scientific reports

Check for updates

OPEN

A predictive model to explore risk factors for severe COVID-19

Fen-Hong Qian[1,2✉], Yu Cao[1,2], Yu-Xue Liu[1], Jing Huang[1] & Rong-Hao Zhu[1]

With the rapid spread of the novel coronavirus (COVID-19), a sustained global pandemic has emerged. Globally, the cumulative death toll is in the millions. The rising number of COVID-19 infections and deaths has severely impacted the lives of people worldwide, healthcare systems, and economic development. We conducted a retrospective analysis of the characteristics of COVID-19 patients. This analysis includes clinical features upon initial hospital admission, relevant laboratory test results, and imaging findings. We aimed to identify risk factors for severe illness and to construct a predictive model for assessing the risk of severe COVID-19. We collected and analyzed electronic medical records of confirmed COVID-19 patients admitted to the Affiliated Hospital of Jiangsu University (Zhenjiang, China) between December 18, 2022, and February 28, 2023. According to the WHO diagnostic criteria for the novel coronavirus, we divided the patients into two groups: severe and non-severe, and compared their clinical, laboratory, and imaging data. Logistic regression analysis, the least absolute shrinkage and selection operator (LASSO) regression, and receiver operating characteristic (ROC) curve analysis were used to identify the relevant risk factors for severe COVID-19 patients. Patients were divided into a training cohort and a validation cohort. A nomogram model was constructed using the "rms" package in R software. Among the 346 patients, the severe group exhibited significantly higher respiratory rates, breathlessness, altered consciousness, neutrophil-to-lymphocyte ratio (NLR), and lactate dehydrogenase (LDH) levels compared to the non-severe group. Imaging findings indicated that the severe group had a higher proportion of bilateral pulmonary inflammation and ground-glass opacities compared to the non-severe group. NLR and LDH were identified as independent risk factors for severe patients. The diagnostic performance was maximized when NLR, respiratory rate (RR), and LDH were combined. Based on the statistical analysis results, we developed a COVID-19 severity risk prediction model. The total score is calculated by adding up the scores for each of the twelve independent variables. By mapping the total score to the lowest scale, we can estimate the risk of COVID-19 severity. In addition, the calibration plots and DCA analysis showed that the nomogram had better discrimination power for predicting the severity of COVID-19. Our results showed that the development and validation of the predictive nomogram had good predictive value for severe COVID-19.

Abbreviations

COVID-19	The novel coronavirus
ROC	Receiver operating characteristic
NLR	Neutrophil-to-lymphocyte ratio
LDH	Lactate dehydrogenase
RR	Respiratory rate
SE	Standard error
OR	Odds ratio
CI	Confidence interval
GGO	Ground glass opacities
PLR	Platelet-to-lymphocyte ratio
MLR	Monocyte-to-lymphocyte ratio
LMR	Lymphocyte-to-monocyte ratio

[1]Department of Respiratory and Critical Care Medicine, Affiliated Hospital of Jiangsu University, No.438, Jiefang Road, Jingkou District, Zhenjiang, Jiangsu, China. [2]These authors contributed equally: Fen-Hong Qian and Yu Cao. ✉email: zhaoqian604@126.com

MRR	Monocyte-to-red blood cell ratio
NRR	Neutrophil-to-red blood cell ratio
LRR	Lymphocyte-to-red blood cell ratio
SII	Systemic immune-inflammation index
SIRI	Systemic immune response index
WBC	White blood cell count
ALC	Absolute lymphocyte count
ANC	Absolute neutrophil count
PLT	Platelet
ESR	Erythrocyte sedimentation rate
CRP	C-reactive protein
TBIL	Total bilirubin
DBIL	Direct bilirubin
IBIL	Indirect bilirubin
ALT	Alanine aminotransferase
AST	Aspartate transaminase
BUN	Blood urea nitrogen
Cre	Creatinine
TG	Triglycerides
TC	Total cholesterol
PT	Prothrombin time
PT-INR	Prothrombin international normalization ratio
APTT	Activated partial thromboplastin time
TT	Thrombin time
FIB	Fibrinogen
DD	D-dimmer
CK	Creatine kinase
CK-MB	Creatine kinase-MB

Since December 2019, an outbreak of unidentified viral pneumonia occurred in Wuhan. The causative virus was identified as a novel coronavirus distinct from the six known coronaviruses[1]. Compared to other influenza viruses, COVID-19 spreads faster, has a wider reach, and presents more severe symptoms and outcomes. Individuals infected with COVID-19 present a wide range of symptoms, including fever, fatigue, sore throat, dry cough, and more. Among these, fever, dry cough, and pulmonary imaging changes are often common clinical manifestations in COVID-19 patients. The severity of infection can vary, with some people being asymptomatic or non-severe, others developing severe symptoms, and in some cases, it may progress rapidly to cause complications and even be life-threatening. The rising number of COVID-19 infections and deaths has severely impacted the lives of people worldwide, healthcare systems, and economic development. To alleviate the burden on the healthcare system, while providing more precise treatment and minimizing the occurrence of severe cases and fatalities, healthcare professionals need to identify risk factors for severe illness in COVID-19 patients at an early stage and engage in timely and effective disease management. Developing predictive models that incorporate multiple variables or features to assess the risk of severe illness in individuals infected with or post-infection by COVID-19 can assist healthcare providers in managing patients systematically while allocating limited medical resources. Several studies have already developed diagnostic and predictive models for COVID-19. For instance, there are COVID-19 diagnostic prediction models based on symptoms like loss of smell and taste[2], as well as diagnostic models utilizing high-resolution computer tomography scans with deep learning techniques[3]. Machine learning methods have been employed to classify COVID-19 using CT images[4]. However, these predictive models exhibit varying degrees of inadequacy in terms of discriminative power and accuracy[5]. Due to factors such as ethnicity, region, and other unassessed variables, these models unavoidably possess limitations in terms of their applicability.

With an increasing number of predictive models being developed, inflammatory markers are considered one of the key biological indicators for assessing disease severity and play a vital role in diagnosing and evaluating inflammatory conditions[6–9]. Research suggests that the Neutrophil-to-Lymphocyte Ratio (NLR) can be used for the diagnosis and assessment of the severity of COVID-19 in patients[10]. Elevated levels of LDH have been significantly associated with the severity and mortality rates of COVID-19[11]. In chest CT scans, severe COVID-19 patients often exhibit bilateral lung involvement, while non-severe cases are more likely to display ground-glass opacities[12].

Hence, combining patient-specific disease characteristics, laboratory test results, and imaging findings to identify risk factors for severe illness and construct a predictive model for the severity of COVID-19 is of paramount clinical significance. This approach aids in the early identification of severe COVID-19 patients and allows for more proactive treatment strategies.

Materials and methods
Participants
Confirmed cases of COVID-19 admitted to the Affiliated Hospital of Jiangsu University (Zhenjiang, China) between December 18, 2022, and February 28, 2023, were selected as study subjects according to the following inclusion and exclusion criteria. All eligible patients met the following: (i) Real-time fluorescent RT-PCR detection of the novel coronavirus nucleic acid was positive at Zhenjiang Disease Control and Prevention Center,

various levels of hospitals in Zhenjiang, and Jiangsu University Affiliated Hospital, (ii) ≥ 18 years old, (iii) Positive patients with radiological examination results, (iv) Patients who have not received treatment for novel coronavirus infection before their visit, (v) The latest peripheral blood sample results were collected from fasting patients before treatment. All eligible patients should exclude the following: (i) Accompanied by acute infections in other parts (acute pancreatitis, acute cholecystitis, liver abscess, etc.), (ii) Infection of the lungs with other known pathogens, (iii) Pregnant, (iv) Recently used antiplatelet, anticoagulant drugs, immunosuppressants, or other conditions that researchers believe may affect the study results, (v) Patients with missing baseline data or those transferred to other designated hospitals during hospitalization.

According to the inclusion and exclusion criteria, a total of 346 patients were included in this study. Based on the WHO diagnostic criteria, all study subjects were further divided into non-severe and severe groups (Fig. 1).

This study was approved and registered by the Ethics Committee of the Affiliated Hospital of Jiangsu University (Approval number: KY2023K1005). In this retrospective study, all participants provided informed consent. We protected the confidentiality of patient information by recognizing and minimizing data collection. The collected data were anonymized— to the greatest degree to ensure the confidentiality of patient information intact.

Research method

Baseline characteristics, laboratory data, and radiological results for each eligible patient were obtained from the electronic medical records system of Jiangsu University Affiliated Hospital. Electronic medical records for each patient were extracted and analyzed by two independent researchers using standardized data collection forms. The present study was approved by The Ethical Review Committee of Jiangsu University Affiliated Hospital (Zhenjiang, China). Approval number: KY2023K1005. All patients provided informed consent. All experiments were performed in accordance with relevant guidelines and regulations.

Clinical baseline data mainly included the following information for each patient: (i) General information: age (years), gender (male/female). (ii) Smoking history (yes/no), alcohol consumption history (yes/no). (iii) Past medical history: presence or absence of comorbidities such as hypertension, diabetes, heart diseases, nerve system disease, chronic lung disease, liver and kidney disease, and cancer. Laboratory data include complete blood count, biochemical parameters, myocardial enzyme profile, and coagulation function. The latest peripheral blood samples were collected from patients with an empty stomach in the early morning before diagnosis and any treatment. Blood cell analysis was performed using the Sysmex XN3000 automated hematology analyzer (Sysmex Corporation, Japan). Biochemical parameters and blood myocardial enzyme spectrum were measured using the Beckman AU5800 fully automated biochemistry analyzer (Beckman Coulter, Inc.). The coagulation function was analyzed using the automated coagulation analyzer Sekisui CP3000 (Sekisui Medical Corporation, Japan). Chest imaging is done using computed tomography (SOMATOM Definition, Germany). The normal ranges for all indicators were recorded according to the manufacturer's instructions.

