Posted on 14 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167879588.86752560/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Development and validation of a model for the prediction of the risk of pneumonia in patients with SARS-CoV-2 infection

Xi Yi[1], Daiyan Fu[1], Guiliang Wang[1], Lile Wang[1], and Jirong LI[1]

[1]Hunan Provincial People's Hospital

March 14, 2023

Abstract

[Abstract] Objective: To develop a pneumonia risk prediction model for SARS-CoV-2 infected patients to reduce unnecessary chest CT scans; **Materials and Methods:** Retrospective analysis was performed on the clinical data of SARS-CoV-2-positive patients who visited outpatient and emergency clinics and underwent chest CT scans at the Mawangdui Branch of Hunan Provincial People's Hospital from 20 December 2022 to 23 December 2022 and at the Tianxinge Branch of Hunan Provincial People's Hospital from 1 January 2023 to 4 January 2023. A retrospective analysis of imaging and clinical data from 205 cases (training cohort) and 94 cases (validation cohort) of SARS-CoV-2-positive patients who visited outpatient and emergency clinics was conducted. The predictor variables were screened using the "univariate and then multivariate logistic regression" and "least absolute shrinkage and selection operator (LASSO)" approaches, and the predictive model was constructed using multifactorial logistic regression and represented as a nomogram. The diagnostic effectiveness of the pneumonia risk model was evaluated using receiver operating characteristic (ROC) curves; the Delong test and Integrated Discrimination Improvement Index (IDI) were used to compare the AUC of the pneumonia risk model with the AUCs for predictors incorporated in the model alone. The calibration of the pneumonia risk model was assessed using calibration curves; Decision curve analysis (DCA) was used to evaluate the clinical validity of the pneumonia risk model. In addition, a smoothed curve was fitted using a generalized additive model (GAM) to explore the relationship between the pneumonia grade and the model's predicted probability of pneumonia; **Results:** "univariate and then multivariate logistic regression " and Lasso regression together show that age, natural log-transformed value (lnCRP), Monocytes percentage (%Mon) are valid predictors of pneumonia risk; the AUC of the pneumonia risk model was 0.7820 (95% CI: 0.7254-0.8439) in the training cohort and 0.8432 (95% CI: 0.7588-0.9151) in the validation cohort; at the cut-off value of 0.5, the sensitivity and specificity of the pneumonia risk model were 70.75%, 66.33% (training cohort), 76.09%, and 73.91% (validation cohort), the calibration curves showed that the pneumonia risk model has good calibration accuracy. The decision curve analysis showed that the pneumonia risk model has high clinical value in predicting the probability of pneumonia in SARS-CoV-2 infected patients. **Conclusion:** The pneumonia risk prediction model developed in this study can be used to predict the risk of pneumonia in SARS-CoV-2 infected patients diagnostically.

Development and validation of a model for the prediction of the risk of pneumonia in patients with SARS-CoV-2 infection

Xi Yi[1] ,Daiyan Fu[2],Guiliang Wang[1],Lile Wang[2] ,Jirong LI[1*]

[1]Department of Radiology, Hunan Provincial People's Hospital/The First Affiliated Hospital of Hunan Normal University, Changsha 410016, China;

[2]Department of Respiratory Medicine, Hunan Provincial People's Hospital/The First Affiliated Hospital of Hunan Normal University, Changsha 410016, China;

[*]*Correspondence to: Jirong LI, E-mail:304353448@qq.com*

[Abstract] Objective: To develop a pneumonia risk prediction model for SARS-CoV-2 infected patients to

