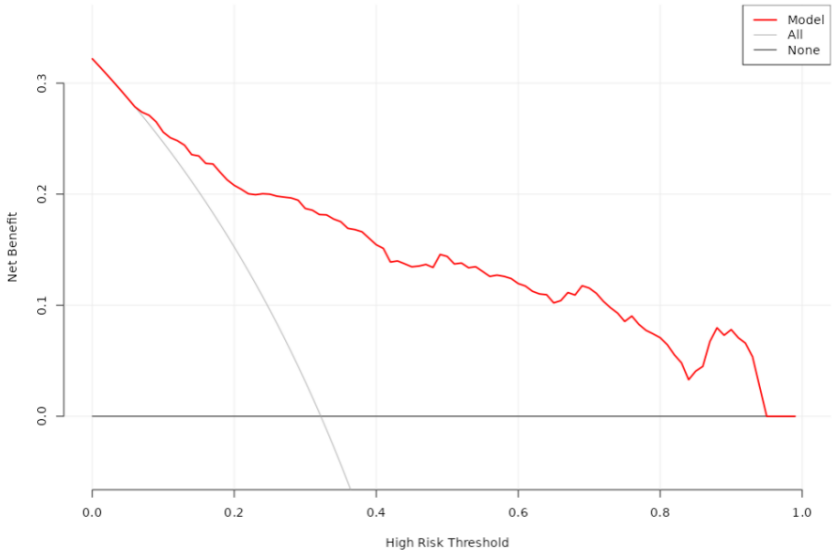

Figure 7 Decision curve analysis of serious COVID-19 prediction model

Discussion

Since January 8, 2023, China has officially implemented the " Class B and B tube " policy for patients with COVID-19. Given the high infectivity of the SARS-CoV-2 virus, which may lead to seasonal epidemic peaks, the healthcare system will undoubtedly face more severe challenges. Whether patients are admitted to the hospital with COVID-19 infection in the community or those infected in the hospital, dynamic assessment of their condition is critical. This study revealed that R, SBP, LDH levels, decreased ALB, and CPR were independent risk factors for predicting severe or critical COVID-19 patients. The five indicators were combined to construct a clinical diagnostic prediction model. The model is composed of common clinical laboratory test indicators and two crucial vital signs and finally meets the original intention of the model - simple and easy to obtain, but also shows good diagnostic efficiency.

Compared with other prediction models, the model constructed in this study is more concise and clear when selecting variables. Unlike those studies that rely on chest imaging examination [12-13], imaging examination is expensive, and some patients may be unable to accept it due to economic reasons, making dynamic monitoring more difficult. Secondly, chest CT imaging has a certain lag in reflecting the current condition of patients. In addition, for some patients with other serious diseases, leaving the ward for CT examination will be restricted. In addition, the data of this study came from 2023, that is, after the prevalence of Omicron variants and the implementation of the "Class B and B tube" policy in China. Therefore, compared with the studies in the early phase of the pandemic, the data in this study are more closely related to the current prevalence of Omicron variants, thus making its findings more relevant and applicable.

SARS-CoV-2 virus can cause a variety of clinical manifestations, ranging from upper respiratory discomfort to pneumonia and even multiple organ failure [14]. This pathological mechanism of multi-organ failure may be related to the unbalanced immune response during COVID-19. Therefore, in patients with COVID-19, the levels of a series of biomarkers will change, some of which are critical for risk stratification, diagnosis, and prognosis assessment of COVID-19 [15]. LDH, one of the key enzymes that catalyze the conversion between pyruvate and lactic acid, plays a central role in the anaerobic metabolism of glucose. This process is critical when the oxygen supply is insufficient or limited. As an essential enzyme in the glycolysis pathway, LDH is widely distributed in myocardium, liver, kidney, and lung tissues, and its concentration in tissues is much higher than that in serum. When the lung tissue suffers from hypoxic necrosis, a large amount of LDH will be released into the blood, resulting in a sharp rise in serum LDH levels [16]. Consistent with the conclusions of this study, serum LDH has been proven to be a sensitive biomarker for assessing the dynamic changes of COVID-19 patients [16-18]. ALB is a harmful acute-phase protein synthesized by the liver and has many vital functions: it is responsible for transporting various substances and maintaining normal plasma colloid osmotic pressure [19]. Previous studies have shown that hypoalbuminemia is a practical, dose-dependent, independent predictor of adverse outcomes. Each 10 g/L decrease in serum albumin concentration was associated with a 137% increase in mortality, 89% increase in morbidity, 28% and 71% increase in intensive care unit and hospital stay, respectively, and 66% increase in resource utilization [20]. In addition, a retrospective cohort study of COVID-19 patients in Spain showed an association between low albumin (ALB) levels and poor prognosis [21].