The following parameters were calculated for each group: NLR (neutrophil-to-lymphocyte ratio), PLR (platelet-to-lymphocyte ratio), MLR (monocyte-to-lymphocyte ratio), LMR (lymphocyte-to-monocyte ratio), MRR (monocyte-to-red blood cell ratio), NRR (neutrophil-to-red blood cell ratio), LRR (lymphocyte-to-red blood cell ratio), SII (systemic immune-inflammation index), and SIRI (systemic immune response index). $NLR = ANC(\times 10^9/L)/ALC(\times 10^9/L)$; $PLR = PLT(\times 10^9/L)/ALC(\times 10^9/L)$; $MLR = AMC(\times 10^9/L)/ALC(\times 10^9/L)$; $LMR = ALC(\times 10^9/L)/AMC(\times 10^9/L)$; $MRR = AMC(\times 10^9/L)/RBC(\times 10^9/L)$; $NRR = ANC(\times 10^9/L)/RBC(\times 10^9/L)$; $LRR = ALC(\times 10^9/L)/RBC(\times 10^9/L)$; $SII = PLT(\times 10^9/L) \times ANC(\times 10^9/L)/ALC(\times 10^9/L)$; $SIRI = ANC(\times 10^9/L) \times AMC(\times 10^9/L)/ALC(\times 10^9/L)$.
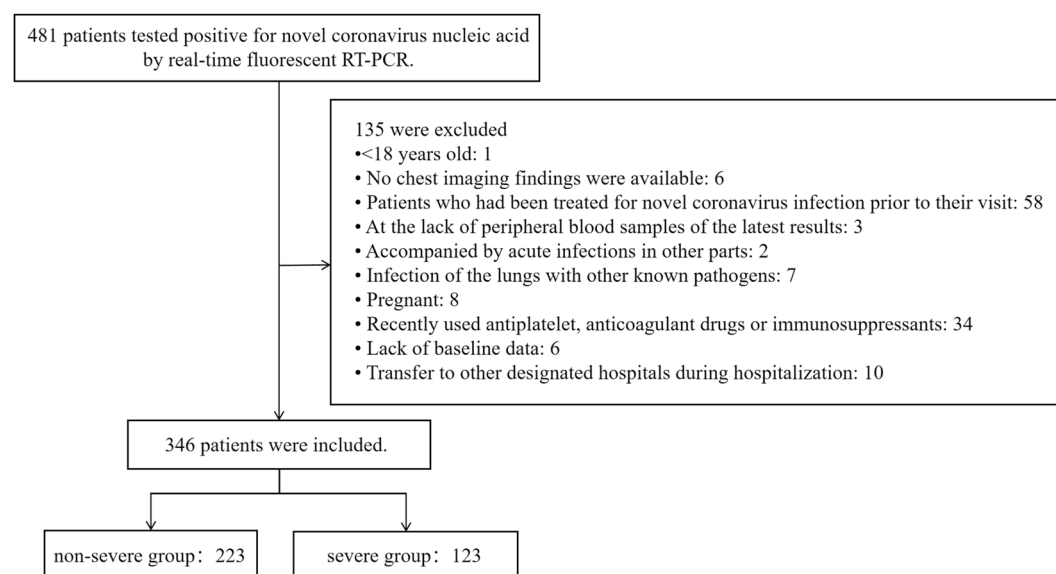

Figure 1. Flowchart of patient selection for this study.

Ethical approval and consent to participate

The present study was approved by The Ethical Review Committee of Jiangsu University Affiliated Hospital (Zhenjiang, China). Approval number: KY2023K1005. All patients provided informed consent.

Statistical analysis

Statistical analysis was performed using IBM SPSS statistical software 25.0 (IBM, USA) and R software (version 4.2.2). The Kolmogorov-Smirnov test was used to evaluate the distribution characteristics of the data. Count data were expressed as percentages (%). Intergroup comparisons were performed using the chi-square test or Fisher's exact test. If the data followed a normal distribution, they were expressed as mean ± standard deviation ($\bar{x} \pm s$). For non-normally distributed continuous data, logarithmic transformation was applied, and the distribution characteristics were evaluated again. If the data followed a normal distribution after transformation, they were expressed as median (interquartile range) [M (P25, P75)]. The process of taking the logarithm of variables can transform the data into a relatively uniform scale, thereby avoiding the effects of magnitude differences and reducing the correlation between variables, which can better reveal the true relationship between variables. The data after taking the logarithm still retains some characteristics of the original data, such as the central trend of the data and the relative size relationship. However, taking the logarithm will reduce the volatility of the data and make the data more stable, which is conducive to subsequent data analysis and model establishment. For the measurement data, two independent sample t-tests were used for the between-group comparison. LASSO regression analysis was employed to determine the basic variables associated with the risk of severe COVID-19. For risk factors with $p < 0.05$ in the univariate logistic regression analysis, stepwise backward-conditional logistic regression analysis was performed to select independent risk factors associated with non-severe and severe COVID-19. The likelihood ratio test was used to analyze the overall effectiveness of the model. The Hosmer–Lemeshow goodness-of-fit test was used to evaluate the fit of the model. ROC curves were used to evaluate the predictive value of individual or combined markers for the severity of COVID-19. The patients were divided into the training and validation cohorts with a ratio of 7:3 using the R function "createDataPartition" to ensure that outcome events were distributed randomly between the two cohorts. The training cohort was used to construct the model. The validation cohort was used to validate the results obtained using the training cohort. Welch's two-sample t-test and Pearson's chi-square test were used to analyze the data distribution characteristics of the training cohort and the validation cohort. A nomogram model was constructed using the "rms" package (version 6.7–1) in R software. Each patient's clinical and laboratory data were plotted in the nomogram, and the corresponding scores for each variable were obtained. The scores for all variables were summed to obtain a total score, and the vertical line corresponding to the final row of numbers represented the predicted probability, indicating the risk of severe COVID-19 in patients. Calibration was evaluated using the calibration curve. Calibration curves of this model were plotted using R software, and calibration curve analysis can be viewed as a visual Hosmer–Lemeshow test. The data analysis phase flowchart is shown in Fig. 2.
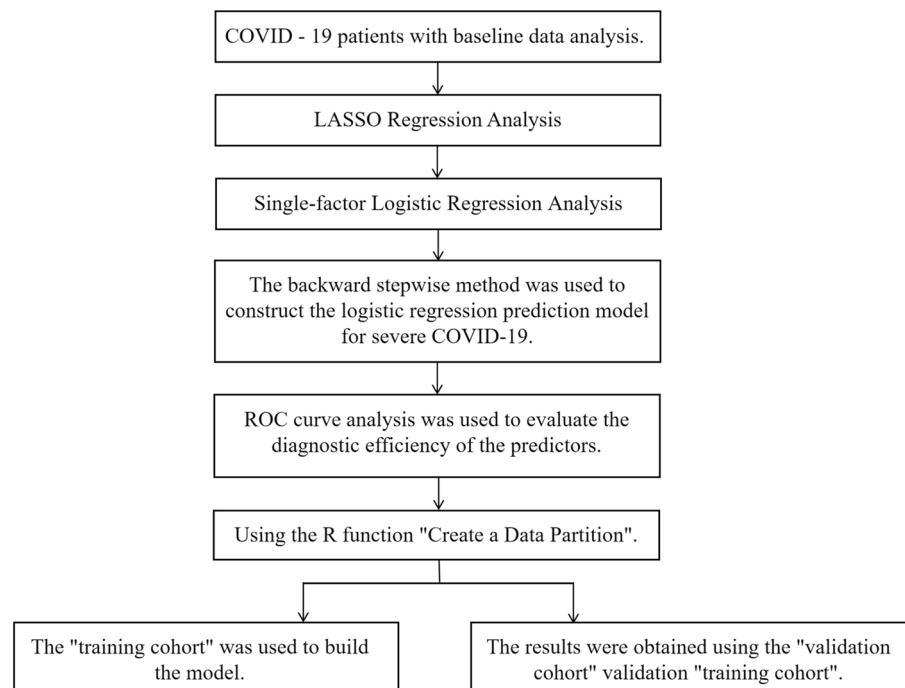

Figure 2. Flowchart of the data analysis phase.

Results

Baseline characteristics of the study participants

Baseline general information

A total of 346 patients with positive nucleic acid testing for the novel coronavirus were included in this study. General data on the patient is shown in Table 1. Among them, 123 cases (35.5%) were classified as severe, and 223 cases (64.5%) as non-severe. The average age of the patients in the severe group was 78.6 ± 10.9 years, while in the non-severe group, it was 73.0 ± 13.9 years. The average age in the severe group was significantly higher than in the non-severe group ($p < 0.05$). There were significant differences in the gender distribution between the two groups. In the severe group, there were 94 males (76.4%) and 29 females (23.6%), while in the non-severe group, there were 143 males (64.1%) and 80 females (35.9%). The proportion of male patients in the severe group was significantly higher than in the non-severe group ($p < 0.05$). The heart rate was 83.13 ± 12.2 breaths per minute in the non-severe group and 92.8 ± 14.0 breaths per minute in the severe group. The respiratory rate in the severe group was significantly higher than in the non-severe group ($p < 0.05$). There were also significant differences in oxygen saturation between the two groups. The oxygenation index in the non-severe group was 96.5 ± 1.5%, while in the severe group, it was 88.4 ± 6.8%. The oxygenation index in the severe group was significantly lower than in the non-severe group ($p < 0.05$).

Initial symptoms

The patient's symptoms and chest imaging findings are shown in Table 2. Among the 346 patients with COVID-19 upon admission, the most common initial symptom was a cough, reported by 296 patients, accounting for 85.5% of the cases. Fever was reported by 249 patients, accounting for 72.0% of the cases. Compared to the non-severe group, the severe group had a higher proportion of patients with symptoms such as wheezing, respiratory distress, and altered consciousness, and these differences were statistically significant ($p < 0.05$). The non-severe group had a significantly higher proportion of patients with fatigue as their initial symptom than the severe group ($p < 0.05$). The two groups had no significant differences in other initial symptoms ($p \geq 0.05$). In terms of radiology, there were 192 cases (55.5%) with ground glass opacities (GGO) in the patients and 167 cases (48.3%) with subpleural lesions. Both of these are common radiological features in COVID-19 patients. Compared with the non-severe group, the severe group had a higher proportion of bilateral lung inflammation, which was statistically significant ($P < 0.05$).