Posted on 14 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167879588.86752560/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

reduce unnecessary chest CT scans; **Materials and Methods:**Retrospective analysis was performed on the clinical data of SARS-CoV-2-positive patients who visited outpatient and emergency clinics and underwent chest CT scans at the Mawangdui Branch of Hunan Provincial People's Hospital from 20 December 2022 to 23 December 2022 and at the Tianxinge Branch of Hunan Provincial People's Hospital from 1 January 2023 to 4 January 2023. A retrospective analysis of imaging and clinical data from 205 cases (training cohort) and 94 cases (validation cohort) of SARS-CoV-2-positive patients who visited outpatient and emergency clinics was conducted. The predictor variables were screened using the "univariate and then multivariate logistic regression" and "least absolute shrinkage and selection operator (LASSO)" approaches, and the predictive model was constructed using multifactorial logistic regression and represented as a nomogram. The diagnostic effectiveness of the pneumonia risk model was evaluated using receiver operating characteristic (ROC) curves; the Delong test and Integrated Discrimination Improvement Index (IDI) were used to compare the AUC of the pneumonia risk model with the AUCs for predictors incorporated in the model alone. The calibration of the pneumonia risk model was assessed using calibration curves; Decision curve analysis (DCA) was used to evaluate the clinical validity of the pneumonia risk model. In addition, a smoothed curve was fitted using a generalized additive model (GAM) to explore the relationship between the pneumonia grade and the model's predicted probability of pneumonia; **Results:**"univariate and then multivariate logistic regression " and Lasso regression together show that age, natural log-transformed value (InCRP), Monocytes percentage (%Mon) are valid predictors of pneumonia risk; the AUC of the pneumonia risk model was 0.7820 (95% CI: 0.7254-0.8439) in the training cohort and 0.8432 (95% CI: 0.7588-0.9151) in the validation cohort; at the cut-off value of 0.5, the sensitivity and specificity of the pneumonia risk model were 70.75%, 66.33% (training cohort), 76.09%, and 73.91% (validation cohort), the calibration curves showed that the pneumonia risk model has good calibration accuracy. The decision curve analysis showed that the pneumonia risk model has high clinical value in predicting the probability of pneumonia in SARS-CoV-2 infected patients.**Conclusion:** The pneumonia risk prediction model developed in this study can be used to predict the risk of pneumonia in SARS-CoV-2 infected patients diagnostically.

【**Key words**】 SARS-CoV-2;COVID-19;prediction model;nomogram;pneumonia

INTRODUCTION

SARS-CoV-2 infection has spread globally since 2020, leading many countries to impose recurring quarantines, significantly impacting public health and the global economy[1-2]. Globally, as of 10 February 2023, there have been 755,385,709 confirmed cases of COVID-19, including 6,833,388 deaths, reported to WHO. Omicron, the mutant strain, entered the community in November 2021 and is far more contagious and escape-resistant than the previous variants of concern (VOC), like Delta[3-8]. At the beginning of 2022, the Omicron version quickly surpasses the Delta variant as the prevalent strain worldwide[9].

During the early period of the COVID-19 pandemic, SARS-CoV-2 primarily affected the lung and caused pneumonia[10-13]. As one of the most representative and accurate diagnostic methods for COVID-19[14], chest computed tomography (CT) scans are widely used in mainland China.

However, recent studies have demonstrated that the most recent VOC Omicron variant is much less likely to cause pulmonary infections[3-5,15-16], suggesting potential implications for adapting management strategies for these infections.

In clinical practice, we found that due to the apprehension of contracting severe pneumonia from the SARS-CoV-2, many people with mild symptoms are choosing to receive CT scans, causing excessive CT scans and putting a strain on the availability of healthcare resources, which is particularly true when SRAS-CoV-2 localized epidemic outbreaks occur. Consequently, a strategy to evaluate the risk of pneumonia among recently infected people is essential to ensure the efficient use of healthcare resources and decrease unnecessary exposure to electromagnetic radiation.

With the aim of improving the classification of the risk of pneumonia in individuals with the most recent VOC of SARS-CoV-2 infections, reducing the overuse of CT scans, reducing non-essential ionizing radiation in individuals, as well as reducing the associated financial burden on patients, and optimizing the allocation

of healthcare resources, we have developed and externally validated a pneumonia risk prediction model based on general patient data and blood routine test, to meet the needs of the new phase of the COVID-19 epidemic control.