Similarly, Beimdiek et al[22] concluded in agreement with the present study that ALB levels were lower in COVID-19 critically ill patients. It may be that during infection, the distribution rate of serum albumin from intravascular to extravascular increases significantly, as does the decomposition rate [23]. CRP is a nonspecific IL-6-induced acute phase reactant in the liver. Clinically, it is a biomarker for different inflammatory and infectious diseases. Elevated CRP levels are directly related to inflammation and disease severity [24]. As such, it is an essential biomarker for assessing the severity of COVID-19. Other studies have also confirmed this conclusion [25-26].

In the conclusion of this study, the proportion of patients with pre-existing hypertension in the serious group was 52.27%, while that in the non-serious group was 39.57%, showing a statistical difference but not

as an independent risk factor for serious COVID-19 patients. However, elevated SBP is an independent risk factor for serious COVID-19 patients. Hypertension is a risk factor for increased COVID-19-related deaths [27]. However, the independent role of hypertension remains controversial, as hypertension is often associated with age and cardiovascular disease, which can also make COVID-19 more severe. In addition, antihypertensive drugs, angiotensin conversion inhibitors (ACEI), and angiotensin receptor blockers (ARBs) may be associated with the condition of COVID-19 patients. ACE2 is a crucial binding receptor that facilitates the entry of SARS-CoV-2 into the organism, and renin-angiotensin system (RAS) blockers may up-regulate ACE2. Therefore, these drugs may contribute to the increased risk of SARS-CoV-2 infection and COVID-19 progression, but this idea is controversial [28]. However, elevated SBP levels may be a marker of pre-existing hypertension-mediated subclinical organ damage (HMOD) [29]. This may be due to inadequate treatment or control, so elevated SBP is a risk factor.

Our study selected three serum biomarkers: LDH, CRP, and ALB. Combined with a direct vital sign that reflects lung conditions, R, and SBP level, a simple prediction model was constructed to evaluate the severity of COVID-19 patients' illness with good sensitivity and specificity.

We have to admit that our study has certain limitations. First, due to the restriction of retrospective study, the collection of variables, such as BMI, vaccination status, peripheral blood oxygen status, etc., is incomplete, as is the collection time of blood indicators. Prospective studies can be conducted subsequently to dynamically observe and compare the blood test results at specific time points before and after infection. Secondly, our study is single-center, and cases are not collected for external verification to verify the model's generalization ability. We can develop web calculators for clinical use after our nomogram is prospectively or externally validated.

Conclusions

When the R, SBP, LDH, and CRP are increased, and ALB are reduced, the risk of being serious COVID-19 is high. Combined with R, SBP, ALB, LDH, and CRP, a nomogram prediction model was established to predict the risk of serious COVID-19 patients, and this model has good prediction ability.

Data availability statement

Reasonable requests for data from this study can be obtained by contacting the corresponding author.

Funding

This work was supported by the Medical Applicable Technology Tracking Project of Hebei Provincial Health Commission (GZ20250080).

Competing interests

The authors declare no competing interests.

Ethics declarations

The Institutional Review Committee of the Affiliated Hospital of North China University of Science and Technology has approved this study. The research followed the 1964 Declaration of Helsinki principles and its later amendments or similar ethical standards. The Institutional Review Committee of the Affiliated Hospital of North China University of Science and Technology waived informed consent due to the retrospective nature of our study. Measures were implemented to ensure the confidentiality and anonymity of the data collected by appropriate privacy protection protocols.

Acknowledgments

We thank the other members of the North China University of Technology Hospital and Shijiazhuang Third Hospital for their constructive criticism.