Hematological and inflammatory marker data of the two patient groups

The hematological and immunological marker data were compared between the severe and non-severe groups of patients with COVID-19. The results of the laboratory examinations are presented in Tables 3, 4 and 5. Among them, Table 4 shows the results of the Kolmogorov–Smirnov test with normal distribution of inflammation index in patients. The results showed that the inflammatory index of the patients did not follow the normal distribution. There were no statistically significant differences between the two groups in terms of platelet (PLT) and erythrocyte sedimentation rate (ESR) levels ($p \geq 0.05$). However, the severe group exhibited significantly higher levels of white blood cell count (WBC), absolute lymphocyte count (ALC), absolute neutrophil count (ANC), neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), monocyte-to-lymphocyte ratio (MLR), lymphocyte-to-red blood cell ratio (LRR), monocyte-to-red blood cell ratio (MRR), neutrophil-to-red blood cell ratio (NRR), systemic immune-inflammation index (SII), and systemic immune response index (SIRI) compared to the non-severe group. Conversely, the severe group had significantly lower lymphocyte-to-monocyte ratio (LMR) levels than the non-severe group. These differences were all statistically significant ($p < 0.05$).

	Total (n = 346)	Non-severe group (n = 223)	Severe group (n = 123)	P value
Gender (male/female)	237/109	143/80	94/29	0.018
Smoking history, n (%)	73 (21.1%)	47 (21.1%)	26 (21.1%)	0.989
Alcohol consumption history, n (%)	42 (12.1%)	30 (13.5%)	12 (9.8%)	0.314
Age, years (mean ± SD)	75.0 ± 13.1	73.1 ± 13.9	78.6 ± 10.9	0.013
Respiratory rate, breaths per min	21.1 ± 3.7	19.8 ± 2.1	23.4 ± 4.8	0.184
Heart rate beats per min	86.6 ± 13.6	83.1 ± 12.2	92.8 ± 14.0	< 0.001
Oxygenation index, %	93.6 ± 5.7	96.5 ± 1.5	88.4 ± 6.8	< 0.001
Comorbidities, n (%)				
Hypertension	205 (59.2%)	120 (53.8%)	85 (69.1%)	0.006
Diabetes	98 (28.3%)	51 (22.9%)	47 (38.2%)	0.002
Heart diseases	80 (23.1%)	42 (18.8%)	38 (30.8%)	0.011
Nerve system disease	68 (19.7%)	46 (20.6%)	22 (17.8%)	0.539
Chronic lung disease	67 (19.4%)	37 (16.6%)	30 (24.4%)	0.079
Liver and kidney disease	43 (12.4%)	26 (11.7%)	17 (13.8%)	0.560
Cancer	63 (18.2%)	40 (17.9%)	23 (18.7%)	0.860

Table 1. General data comparing non-severe and severe COVID-19 groups.

Symptom, n (%)	Total (n = 346)	Non-severe group (n = 223)	Severe group (n = 123)	P value
Fever	249 (72.0%)	157 (70.4%)	92 (73.2%)	0.411
Cough	296 (85.5%)	195 (87.4%)	101 (82.1%)	0.208
Sputum	273 (78.9%)	177 (79.4%)	96 (78.0%)	0.817
Rhinorrhoea	7 (2.0%)	5 (2.2%)	2 (1.6%)	1.000
Throat pain	26 (7.5%)	21 (9.4%)	5 (4.1%)	0.159
Myalgia	40 (11.6%)	30 (13.5%)	10 (8.1%)	0.118
Fatigue	65 (18.8%)	51 (22.9%)	14 (11.4%)	0.009
Headache	4 (1.2%)	4 (1.8%)	0 (0.0%)	0.333
Dizzy	8 (2.3%)	7 (3.1%)	1 (0.8%)	0.315
Chest distress	194 (56.1%)	121 (54.3%)	73 (59.3%)	0.357
Pectoralgia	9 (2.6%)	7 (3.1%)	2 (1.6%)	0.622
Poor appetite	37 (10.7%)	25 (11.2%)	12 (9.8%)	0.607
Nausea or vomiting	26 (7.5%)	21 (9.4%)	5 (4.1%)	0.074
Abdominal pain	5 (1.4%)	5 (2.2%)	0 (0.0%)	0.229
Diarrhea	8 (2.3%)	8 (3.6%)	0 (0.0%)	0.08
Dyspnea	198 (56.9%)	116 (52.0%)	82 (66.7%)	0.007
Gasp	101 (29.0%)	48 (21.5%)	53 (43.1%)	0.001
Unconsciousness	29 (8.3%)	9 (4.0%)	20 (16.3%)	0.001
Chest CT images				
Normal	121 (35.0%)	89 (40.0%)	32 (26.0%)	0.013
Abnormal	225 (65.0%)	134 (60.0%)	91 (74.0%)	-
Bilateral lung	121 (35.0%)	54 (24.2%)	67 (54.5%)	< 0.001
Unilateral lung	104 (30.0%)	80 (35.9%)	24 (19.5%)	-
GGO	192 (55.5%)	106 (47.5%)	86 (69.9%)	< 0.001
Subpleural lesions	167 (48.3%)	146 (65.5%)	21 (17.1%)	< 0.001

Table 2. Symptoms and chest imaging of COVID-19 in the non-severe and severe care group.

Variables	Non-severe group (n = 225)	Severe group (n = 123)	t value	P value
WBC	7.62 ± 4.29	8.86 ± 4.50	2.519	0.012
ALC	1.02 ± 0.59	0.70 ± 0.48	− 5.162	0.001
ANC	5.99 ± 4.14	7.64 ± 4.38	3.475	0.001
PLT	195.28 ± 90.60	178.20 ± 93.78	− 1.661	0.098
ESR	53.88 ± 33.28	56.71 ± 33.67	0.755	0.451
CRP	69.82 ± 68.23	104.62 ± 78.42	− 4.336	0.001

Table 3. Comparison of blood routine and inflammatory markers in the non-severe and severe COVID-19 groups ($\bar{x} \pm s$). *WBC* white blood cell count, *ALC* absolute lymphocyte count, *ANC* absolute neutrophil count, *PLT* platelet, *ESR* erythrocyte sedimentation rate, *CRP* C-reactive protein.

Variables	Mean value	Standard deviation	P value
NLR	10.83	12.57	< 0.001
PLR	284.11	227.68	< 0.001
MLR	0.76	0.58	< 0.001
LMR	2.15	3.04	< 0.001
LRR	0.24	0.18	< 0.001
MRR	0.15	0.09	< 0.001
NRR	1.77	1.21	< 0.001
SII	2145.30	3442.39	< 0.001
SIRI	5.79	8.06	< 0.001

Table 4. The inflammatory index of the total patients was tested for normal distribution.

Variables	Non-severe group (n = 223)	Severe group (n = 123)	P value
NLR	5.40 (3.14, 9.00)	11.64 (6.25, 19.67)	< 0.001
PLR	203.16 (125.71, 318.33)	275.00 (170.00, 460.00)	0.001
MLR	0.57 (0.33, 0.86)	0.75 (0.42, 1.09)	0.004
LMR	1.75 (1.17, 3.00)	1.33 (0.92, 2.40)	0.006
LRR	0.24 (0.17, 0.35)	0.15 (0.10, 0.22)	< 0.001
MRR	0.13 (0.10, 0.19)	0.11 (0.06, 0.19)	0.003
NRR	1.29 (0.81, 2.08)	1.83 (1.10, 2.76)	0.011
SII	1020.44 (523.64, 1991.67)	1976.00 (826.42, 3103.00)	< 0.001
SIRI	2.82 (1.32, 5.48)	5.04 (2.00, 9.35)	< 0.001

Table 5. Comparison of inflammation index between non-severe and severe COVID-19 groups [M (P25, P75)]. NLR = ANC($\times 10^9$/L)/ALC($\times 10^9$/L); PLR = PLT($\times 10^9$/L)/ALC($\times 10^9$/L); MLR = AMC($\times 10^9$/L)/ALC($\times 10^9$/L); LMR = ALC($\times 10^9$/L)/AMC($\times 10^9$/L); MRR = AMC($\times 10^9$/L)/RBC($\times 10^9$/L); NRR = ANC($\times 10^9$/L)/RBC($\times 10^9$/L); LRR = ALC($\times 10^9$/L)/RBC($\times 10^9$/L); SII = PLT($\times 10^9$/L) $\times$ ANC($\times 10^9$/L)/ALC($\times 10^9$/L); SIRI = ANC($\times 10^9$/L) $\times$ AMC($\times 10^9$/L)/ALC($\times 10^9$/L).

Peripheral blood biomarker data of the two patient groups
We compare hematological and biochemical parameters between severe and non-severe COVID-19 patients:

The analysis results, as shown in Table 5, indicate that the levels of total bilirubin (TBIL), direct bilirubin (DBIL), aspartate aminotransferase (AST), blood urea nitrogen (BUN), and creatinine (Cre) were significantly higher in the severe group compared to the non-severe group, with all differences being statistically significant ($p < 0.05$). However, there were no significant differences in indirect bilirubin (IBIL), alanine aminotransferase (ALT), triglycerides (TG), and total cholesterol (TC) between the two groups, with no statistical significance ($p \geq 0.05$).

We compare coagulation function parameters between severe and non-severe COVID-19 patients:

The analysis results, as presented in Table 6, reveal that the levels of international normalized ratio (INR) and D-dimer (DD) were significantly higher in the severe group compared to the non-severe group, with both differences being statistically significant ($p < 0.05$). However, there were no significant differences in prothrombin time (PT), activated partial thromboplastin time (APTT), thrombin time (TT), and fibrinogen (FIB) levels between the two groups, with no statistical significance ($p \geq 0.05$).