Material and Methods

Materials

A retrospective analysis was performed on the clinical data of SARS-CoV-2-positive patients who visited outpatient and emergency clinics and underwent chest CT scans at the Mawangdui Branch of Hunan Provincial People's Hospital from 20 December 2022 to 23 December 2022 and at the Tianxinge Branch of Hunan Provincial People's Hospital from 1 January 2023 to 4 January 2023.Inclusion criteria: (1) Attendance as an outpatient or emergency (not including inpatients); (2) Patients had completed chest CT scans, and CT image quality meets diagnostic requirements; (3) SARS-CoV-2 infection positive was diagnosed by antigen test or nucleic acid test within 3 days before the current chest CT; (4) Complete blood routine examination results. Exclusion criteria: (1) Inflammation of a body part other than the lungs has been diagnosed at the time of the current blood routine tests;(2) the Patient was already on antiviral medication at the time of the visit. The patient recruitment pathway is detailed in FIGURE 1. The study complies with the Declaration of Helsinki. It was approved by the Medical Ethics Committee of Hunan Provincial People's Hospital (The First Affiliated Hospital of Hunan Normal University), exempting the subjects from informed consent.

Methods

Device parameters and image analysis

The Mawangdui Branch (training cohort) used CT scans with a field of view (FOV) of 230 mm × 230 mm, a layer thickness of 5 mm, and layer spacing of 5 mm using the United Imaging uCT 760GE 128-slice CT; the Tianxinge Branch (validation cohort) used CT scans with a field of view (FOV) of 230 mm × 230 mm, a layer thickness of 5 mm, and layer spacing of 5 mm using the United Imaging uCT 860 160-slice CT or United Imaging uCT 960+ 640-slice CT. Two attending radiologists performed image analysis separately, and the final decision in case of a dispute was determined by consultation between the two physicians. CT Diagnosis of COVID-19 was referred to the report published by the RSNA[17]; typical findings were as follows: peripheral distribution, ground-glass opacity, fine reticular opacity, vascular thickening, and reverse halo sign. Patients with pneumonia were also classified into grades 0, 1, 2, 3, and 4 according to the extent and distribution of lung involvement (no lung involvement was categorized as grade 0).

Statistical analysis & construction and evaluation of predictive models

Statistical analysis was performed using Empower Stats, version 5.0 (http://www.empowerstats.com, X&Y Solutions, Inc., Boston, MA, USA), R statistical software, version 4.2.0 (http://www.R-project.org, The R Foundation), and the SPSS statistical software, version 27.0 (SPSS Inc., Chicago, IL, USA) with continuity variables expressed as medians (min, max) and categorical variables expressed as frequencies (percentages). Kruskal Wallis rank sum test or Fisher exact probability test was used to compare differences between groups of continuity variables. Chi-square tests are used for comparisons of categorical variables. After the natural log transformation of some continuity variables, to reduce irrelevant and redundant information, the predictor variables of the training cohort are filtered by both " univariate and then multivariate logistic regression " and " least absolute shrinkage and selection operator (LASSO)" methods, the variables selected by both screening methods were used as the final predictor variables. The prediction model was constructed using multivariate logistic regression and presented in a nomogram. The ROC curves were used, and 500 in eternal resamples were performed by Bootstrap to evaluate the discrimination of the pneumonia risk model between the training and validation cohorts. Delong test and Integrated Discrimination Improvement Index (IDI) were used to compare the AUC of the pneumonia risk model with the AUCs for predictors incorporated in the model alone. Calibration curves were plotted to assess the calibration of the model. The clinical validity of the model was evaluated by the net benefit of DCA at different threshold probabilities; in addition, a smoothed curve was fitted using a generalized additive model (GAM) to explore the relationship between

Posted on 14 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167879588.86752560/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on 14 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167879588.86752560/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

the pneumonia grade and the model's predicted probability of pneumonia; a difference of $p < 0.05$ was considered statistically significant.

Results

General information.

A total of 205 cases were enrolled in the training cohort, with a median age of 47 years, the youngest being 14 years and the oldest 97 years, of which 105 cases (51.22%) were female and 100 cases (48.78%) were male, 99 cases (48.29%) without pneumonia and 106 cases (51.71%) with pneumonia; a total of 94 cases were enrolled in the validation cohort, with a median age of 56 years, the youngest being 2 years and the oldest 89 years, of which 60 (63.83%) were female and 34 (36.17%) were male, 47 (50.00%) were without pneumonia, and 47 (50.00%) were with pneumonia; the distribution of the remaining baseline indicators is shown in TABLE 1.