Reference

[1] Wang, Dawei et al. "Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China." JAMA vol. 323,11 (2020): 1061-1069. doi:10.1001/jama.2020.1585. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y, Li Y, Wang X, Peng Z. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA. 2020 Mar 17;323(11):1061-1069.

[2] Zhu, Na et al. "A Novel Coronavirus from Patients with Pneumonia in China, 2019." *The New England journal of medicine* vol. 382,8 (2020): 727-733. doi:10.1056/NEJMoa2001017.

[3] Bálint, Gábor et al. "Omicron: increased transmissibility and decreased pathogenicity." Signal transduction and targeted therapy vol. 7,1 151. May 7. 2022, doi:10.1038/s41392-022-01009-8.

[4] Cheung, Pak-Hin Hinson et al. "Lessons learned from the fifth wave of COVID-19 in Hong Kong in early 2022." Emerging microbes & infections vol. 11,1 (2022): 1072-1078. doi:10.1080/22221751.2022.2060137.

[5] Hippisley-Cox, Julia et al. "Risk prediction of covid-19 related death or hospital admission in adults testing positive for SARS-CoV-2 infection during the omicron wave in England (QCOVID4): cohort study." BMJ (Clinical research ed.) vol. 381 e072976. June 21. 2023, doi:10.1136/bmj-2022-072976.

[6] Baker, James B et al. "Development of a Novel Clinical Risk Score for COVID-19 Infections." The American journal of medicine vol. 136,12 (2023): 1169-1178.e7. doi:10.1016/j.amjmed.2023.08.016.

[7] Weng, Zhihong et al. "ANDC: an early warning score to predict mortality risk for patients with Coronavirus Disease 2019." Journal of translational medicine vol. 18,1 328. August 31 2020, doi:10.1186/s12967-020-02505-7.

[8] Moon, Hui Jeong et al. "Prediction of COVID-19-related Mortality and 30-Day and 60-Day Survival Probabilities Using a Nomogram." Journal of Korean medical science vol. 36,35 e248. 6 Sep. 2021, doi:10.3346/jkms.2021.36.e248.

[9] Yang, Tianning et al. "Identifying severe community-acquired pneumonia using radiomics and clinical data: a machine learning approach." Scientific reports vol. 14,1 21884. September 19. 2024, doi:10.1038/s41598-024-72310-5.

[10] Zhu, Xiaofeng et al. "Joint prediction and time estimation of COVID-19 developing severe symptoms using chest CT scan." Medical image analysis vol. 67 (2021): 101824. doi:10.1016/j.media.2020.101824.

[11] Hunter, David J, and Christopher Holmes. "Where Medical Statistics Meets Artificial Intelligence." The New England journal of medicine vol. 389,13 (2023): 1211-1219. doi:10.1056/NEJMra2212850.

[12] Zysman, Maéva et al. "Development and external validation of a prediction model for the transition from mild to moderate or severe form of COVID-19." European radiology vol. 33,12 (2023): 9262-9274. doi:10.1007/s00330-023-09759-x.

[13] Qiu, Jiajun et al. "A Radiomics Signature to Quantitatively Analyze COVID-19-Infected Pulmonary Lesions." Interdisciplinary sciences, computational life sciences vol. 13,1 (2021): 61-72. doi:10.1007/s12539-020-00410-7.

[14] Robba, Chiara et al. "Multiple organ dysfunction in SARS-CoV-2: MODS-CoV-2." Expert review of respiratory medicine vol. 14,9 (2020): 865-868. doi:10.1080/17476348.2020.1778470.

[15] Battaglini, Denise et al. "Laboratory Biomarkers for Diagnosis and Prognosis in COVID-19." Frontiers in immunology vol. 13 857573. 27 Apr. 2022, doi:10.3389/fimmu.2022.857573.

[16] Mu, Tong et al. "Implications of LDH in patients with coronavirus disease 2019 pneumonia." Frontiers in cellular and infection microbiology vol. 13 1180187. 30 Oct. 2023, doi:10.3389/fcimb.2023.1180187.