We compare cardiac enzyme profile parameters between severe and non-severe COVID-19 patients:

Variables	Non-severe group (n = 223)	Severe group (n = 123)	P value
TBIL	10.30 (8.00, 14.40)	11.40 (8.10, 16.80)	0.018
DBIL	3.30 (2.60, 4.70)	4.40 (3.10, 7.40)	< 0.001
IBIL	6.50 (4.80, 9.30)	6.30 (4.20, 10.40)	0.863
ALT	21.70 (15.20, 33.10)	24.20 (16.00, 41.80)	0.123
AST	23.70 (15.20, 36.80)	30.50 (20.80, 54.10)	< 0.001
BUN	6.07 (4.39, 9.09)	9.91 (6.07, 16.06)	< 0.001
Cre	69.10 (56.60, 93.70)	85.90 (63.70, 157.20)	< 0.001
PT	12.00 (11.40, 12.80)	12.20 (11.60, 13.00)	0.519
INR	1.05 (0.99, 1.13)	1.06 (1.01, 1.16)	0.039
APTT	28.80 (27.00, 31.90)	30.00 (27.10, 33.20)	0.391
TT	17.00 (16.40, 17.80)	17.50 (16.50, 18.50)	0.108
FIB	4.75 (3.87, 5.74)	4.90 (4.02, 6.59)	0.842
DD	0.97 (0.52, 2.42)	2.41 (1.03, 5.75)	< 0.001
CK	65.00 (35.00, 136.00)	136.00 (57.00, 313.00)	< 0.001
CK-MB	9.60 (6.60, 14.00)	16.50 (10.10, 29.40)	< 0.001
LDH	197.00 (164.00, 238.00)	300.00 (234.00, 400.00)	< 0.001

Table 6. Comparison of severe and non-severe COVID-19 groups Peripheral blood biomarkers [M (P25, P75)]. *TBIL* total bilirubin, *DBIL* direct bilirubin, *IBIL* indirect bilirubin, *ALT* alanine aminotransferase, *AST* aspartate transaminase, *BUN* blood urea nitrogen, *Cre* creatinine, *TG* triglycerides, *TC* total cholesterol, *PT* prothrombin time, *PT-INR* prothrombin international normalization ratio, *APTT* activated partial thromboplastin time, *TT* thrombin time, *FIB* fibrinogen, *DD* D-dimmer, *CK* creatine kinase, *CK-MB* creatine kinase-MB, *LDH* lactate dehydrogenase.

The results indicate that the levels of creatine kinase (CK), creatine kinase-MB (CK-MB), and lactate dehydrogenase (LDH) were significantly higher in the severe group compared to the non-severe group, with all differences being statistically significant ($p < 0.05$).

Selection of risk prediction factors for COVID-19

LASSO regression analysis for COVID-19
In the process of building the regression model, a large number of independent variables can lead to inflated coefficients, potentially causing overfitting. To efficiently extract important variables, LASSO regression was used for the regularization and selection of variables. The degree of complexity adjustment in LASSO regression was controlled by the parameter λ, where a larger λ value indicates a stronger penalty on the variables. The selection of variable combinations depends on the adjustment of λ.

Figure 3 presents the LASSO regression path plot obtained through the R software for variable selection. The changes in each variable's trajectory were shown, with the logarithm of λ on the x-axis and the regression coefficients of the variables on the y-axis. As λ increases, the regression coefficients of the variables gradually shrunk toward zero. A non-zero coefficient suggested a greater contribution of the variable to the outcome, making it more likely to be retained.

Figure 4 displayed the tenfold cross-validation results of the LASSO regression, showing the relationship between the logarithm of λ (log (λ)), the mean squared error (MSE), and the number of variables in the model. When cross-validation is performed, the function will automatically divide the original data set into 10 parts. The function will use 9 of these data sets to train the model and use the remaining 1 data set to test the training results and give the error. This process will be repeated 10 times. In each cross-validation, the function will try to substitute different λ to build the model, so that the model error under different λ is obtained. The dotted line in the middle of the value of positive and negative standard deviation of the logarithm (lambda) range. On the left side of the dotted line the model error logs the most hours of harmonic parameters (lambda) value. For the clinical prediction model, we tend to choose a higher precision of the model. Model error of the hour is the optimal value, when lambda is 0.012, get excellent performance with the least variable number of models.

In the end, 26 variables were selected as predictive factors for severe COVID-19, categorized as Age, Height, Day, RR, Heart rate, Oxygen, Mechanical Ventilation, Organ Failure, Fatigue, Eosinophilic Granulocyte%, Basophilic Granulocyte%, ALC, RDW, MPV, CRP, PCT, TBIL, ADA, UA, TG, TC, APO-B, CK, FIB, DD, and LDH.

Multivariable logistic regression analysis of severe and non-severe cases of COVID-19
Based on the LASSO regression, a single-factor logistic regression analysis was conducted on the identified risk factors. Figure 5 shows that risk factors with a significance level of $P < 0.05$ were selected, and a backward stepwise method was used to construct the logistic regression prediction model for severe COVID-19. The predictive model achieved an overall accuracy of 80.9%, with an accuracy of 91.0% for non-severe cases and 62.6% for severe cases, indicating a high level of accuracy. The likelihood ratio test demonstrated the effectiveness of the included independent variables in constructing the model ($P < 0.001$), indicating the significance of the model construction. The Hosmer–Lemeshow goodness-of-fit test indicated a good fit of the model to the prediction results, with no significant difference between the predicted probabilities and the actual probabilities ($P = 0.118$).
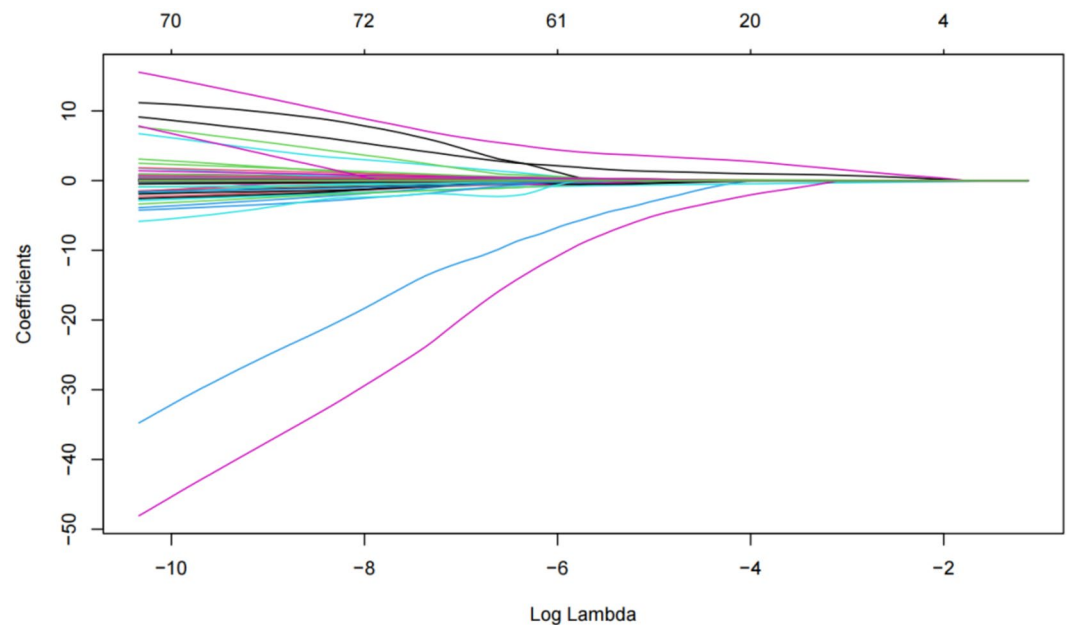

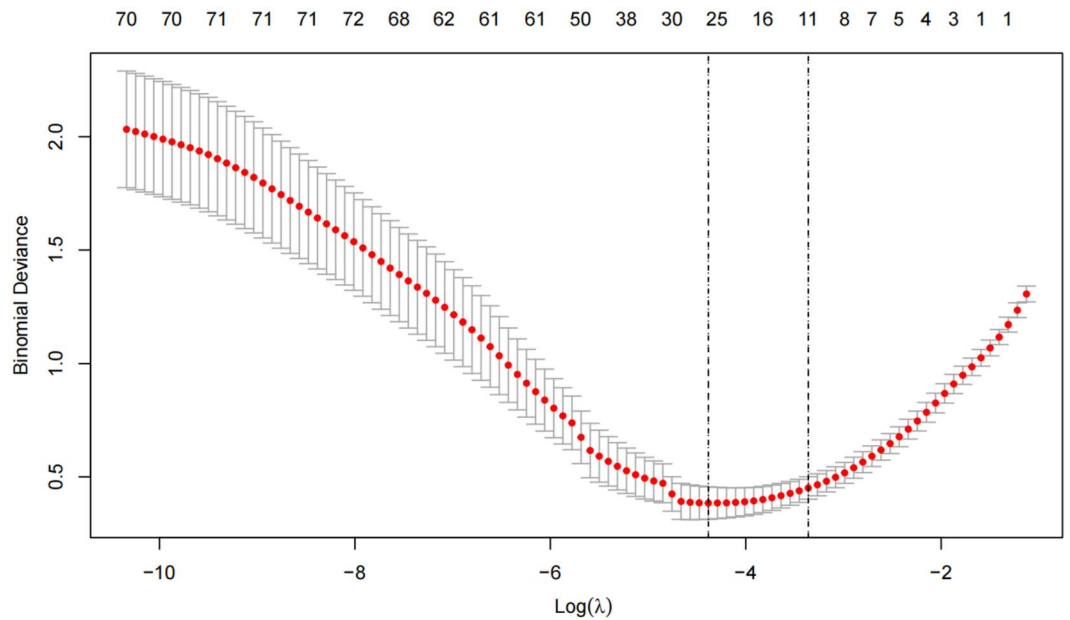**Figure 3.** LASSO regression path plot: LASSO regression path plot for variable selection obtained by R software.