TABLE 1. Baseline indicators in the Training cohort and Validation cohort.

Characteristic	Training cohort	Training cohort	Training cohort	Validation cohort	Validation
	No pneumonia	Incident pneumonia	P-value	No pneumonia	Incident pn
participants	99	106		47	47
Age(year)	32.00 (14.00-86.00)	61.50 (17.00-97.00)	<0.001	39.00 (17.00-79.00)	69.00 (2.00-
Gender			0.718		
Female	52 (52.53%)	53 (50.00%)		29 (61.70%)	31 (65.96%
Male	47 (47.47%)	53 (50.00%)		18 (38.30%)	16 (34.04%
CRP(mg/L)	6.82 (0.00-154.30)	18.60 (0.47-359.14)	<0.001	1.66 (0.00-115.27)	28.21 (0.00-
WBC(10^9/L)	5.40 (1.41-10.36)	5.53 (2.36-15.56)	0.038	7.35 (2.43-9.42)	6.82 (2.54-1
RBC(10^{12}/L)	4.86 (3.31-6.44)	4.79 (2.78-6.55)	0.178	4.90 (3.84-6.29)	4.46 (3.30-5
HGB(g/L)	146.00 (67.00-184.00)	145.00 (93.00-192.00)	0.224	141.00 (116.00-173.00)	132.00 (100
PLT(10^9/L)	185.00 (98.00-361.00)	173.00 (51.00-460.00)	0.090	269.00 (136.00-424.00)	221.00 (59.0
%Neu(%)	69.10 (29.00-89.30)	70.65 (38.80-94.10)	0.019	61.30 (44.70-82.10)	71.40 (44.70
%Lymph(%)	19.70 (4.40-60.60)	20.00 (2.40-53.80)	0.189	30.70 (8.90-47.60)	19.40 (5.80-
%Mon(%)	9.40 (3.70-29.50)	7.90 (1.40-20.10)	<0.001	7.00 (3.40-14.50)	7.20 (1.50-1
%Eos(%)	0.60 (0.00-8.00)	0.25 (0.00-5.90)	0.016	1.20 (0.00-11.20)	0.70 (0.00-3
#Eos(10^9/L)	0.03 (0.00-0.35)	0.01 (0.00-0.36)	0.057	0.08 (0.00-0.91)	0.04 (0.00-0
%Bas(%)	0.20 (0.00-7.50)	0.10 (0.00-0.80)	0.002	0.20 (0.00-0.70)	0.10 (0.00-0
#Bas(10^9/L)	0.01 (0.00-0.32)	0.01 (0.00-0.03)	0.049	0.01 (0.00-0.04)	0.01 (0.00-0
#Neu(10^9/L)	3.45 (0.72-8.25)	3.75 (0.91-14.36)	0.012	4.26 (1.47-6.85)	4.67 (1.46-1
#Lymph(10^9/L)	1.05 (0.23-3.56)	1.06 (0.24-5.47)	0.864	2.05 (0.73-3.34)	1.19 (0.55-3
#Mon(10^9/L)	0.48 (0.19-1.50)	0.43 (0.08-1.51)	0.216	0.48 (0.17-0.74)	0.52 (0.14-1
HCT(%)	42.10 (22.40-52.10)	41.15 (26.00-53.30)	0.135	42.50 (36.20-52.00)	39.80 (30.9
MCV(fL)	87.60 (55.30-98.10)	87.55 (62.00-99.40)	0.597	89.30 (66.10-105.00)	89.70 (69.3
MCHC(g/L)	350.00 (301.00-369.00)	352.00 (298.00-376.00)	0.147	331.00 (311.00-352.00)	333.00 (315
MCH(Pg)	30.70 (16.70-33.90)	30.90 (20.10-35.20)	0.286	30.10 (21.00-34.40)	30.10 (22.3
RDW-SD(fL)	39.40 (33.30-55.50)	40.15 (33.70-49.60)	0.040	40.50 (36.00-48.50)	40.20 (32.5
RDW-CV(%)	12.20 (11.30-19.90)	12.40 (11.30-17.40)	0.096	12.10 (11.10-15.90)	12.20 (10.9
PDW(%)	16.20 (10.20-17.10)	16.30 (15.40-17.80)	0.014	16.20 (15.60-16.80)	16.30 (15.7
MPV(fL)	9.50 (7.30-11.90)	9.80 (8.00-12.40)	0.020	9.30 (7.70-11.70)	9.50 (7.20-1
PCT(%)	0.18 (0.10-0.34)	0.17 (0.05-0.41)	0.247	0.25 (0.14-0.36)	0.20 (0.06-0
P-LCR (%)	23.80 (8.80-41.90)	26.10 (12.40-44.90)	0.033	21.70 (12.60-39.20)	24.20 (10.0