[17] Wu, Mei-Ying et al. "Clinical evaluation of potential usefulness of serum lactate dehydrogenase (LDH) in 2019 novel coronavirus (COVID-19) pneumonia." Respiratory research vol. 21,1 171. July 6 2020,

doi:10.1186/s12931-020-01427-8.

[18] Chen, Zhufeng et al. "Clinical laboratory evaluation of COVID-19." Clinica chimica acta; international journal of clinical chemistry vol. 519 (2021): 172-182. doi:10.1016/j.cca.2021.04.022.

[19] Jellinge, Marlene Ersgaard et al. "Hypoalbuminemia is a strong predictor of 30-day all-cause mortality in acutely admitted medical patients: a prospective, observational, cohort study." PloS one vol. 9,8 e105983. August 22 2014, doi:10.1371/journal.pone.0105983.

[20] Vincent, Jean-Louis et al. "Hypoalbuminemia in acute illness: is there a rationale for intervention? A meta-analysis of cohort studies and controlled trials." Annals of surgery vol. 237,3 (2003): 319-34. doi:10.1097/01.SLA.0000055547.93484.87.

[21] de la Rica, Roberto et al. "Low Albumin Levels Are Associated with Poorer Outcomes in a Case Series of COVID-19 Patients in Spain: A Retrospective Cohort Study." Microorganisms vol. 8,8 1106. July 24 2020, doi:10.3390/microorganisms8081106.

[22] Beimdiek, Julia et al. "Plasma markers of COVID-19 severity: a pilot study." Respiratory research vol. 23,1 343. December 13 2022, doi:10.1186/s12931-022-02272-7.

[23] Li, Wei-qin et al. "Albumin kinetics in patients with severe sepsis. "Zhonghua wai ke za zhi [Chinese journal of surgery] vol. 41,6 (2003): 423-6.

[24] Malik, Preeti et al. "Biomarkers and outcomes of COVID-19 hospitalizations: systematic review and meta-analysis." BMJ evidence-based medicine vol. 26,3 (2021): 107-108. doi:10.1136/bmjebm-2020-111536.

[25] Zhang, Jin-Jin et al. "Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China." Allergy vol. 75,7 (2020): 1730-1741. doi:10.1111/all.14238.

[26] Ali, Ayad M et al. "Serum troponin, D-dimer, and CRP level in severe coronavirus (COVID-19) patients." Immunity, inflammation and disease vol. 10,3 (2022): e582. doi:10.1002/iid3.582.

[27] Semenzato, Laura et al. "Chronic diseases, health conditions and risk of COVID-19-related hospitalization and in-hospital mortality during the first wave of the epidemic in France: a cohort study of 66 million people." The Lancet regional health. Europe vol. 8 (2021): 100158. doi:10.1016/j.lanepe.2021.100158.

[28] Kreutz, Reinhold et al. "Hypertension, the renin-angiotensin system, and the risk of lower respiratory tract infections and lung injury: implications for COVID-19." *Cardiovascular research* vol. 116,10 (2020): 1688-1699. doi:10.1093/cvr/cvaa097.

[29] Savoia, Carmine et al. "Hypertension, a Moving Target in COVID-19: Current Views and Perspectives." *Circulation research* vol. 128,7 (2021): 1062-1079. doi:10.1161/CIRCRESAHA.121.318054.

legend

Figure 1 The overall process of research

Figure 2 Lasso regression analysis and tenfold cross-validation

A Fig Lasso regression coefficient path plots, vertical coordinates represent coefficients, and horizontal coordinates represent log(λ).

B Fig 10-fold cross-validation curves for lasso regression, vertical coordinates represent likelihood deviations, horizontal coordinates represent log(λ), the left dashed line represents λ with the minor deviation (lambda-min), and the right dashed line represents one standard error to the right of the smallest λ (lambda- 1-SE) (chosen for this study).

Figure 3 nomogram of predicted risk of severe or critical cases in COVID-19 patients.

Figure 4 ROC curve analysis of the predictive model for serious COVID-19.

Figure 5 Confusion matrix predicted by the model.

Figure 6 Calibration curve analysis of serious COVID-19 prediction model.

Figure 7 Decision curve analysis of serious COVID-19 prediction model.