Figure 4. Tenfold cross-validation results of LASSO regression: show the relationship between log (λ), mean square error (MSE), and the number of variables in the model.
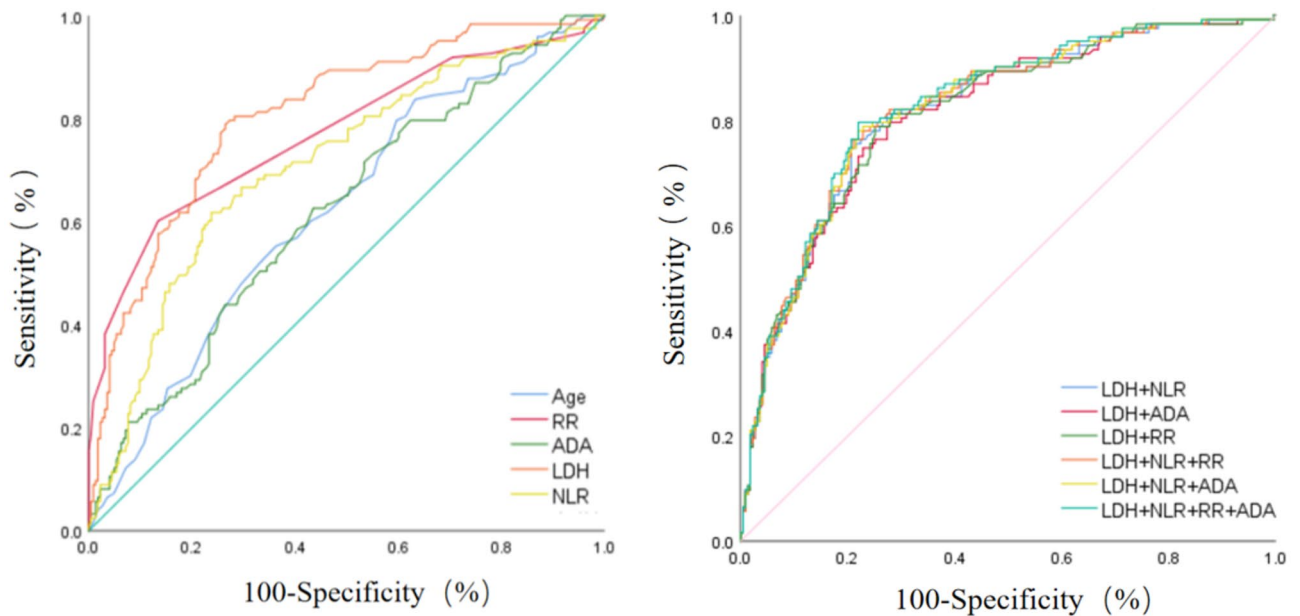

Figure 5. A logistic regression prediction model for severe COVID-19.

The results, as shown in Table 7, indicate that increasing age, accelerated respiratory rate, elevated ADA, LDH, and NLR levels were associated with an increased risk of severe COVID-19, with statistically significant differences ($P < 0.05$). In conclusion, age, respiratory rate, ADA level, LDH level, and NLR level are independent predictive factors for severe COVID-19.

ROC curve analysis of biomarkers for severe and non-severe cases of COVID-19
ROC curve analysis was performed to evaluate the discriminative ability of Age, RR, LDH, and NLR for distinguishing between non-severe and severe cases of COVID-19. The results, as shown in Table 8. The results showed that LDH, RR, and NLR exhibited the expected diagnostic value for severe COVID-19, with LDH demonstrating higher diagnostic efficiency. The AUC value for LDH was 0.809 (95% CI 0.761–0.856), with a sensitivity of 78.9% and specificity of 73.5%. The AUC value for RR was 0.772 (95% CI 0.716–0.827), with a sensitivity of 60.2% and specificity of 86.4%, indicating higher specificity in diagnosing severe COVID-19. NLR had lower diagnostic

	β	S.E	Wald X²	P value	OR	95% CI	
						Lower	Upper
Age	0.037	0.013	7.391	0.007	1.037	1.010	1.065
RR	0.411	0.066	38.817	0.000	1.508	1.325	1.716
ADA	0.058	0.026	4.976	0.026	1.059	1.007	1.115
LDH	0.007	0.002	20.793	0.000	1.007	1.004	1.010
NLR	0.034	0.015	4.967	0.026	1.035	1.004	1.066
Constant	− 15.145	1.897	63.712	0.000	0.000		

Table 7. Multivariate binary logistic regression analysis of severe and non-severe COVID-19 groups.

Variables	AUC	Cut-off	Sensitivity, %	Specificity, %	95% CI	
					Lower limit	Upper limit
NLR	0.710	9.17	61.8	76.2	0.652	0.767
LDH	0.809	230.50	78.9	73.5	0.761	0.856
RR	0.772	21.50	60.2	86.4	0.716	0.827
ADA	0.620	13.35	62.6	56.5	0.559	0.681
Age	0.621	69.50	83.7	36.8	0.561	0.682
LDH + NLR	0.817	–	74.8	79.4	0.770	0.864
LDH + RR	0.814	–	78.9	74.9	0.767	0.861
LDH + NLR + RR	0.823	–	76.4	78.9	0.777	0.869

Table 8. ROC curve analysis of severe and non-severe COVID-19 markers.

efficiency compared to LDH and RR, with an AUC value of 0.710 (95% CI 0.652–0.767), sensitivity of 61.8%, and specificity of 76.2%. Due to their good performance in ROC curve analysis, LDH, RR, and NLR were selected for further analysis. The AUC values for LDH combined with NLR or RR were 0.817 and 0.814, respectively, which were higher than the diagnostic efficiency of NLR alone (AUC, 0.710) and RR alone (AUC, 0.772), and the sensitivity was also improved (sensitivity of 74.8% and 78.9% respectively). The combined analysis demonstrated that the diagnostic efficiency of the LDH + NLR + RR combined index was higher than that of single indices, with an AUC value of 0.823 (95% CI 0.777–0.869).

As shown in Fig. 5 for the risk factors with a significance level of $P < 0.05$, a logistic regression prediction model for severe COVID-19 was constructed by using a reverse step-by-step method.

Construction and validation of the nomogram
The patients were divided into the training and validation cohorts with a ratio of 7:3 using the R function "create-DataPartition" to ensure that outcome events were distributed randomly between the two cohorts. The training cohort was used to construct the model. The validation cohort was used to validate the results obtained using the training cohort. A total of 346 COVID-19 patients were randomly divided into a training cohort and a validation cohort at a ratio of 7:3. The general data and clinical characteristics of these patients are summarized in Table 9. In the training and validation cohorts, the mean age of COVID-19 patients was 74 and 76 years, respectively. The respiratory rate was 21.1 ± 3.8 and 21.1 ± 3.5 breaths per minute, and the heart rate was 86 ± 14 and 87 ± 13 beats per minute. There were 212 patients using mechanical ventilation in the training cohort and 90 in the validation cohort. In the training and validation cohorts, the mean LDH was 267 ± 236 and 271 ± 164, ADA 14.8 ± 6.0 and 15.3 ± 12.7, SII 2,016 ± 2,514 and 2,445 ± 4,977, and NLR 11 ± 12 and 11 ± 15, respectively. In the training and validation cohorts, there were 135 and 57 patients with GGO, 115 and 52 patients with subpleural lesions, and 85 and 36 patients with bilateral lung inflammation in the lung imaging examination. The training and validation cohorts were comparable in terms of general data and clinical characteristics ($P > 0.05$).

As shown in Fig. 6, significant and independent predictors were identified based on regression analysis and clinical considerations to construct a predictive nomogram model. The nomogram model included twelve variables (age, creatinine, respiratory rate, heart rate, mechanical ventilation, lymphocyte count, GGO, subpleural lesions, ADA, LDH, NLR, and SII). The severity of COVID-19 could be estimated by summing the scores of each independent variable and predicting the total score on the lowest scale.

Nomogram validation and evaluation
In this study, ROC analysis, DCA analysis, and calibration plots were used to test the predictive efficiency of the probability of COVID-19, and the results showed that the nomogram had good prediction efficiency. In the ROC curve, the Y-axis is called the sensitivity, which also becomes the true positive rate. Higher values on the Y-axis represent higher model accuracy. The X-axis is 1-specificity, also known as the false positive rate, and the closer the intersection point between the curve and the X-axis is to 0, the higher the accuracy of the model. The

Characteristic	Training cohort n = 242	Validation cohort n = 104	P value
Age, years (mean ± SD)	74 ± 14	76 ± 11	0.193
Respiratory rate, breaths per min	21.1 ± 3.8	21.1 ± 3.5	0.936
Heart rate beats per min	86 ± 14	87 ± 13	0.759
Mechanical Ventilation			0.785
0	212 (87.6%)	90 (86.5%)	
1	30 (12.4%)	14 (13.5%)	
LDH (mean ± SD)	267 ± 236	271 ± 164	0.878
ADA (mean ± SD)	14.8 ± 6.0	15.3 ± 12.7	0.709
SII (mean ± SD)	2016 ± 2,514	2445 ± 4977	0.406
NLR (mean ± SD)	11 ± 12	11 ± 15	0.699
ALC (mean ± SD)	0.92 ± 0.59	0.86 ± 0.54	0.323
GGO			0.867
0	107 (44.2%)	47 (45.2%)	
1	135 (55.8%)	57 (54.8%)	
Bilateral lung			0.947
0	156 (64.7%)	65 (64.4%)	
1	85 (35.3%)	36 (35.6%)	