CRP= C reactive protein; WBC = White Blood Cells; RBC = Red Blood Cells; HGB = Hemoglobin; PLT = Platelets; %Neu = Neutrophils (percentage); %Lymph = Lymphocytes (percentage); %Mon = Monocytes (percentage); %Eos = Eosinophils (percentage); #Eos = Eosinophils (number); %Bas = Basophils (per-

Posted on 14 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167879588.86752560/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

centage); #Bas = Basophils (number);#Neu = Neutrophils (number); #Lymph = Lymphocytes (number); #Mon = Monocytes (number); HCT = Hematocrit; MCV = Mean Corpuscular Volume; MCHC = Mean Corpuscular Hemoglobin Concentration; MCH = Mean Corpuscular Hemoglobin; RDW-SD = Red Cell Distribution Width-Standard Deviation; RDW-CV = Red Cell Distribution Width-Coefficient of Variation; PDW = Platelet Distribution Width; MPV = Mean Platelet Volume; PCT = Plateletcrit; P-LCR = Platelet Large Cell Ratio.

Predictor variable screening results

Among the baseline indicators in the training cohort, univariate logistic regression showed age, White Blood Cells (WBC), Red Blood Cells (RBC), Neutrophils percentage (%Neu), Neutrophils number (#Neu), Lymphocytes percentage (%Lymph), Monocytes percentage (%Mon), Red Cell Distribution Width-Standard Deviation (RDW-SD), Platelet Distribution Width (PDW), Mean Platelet Volume (MPV), Platelet Large Cell Ratio (P-LCR), CRP natural log-transformed value (InCRP), Eosinophils (percentage) (%Eos), Basophils percentage (%Bas), Basophils number (#Bas) as possible predictors($p < 0.1$), further multivariate logistic regression showed age, CRP natural log-transformed value (InCRP), Neutrophils percentage (%Neu), and Monocytes percentage (%Mon) as independent predictors ($p < 0.05$)(TABLE 2). Three predictors with non-zero coefficients were obtained by Lasso regression (screening lambda by 10-fold cross-validation, based on lambda.1se, i.e., the maximum lambda corresponding to an error mean within one standard deviation of the minimum): age, CRP natural log-transformed value (InCRP), Monocytes percentage (%Mon) (FIGURE 2). To reduce irrelevant and redundant information, the variables selected by both screening methods were taken as the final selected predictor variables: age, natural log-transformed value (InCRP), and Monocytes percentage (%Mon).

TABLE 2.Univariate and Multivariate analysis results of the training cohort.