Table 9. General data and clinical characteristics of the patients.
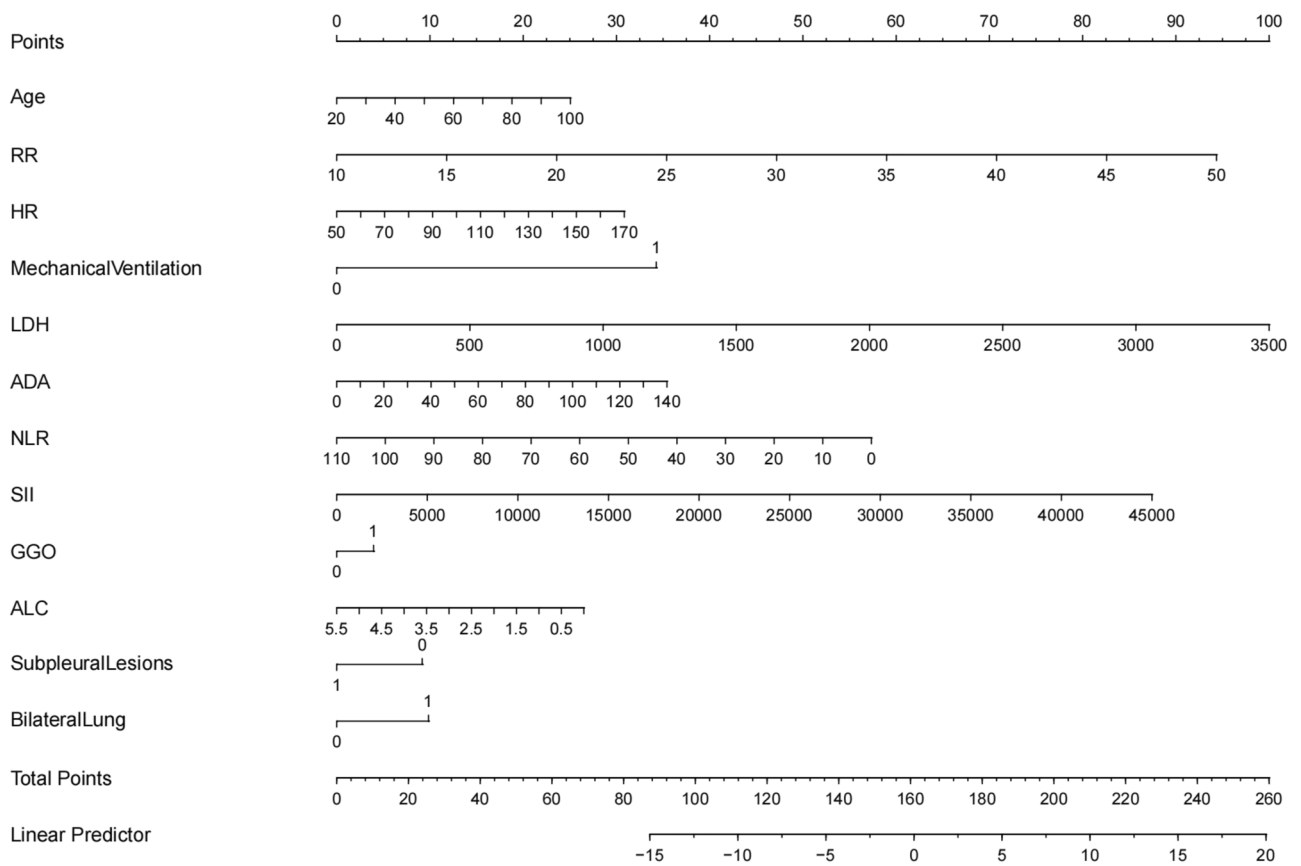
Figure 6. A nomogram to predict the severity of COVID-19.

area under the curve (AUC) ranges from 0.5 to 1, and the closer the AUC is to 1, the better the diagnostic effect of the model in predicting the outcome. As shown in Fig. 7, the AUC of the ROC curve was 0.981 for the training cohort and 0.907 for the internal validation cohort. The AUC values of the two cohorts reflected the good diagnostic effect of the nomogram. In the calibration curve, the X-axis represents the predicted probability of an event and the Y-axis represents the actual probability of an event. The thick gray line represents the ideal reference line when the predicted probability matches the actual probability, while the dashed and solid lines represent the
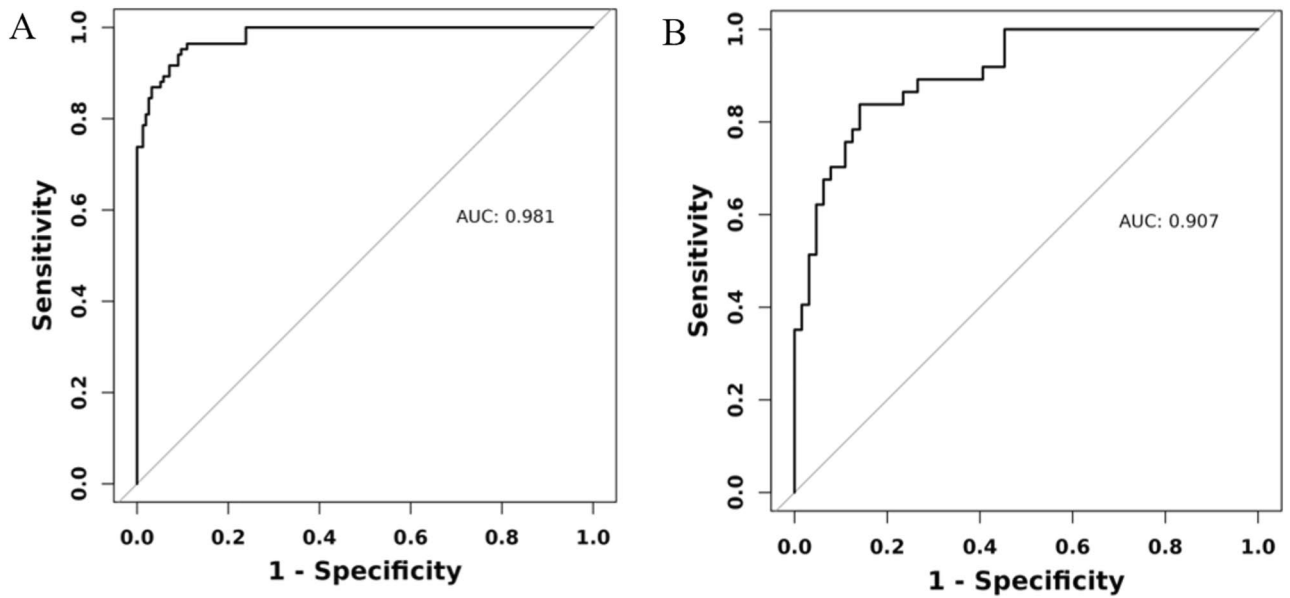
Figure 7. ROC curves for the nomogram. (**A**): Training group; (**B**): Validation group.

calibration curve for the entire cohort and the model curve built through internal validation. A higher degree of fit between the two indicates a better predictive performance of the nomogram model. As shown in Fig. 8, the curve representing the risk of severe COVID-19 disease estimated by the model is in good agreement with the probability curve observed in internal validation, indicating that the nomogram performs better in predicting the probability of COVID-19. DCA assessed the clinical validity of the model. Based on the classification results, the X-axis represents the boundary of the expected likelihood value, and the Y-axis represents the normalized net benefit at this boundary. Gray and black reference lines indicate the "all intervention" and "no intervention" hypotheses, respectively. In the threshold probability range of 0.1 to 0.7, DCA curves lie above the two baselines "none" and "all," indicating that the performance of the model is acceptable in this range. As shown in Fig. 9, this nomogram has clinical utility. In conclusion, the calibration plot and DCA analysis showed that the nomogram had a good predictive effect on the severity of COVID-19.
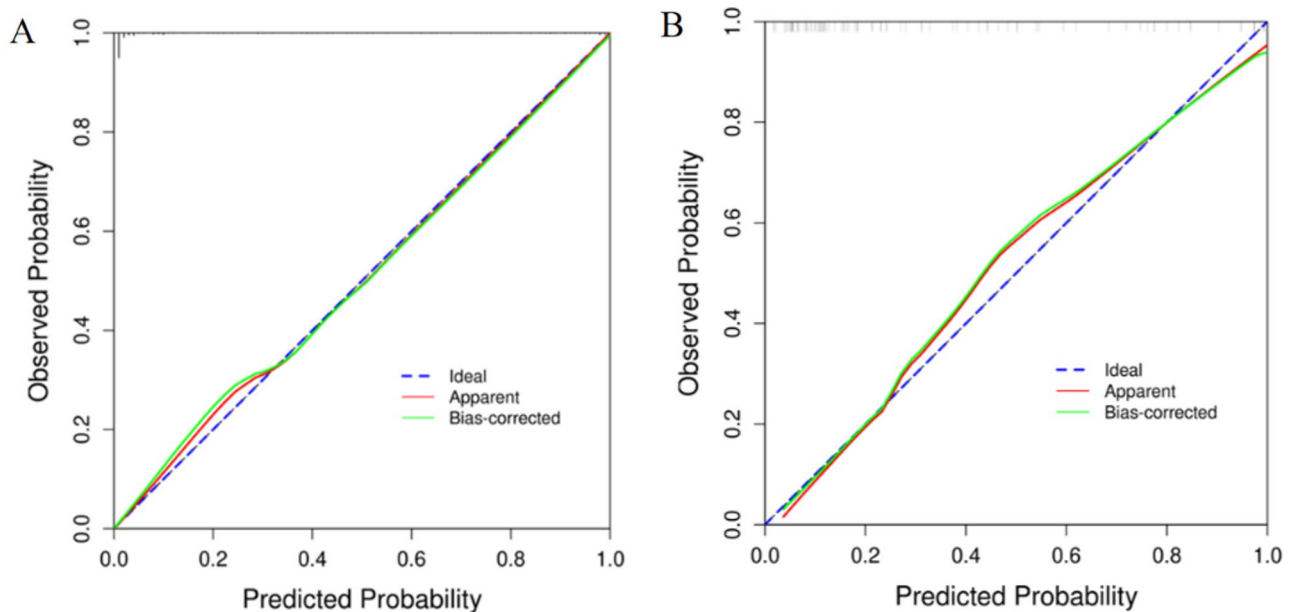
Figure 8. Calibration curve for predicting the probability of COVID-19 severity. (**A**): Training group; (**B**): Validation group.
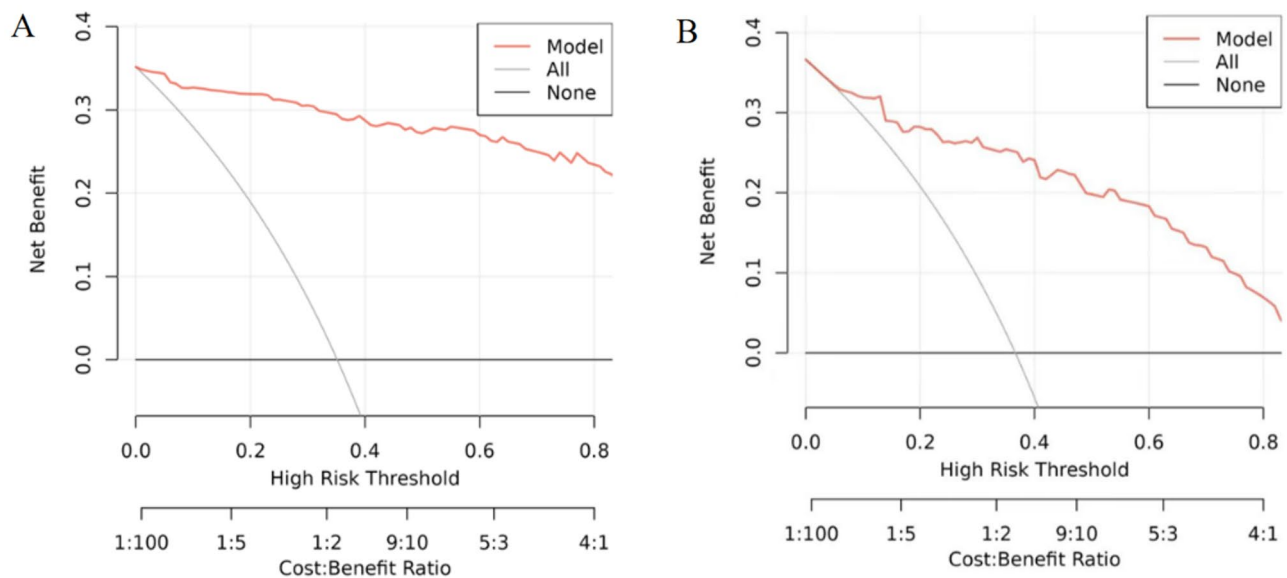
Figure 9. Decision curve analysis in the prediction of COVID-19 severity. (**A**): Training group; (**B**): Validation group.