Characteristic	Univariate analysis	Univariate analysis	Multivariate analysis	Multivariate analysis
	β (95%CI) / OR (95%CI)	P-value	β (95%CI) / OR (95%CI)	P-value
Gender		0.7178		
Female	1.0			
Male	1.11 (0.64, 1.91)			
Age(year)	1.03 (1.02, 1.04)	<0.0001	**1.03 (1.01, 1.05)**	**0.0095**
WBC(10^9/L)	1.18 (1.04, 1.35)	0.0136	1.38 (0.46, 4.09)	0.5663
RBC(10^{12}/L)	0.70 (0.45, 1.07)	0.0978	2.77 (0.68, 11.32)	0.1549
HGB(g/L)	0.99 (0.98, 1.01)	0.1937		
PLT(10^9/L)	1.00 (0.99, 1.00)	0.1753		
%Neu(%)	1.03 (1.01, 1.05)	0.0051	**0.76 (0.59, 0.99)**	**0.0415**
%Lymph(%)	0.98 (0.96, 1.00)	0.0769	0.80 (0.61, 1.04)	0.0991
%Mon(%)	0.85 (0.77, 0.93)	0.0005	**0.66 (0.50, 0.88)**	**0.0041**
#Neu(10^9/L)	1.23 (1.06, 1.42)	0.0056	0.98 (0.24, 3.96)	0.9818
#Lymph(10^9/L)	0.99 (0.65, 1.51)	0.9585		
#Mon(10^9/L)	0.53 (0.17, 1.65)	0.2730		
HCT(%)	0.95 (0.90, 1.01)	0.0989	0.88 (0.75, 1.04)	0.1332
MCV(fL)	1.01 (0.97, 1.05)	0.5854		
MCHC(g/L)	1.02 (0.99, 1.04)	0.2370		
MCH(Pg)	1.05 (0.94, 1.16)	0.3793		
RDW-SD(fL)	1.09 (1.00, 1.19)	0.0522	0.95 (0.81, 1.12)	0.5666
RDW-CV(%)	1.18 (0.90, 1.56)	0.2336		
PDW(%)	2.35 (1.15, 4.82)	0.0195	1.44 (0.65, 3.16)	0.3659
MPV(fL)	1.53 (1.11, 2.11)	0.0097	18.62 (0.98, 355.15)	0.0519
PCT(%)	0.14 (0.00, 22.97)	0.4483		
P-LCR (%)	1.05 (1.01, 1.10)	0.0178	0.70 (0.47, 1.04)	0.0773
InCRP(mg/L)	1.84 (1.46, 2.32)	<0.0001	**1.89 (1.40, 2.55)**	**<0.0001**

Characteristic	Univariate analysis	Univariate analysis	Multivariate analysis	Multivariate analysis
%Eos(%)	0.83 (0.68, 1.01)	0.0594	0.71 (0.50, 1.03)	0.0691
#Eos(10^9/L)	0.09 (0.00, 3.02)	0.1814		
%Bas(%)	0.08 (0.01, 0.49)	0.0058	2.95 (0.97, 90.23)	0.5346
#Bas(10^9/L)	0.00 (0.00, 0.01)	0.0249	0.00 (0.00, inf.)	0.3203

WBC = White Blood Cells; RBC = Red Blood Cells; HGB = Hemoglobin; PLT = Platelets; %Neu = Neutrophils (percentage); %Lymph = Lymphocytes (percentage); %Mon = Monocytes (percentage); #Neu = Neutrophils (number); #Lymph = Lymphocytes (number); #Mon = Monocytes (number); HCT = Hematocrit; MCV = Mean Corpuscular Volume; MCHC = Mean Corpuscular Hemoglobin Concentration; MCH = Mean Corpuscular Hemoglobin; RDW-SD = Red Cell Distribution Width-Standard Deviation; RDW-CV = Red Cell Distribution Width-Coefficient of Variation; PDW = Platelet Distribution Width; MPV = Mean Platelet Volume; PCT = Plateletcrit; P-LCR = Platelet Large Cell Ratio; InCRP = natural log-transformed value of CRP; %Eos = Eosinophils (percentage); #Eos = Eosinophils (number); %Bas = Basophils (percentage); #Bas = Basophils (number).

Construction and evaluation of the nomogram prediction model.