Discussion

The results of this study indicate that in the severe group, respiratory rate, breathlessness, altered consciousness, NLR, and LDH levels were significantly higher compared to the non-severe group. Imaging findings suggest that in the severe group, there was a higher proportion of bilateral pulmonary inflammation and ground-glass opacities. NLR and LDH were identified as independent risk factors for severe patients. The diagnostic efficiency was maximized when NLR, RR, and LDH were combined. In this study, we developed a COVID-19 severity risk prediction model. It includes twelve variables to predict the risk of severe COVID-19. The total score is calculated by adding up the scores for each of the fourteen independent variables. By mapping the total score to the lowest scale, we can estimate the probability of severe COVID-19 risk.

Regarding COVID-19 diagnostic and survival prognosis models, the National Health Commission of the People's Republic of China has reported one of the initial validated survival models, which includes ten independent predictive factors (chest imaging abnormalities, age, hemoptysis, dyspnea, altered consciousness, comorbidity count, cancer history, NLR, LDH, DBIL, and TBIL)[13]. However, this prognosis model only includes 59 cases of severe and fatal patients and has not yet been externally validated in different patient populations and healthcare settings in Western societies. Among the expanding list of other models, researchers from the UK reported one of the largest models. They collected observational data from 57,824 COVID patients across 260 hospitals in England, Scotland, and Wales. Their severity score includes eight variables (age, gender, comorbidity count, respiratory rate, peripheral blood oxygen saturation, consciousness level, BUN, and CRP)[14]. Another approach from Spanish researchers focuses on prognosis features directly related to the pathophysiology of COVID-19 rather than patient characteristics, constructing a model for the mortality of severe patients based on peripheral oxygenation levels during hospitalization, ANC, PLT, LDH, and CRP[15]. Although some common variables are shared among these models, there is significant variation in predictive outcomes. Other studies have also reported models based on deep learning algorithms, using CT images to predict the severity of COVID-19, showing high accuracy[16,17]. However, these models are challenging to construct due to their complex algorithms. Consequently, these models may not be suitable for all institutions and healthcare professionals.

Therefore, we are attempting to develop a novel predictive model for the risk of severe COVID-19. This predictive model relies solely on clinical manifestations, laboratory indicators, and imaging features. These are readily obtainable and identifiable in a clinical setting. Ultimately, based on the observed results, we aim to use a risk score to predict the risk of severe illness in COVID-19 patients.

First, we analyzed the clinical characteristics of the patients in the study. The most common clinical presentations were fever, cough, and phlegm production. Among patients in the severe group, there was a higher prevalence of increased respiratory rate, breathlessness, and altered consciousness as first symptoms compared to the non-severe group, and this difference was statistically significant. Additionally, the severe group had significantly older individuals when compared to the non-severe group. In both groups, the prevalence of severe cases was higher in males than in females, which is consistent with previous research, further underscoring the significant association between age, gender, and disease severity[18–21]. Smoking history has been considered a risk factor for severe COVID-19[22]. Similarly, Mehra et al. demonstrated a higher in-hospital mortality rate among current smokers in COVID-19 patients[23]. However, in our study, no significant difference was observed between the severe and non-severe groups, in contrast to some prior studies. Variations in inclusion criteria or sample size differences between study populations might explain the disparities between our findings and those of previous studies. Our results indicated a significantly higher likelihood of severe illness in patients with comorbidities such as hypertension, diabetes, and heart diseases, which aligns with previous research, supporting

the notion that patients with underlying conditions are more likely to progress to severe illness. While previous studies have shown that chest CT scans of non-severe COVID-19 patients often display ground-glass opacities, our study found a higher proportion of ground-glass opacities in the severe group. This difference may be due to the relatively small sample size in our study, potentially introducing some bias. Further research with larger sample sizes is needed to validate these findings.

After discussing the clinical characteristics of COVID-19, we analyzed the immunological features of peripheral blood in COVID-19 patients. Compared to non-severe patients, severe patients had elevated white blood cell and neutrophil counts upon admission, while lymphocyte counts were significantly reduced. This is in line with results from other related studies[12,20,24,25] and is believed to be an effect of the virus on T cells through ACE2 receptor infection[26]. In comparison to non-severe patients, severe patients had a higher NLR, and this difference was statistically significant. NLR is particularly useful. It is associated with systemic inflammatory status and disease activity. Additionally, NLR has prognostic value in cardiovascular diseases[27], autoimmune diseases[28], tumors[28], and other infectious diseases[29]. Some scholars have indicated that NLR is an early marker of infection in COVID-19 patients[30], as virus-induced inflammation raises NLR levels. Elevated NLR further promotes the progression of COVID-19. Some studies have also identified the role of NLR in distinguishing COVID-19 severity and predicting mortality[10,20,31–33], and our study's results are consistent with these findings. In our multifactorial logistic regression model, NLR emerged as a crucial predictive factor for the severity of COVID-19 in patients. Our data showed a significant increase in LDH levels among severe patients. Some studies have suggested that elevated serum LDH levels are an independent predictive factor for severe cases[11], which aligns with our findings in the current study. To summarize, our research further validates the use of NLR and LDH in predicting COVID-19 severity. In this study, through the observation of ROC curves, we noted that NLR, RR, and LDH have the potential to distinguish between severe and non-severe COVID-19 patients. Particularly, combined LDH and NLR testing exhibits high specificity. The predictive efficiency is maximized when NLR, RR, and LDH are combined.

In summary, we hypothesize that these clinical characteristics, laboratory indicators, and imaging findings combined may be more useful for clinicians as practical tools in assessing the severity and prognosis of COVID-19 patients. Therefore, we used twelve variables, including age, RR, HR, mechanical ventilation, ALC, ADA, LDH, NLR, SII, GGO in chest CT, subpleural lesions, and bilateral pulmonary inflammation, to construct the COVID-19 severity prediction model. Finally, by adding the scores of each of the twelve independent variables, we calculated a total score. By mapping the total score to the lowest scale, we were able to estimate the severity risk of COVID-19 patients. Nomograms are a reliable tool for creating statistical prediction models, resulting in simple and intuitive charts that quantify the risk of clinical events. ROC, calibration curve, and DCA analysis were used to validate the nomogram model, which could be used to judge the prediction effect of the nomogram. Therefore, compared to other clinical prediction models, the model we have established is faster, simpler, and more practical.

Finally, it should be acknowledged that this study has some limitations. First, this is a single-center retrospective study. The study population was relatively small, which inevitably led to some bias. In the future, we can conduct multicenter studies to expand the scope of the study population and validate the results of this study. Secondly, being a retrospective study, data were collected based on electronic records from the hospital, and the accuracy and reliability may vary across different hospitals. We can increase the researcher's follow-up data collection scope, join more hospitals in data collection, and sorting, and the right, as far as possible, improve the accuracy and reliability of the data. Thirdly, we cannot exclude the potential influence of certain treatments received before admission on age, respiratory rate, heart rate, mechanical ventilation use, organ failure comorbidity, absolute lymphocyte count, ADA, LDH, NLR, SII, and chest CT outcomes. Despite these limitations. Despite these limitations, this COVID-19 severity risk prediction model offers the advantage of combined prediction, allowing for a more comprehensive and systematic assessment of the severity of COVID-19 patients. In this regard, we can carry out early medical history tracking when collecting patients' data in the later stage, understand the basic situation of patients before admission in detail, reduce some unnecessary influencing factors as much as possible, and make the research results more accurate and reliable.

In conclusion, the utilization of 12 patient features at the time of their visit can be used to generate a single variable, and the risk score from the line chart helps predict an individual's risk of severity in COVID-19. We also confirmed during the model-building process that the combined use of NLR, RR, and LDH can enhance the predictive efficiency of COVID-19. Using the severity prediction model and assessing relevant parameters aids in identifying severe COVID-19 patients. Early medical intervention and support for these high-risk patients may help reduce the severity and mortality rates of this disease.

Conclusion
This study found significant differences in RR, NLR, and LDH between severe and non-severe COVID-19 patients and demonstrated an enhanced predictive efficiency when combining NLR, RR, and LDH. A nomogram model was constructed by integrating patients' clinical characteristics, laboratory tests, and imaging findings. The calibration plot and DCA analysis showed that the nomogram had better clinical benefit and utility in predicting the severity of COVID-19. It may assist healthcare providers in the early identification of severe cases and the timely implementation of effective treatments.

Data availability
We feel great thanks for your professional review work on our manuscript. The full data set used in this study is available on reasonable request from the corresponding author at zhaoqian604@126.com. As for the conditions of data use, we want the demander to indicate the way and purpose of data use. At the same time, we may require

the demander to keep the data strictly confidential, and only use it for relevant research. The final decision to give the data was made after review by the corresponding author.

References

1. The species Severe acute respiratory syndrome-related coronavirus. classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **5**, 536–544. https://doi.org/10.1038/s41564-020-0695-z (2020).
2. Roland, L. T., Gurrola, J. G. 2nd., Loftus, P. A., Cheung, S. W. & Chang, J. L. Smell and taste symptom-based predictive model for COVID-19 diagnosis. *Int. Forum. Allergy Rhinol.* **10**, 832–838. https://doi.org/10.1002/alr.22602 (2020).
3. Chen, J. *et al.* Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography. *Sci. Rep.* **10**, 19196. https://doi.org/10.1038/s41598-020-76282-0 (2020).
4. Albataineh, Z., Aldrweesh, F. & Alzubaidi, M. A. COVID-19 CT-images diagnosis and severity assessment using machine learning algorithm. *Cluster Comput.* https://doi.org/10.1007/s10586-023-03972-5 (2023).
5. Wynants, L. *et al.* Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *BMJ* **369**, m1328. https://doi.org/10.1136/bmj.m1328 (2020).
6. Zinellu, A. *et al.* Blood cell count derived inflammation indexes in patients with idiopathic pulmonary fibrosis. *Lung* **198**, 821–827. https://doi.org/10.1007/s00408-020-00386-7 (2020).
7. Putzu, C. *et al.* Blood cell count indexes as predictors of outcomes in advanced non-small-cell lung cancer patients treated with Nivolumab. *Cancer Immunol. Immunother.* **67**, 1349–1353. https://doi.org/10.1007/s00262-018-2182-4 (2018).
8. Paliogiannis, P. *et al.* Associations between the neutrophil-to-lymphocyte and the platelet-to-lymphocyte ratios and the presence and severity of psoriasis: A systematic review and meta-analysis. *Clin. Exp. Med.* **19**, 37–45. https://doi.org/10.1007/s10238-018-0538-x (2019).
9. Paliogiannis, P. *et al.* The neutrophil-to-lymphocyte ratio as a marker of chronic obstructive pulmonary disease and its exacerbations: A systematic review and meta-analysis. *Eur. J. Clin. Invest.* **48**, e12984. https://doi.org/10.1111/eci.12984 (2018).
10. Yang, A. P., Liu, J. P., Tao, W. Q. & Li, H. M. The diagnostic and predictive role of NLR, d-NLR and PLR in COVID-19 patients. *Int. Immunopharmacol.* **84**, 106504. https://doi.org/10.1016/j.intimp.2020.106504 (2020).
11. Como, N. *et al.* Severity and mortality predictors of COVID-19 patients with thrombotic events-evidence from the "covid-one" hospital in Albania. *Vaccines (Basel)* https://doi.org/10.3390/vaccines10111851 (2022).
12. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan. *China. Lancet* **395**, 497–506. https://doi.org/10.1016/s0140-6736(20)30183-5 (2020).
13. Liang, W. *et al.* Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients With COVID-19. *JAMA Intern. Med.* **180**, 1081–1089. https://doi.org/10.1001/jamainternmed.2020.2033 (2020).
14. Knight, S. R. *et al.* Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO clinical characterisation protocol: Development and validation of the 4C mortality Score. *Bmj* **370**, m3339. https://doi.org/10.1136/bmj.m3339 (2020).
15. Salto-Alejandre, S. *et al.* A quick prediction tool for unfavourable outcome in COVID-19 inpatients: Development and internal validation. *J. Infect.* **82**, e11–e15. https://doi.org/10.1016/j.jinf.2020.09.023 (2021).
16. Xiao, L. S. *et al.* Development and validation of a deep learning-based model using computed tomography imaging for predicting disease severity of coronavirus disease 2019. *Front. Bioeng. Biotechnol.* **8**, 898. https://doi.org/10.3389/fbioe.2020.00898 (2020).
17. Chen, Y. *et al.* A Multimodality machine learning approach to differentiate severe and nonsevere COVID-19: Model development and validation. *J. Med. Internet. Res.* **23**, e23948. https://doi.org/10.2196/23948 (2021).
18. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *Lancet* **395**, 1054–1062. https://doi.org/10.1016/s0140-6736(20)30566-3 (2020).
19. Chen, L. *et al.* Risk factors for death in 1859 subjects with COVID-19. *Leukemia* **34**, 2173–2183. https://doi.org/10.1038/s41375-020-0911-0 (2020).
20. Fois, A. G. *et al.* The systemic inflammation index on admission predicts in-hospital mortality in COVID-19 Patients. *Molecules* https://doi.org/10.3390/molecules25235725 (2020).
21. Zhang, H. *et al.* Age-related risk factors and complications of patients with COVID-19: A population-based retrospective study. *Front. Med. (Lausanne)* **8**, 757459. https://doi.org/10.3389/fmed.2021.757459 (2021).
22. Clift, A. K. *et al.* Smoking and COVID-19 outcomes: an observational and Mendelian randomisation study using the UK Biobank cohort. *Thorax* **77**, 65–73. https://doi.org/10.1136/thoraxjnl-2021-217080 (2022).
23. Mehra, M. R., Desai, S. S., Kuy, S., Henry, T. D. & Patel, A. N. Retraction: Cardiovascular disease, drug therapy, and mortality in covid-19. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMc2021225 (2020).
24. Gong, J. *et al.* A tool for early prediction of severe coronavirus disease 2019 (COVID-19): A multicenter study using the risk nomogram in Wuhan and Guangdong China. *Clin. Infect. Dis.* **71**, 833–840. https://doi.org/10.1093/cid/ciaa443 (2020).
25. Itelman, E. *et al.* Clinical characterization of 162 COVID-19 patients in Israel: Preliminary report from a large tertiary center. *Isr Med. Assoc. J.* **22**, 271–274 (2020).
26. Chan, A. S. & Rout, A. Use of neutrophil-to-lymphocyte and platelet-to-lymphocyte ratios in COVID-19. *J. Clin. Med. Res.* **12**, 448–453. https://doi.org/10.14740/jocmr4240 (2020).
27. Huguet, E. *et al.* reference values for neutrophil to lymphocyte ratio (NLR), a biomarker of cardiovascular risk, according to age and sex in a Latin American population. *Curr. Probl. Cardiol.* **46**, 100422. https://doi.org/10.1016/j.cpcardiol.2019.04.002 (2021).
28. Wang, X., Qiu, L., Li, Z., Wang, X. Y. & Yi, H. Understanding the multifaceted role of neutrophils in cancer and autoimmune diseases. *Front. Immunol.* **9**, 2456. https://doi.org/10.3389/fimmu.2018.02456 (2018).
29. Zeng, Z. Y., Feng, S. D., Chen, G. P. & Wu, J. N. Predictive value of the neutrophil to lymphocyte ratio for disease deterioration and serious adverse outcomes in patients with COVID-19: A prospective cohort study. *BMC Infect. Dis.* **21**, 80. https://doi.org/10.1186/s12879-021-05796-3 (2021).
30. Ponti, G., Maccaferri, M., Ruini, C., Tomasi, A. & Ozben, T. Biomarkers associated with COVID-19 disease progression. *Crit. Rev. Clin. Lab. Sci.* **57**, 389–399. https://doi.org/10.1080/10408363.2020.1770685 (2020).
31. Seyit, M. *et al.* Neutrophil to lymphocyte ratio, lymphocyte to monocyte ratio and platelet to lymphocyte ratio to predict the severity of COVID-19. *Am. J. Emerg. Med.* **40**, 110–114. https://doi.org/10.1016/j.ajem.2020.11.058 (2021).
32. Tatum, D. *et al.* Neutrophil-to-lymphocyte ratio and outcomes in Louisiana COVID-19 patients. *Shock* **54**, 652–658. https://doi.org/10.1097/shk.0000000000001585 (2020).
33. Prozan, L. *et al.* Prognostic value of neutrophil-to-lymphocyte ratio in COVID-19 compared with Influenza and respiratory syncytial virus infection. *Sci. Rep.* **11**, 21519. https://doi.org/10.1038/s41598-021-00927-x (2021).

Author contributions

F.H.Q. and Y.C. contributed significantly to the conceptualization and design of this work. F.H.Q. and Y.C. contributed equally to this manuscript. Y.C. and Y.X.L. carried out the data collection and data analysis. Y.C. pre-processed the data and participated in drafting and revising the manuscript. F.H.Q. critically revised the manuscript for important content and finally approved the manuscript for publication. J.H. and R.H.Z. confirmed the authenticity of all original data. All authors have read and approved the final version of the manuscript.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to F.-H.Q.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.