Multivariable logistic regression analysis established a nomogram model based on the final selected predictor variables(FIGURE 3. A). The AUC of the pneumonia risk model was 0.7820 (95% CI: 0.7254-0.8439) in the training cohort and 0.8432 (95% CI: 0.7588-0.9151) in the validation cohort(FIGURE 3. B, C); at the cut-off value of 0.5, the sensitivity and specificity of the pneumonia risk model were 70.75%, 66. 33% (training cohort), 76.09%, and 73.91% (validation cohort), respectively; the calibration curve showed good agreement between the predicted probability of pneumonia from the pneumonia risk model and the actually observed probability. Decision curve analysis (DCA) showed good clinical validity of the pneumonia risk model in the training and validation cohort (FIGURE 3. F, G). Other diagnostic parameters of the model are shown in TABLE 3. Comparison of the AUC and DCA for the pneumonia risk model with predictors incorporated in the model alone in the whole study cohort were shown in FIGURE 4, showing that the pneumonia risk model combining multiple predictors has better diagnostic performance than a single predictor.

TABLE 3. Diagnostic parameters ofthe pneumonia risk model.

Variable	Value	Value
	Training cohort	Validation cohort
AUC	0.7820(95%CI:0.7254,0.8439)	0.8432 (95%CI:0.7588,0.9151)
Cutoff value	0.5	0.5
specificity	66.33%	73.91%
sensitivity	70.75%	76.09%
accuracy	68.63%	75.00%
positive-LR	2.10	2.92
negative-LR	0.44	0.32
diagnose-OR	4.77	9.02
positive-pv	69.44%	74.47%
negative-pv	67.71%	75.56%

Correlation between the predicted probability of pneumonia risk and pneumonia grade

We further explored the correlation between the predictive values of the pneumonia risk prediction model constructed in this study and the actual pneumonia severity rating. As mentioned in the method, patients with pneumonia were also classified into grades 0, 1, 2, 3, and 4 according to the extent and distribution of lung involvement (no lung involvement was categorized as grade 0); the actual pneumonia grading results are

Posted on 14 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167879588.86752560/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

shown in TABLE 4. A positive linear correlation was found between the predicted pneumonia probability of the pneumonia risk model and actual pneumonia grade using GAM(FIGURE 5); see FIGURE 6 for examples.

TABLE 4. Distribution of actual pneumonia grades in the training cohort and validation cohort

Actual Pneumonia grade	Participants n(%)	Participants n(%)
	Training cohort	Validation cohort
0	99 (48.29%)	47 (50.00%)
1	70 (34.15%)	22 (23.40%)
2	23 (11.22%)	18 (19.15%)
3	12 (5.85%)	6 (6.38%)
4	1 (0.49%)	1 (1.06%)

Discussion

As SARS-CoV-2 continues to evolve, the lung pathogenicity of the emerging VOCs continues to decline[3-5,15]. As the primary screening test for pneumonia, CT scans play an essential role in the early stages of the epidemic of SARS-CoV-2[18-19]. However, with the reduced lung pathogenicity of the new mutant strain, a system is required to evaluate the risk of pneumonia in recently infected individuals to ensure the effective use of healthcare resources and minimize unnecessary exposure to electromagnetic radiation.

The present study was designed to develop a model for pneumonia risk prediction in patients with SRAS-CoV-2 infection, for classifying the risk of pneumonia in SARS-CoV-2 infected patients, to provide clinicians with an appropriate reference for selecting CT scans by predicting the risk of pneumonia in subjects before chest CT scans, to reduce non-essential medical ionizing radiation and reduce the financial burden on patients.

The pneumonia risk model constructed in this study shows good discrimination, calibration, and clinical validity. In addition, the predictors used in the model are "age" and "blood routine indicators", which are very common, readily available, and inexpensive. This provides a reasonable basis for promoting the use of the model.

To reduce irrelevant and redundant information, we used both "univariate and multivariate logistic regression" and "Lasso regression" to screen for predictor variables; the variables selected for both options were taken as the final predictors, namely: age, InCRP, %Mon.

Previous reports have shown that the severity and fatality rates of COVID-19 significantly vary with age group, and they rise sharply in the elderly[20-22]; this supports the age predictor's inclusion in the pneumonia risk prediction model.

As a general indicator of inflammation, CRP is associated with the clinical severity of COVID-19[22-24]. CRP may indicate COVID-19 changes earlier than chest CT — CRP was significantly elevated before CT findings in severe COVID-19 patients[25].

In our study, %Mon was partially associated with the risk of pneumonia, which is in accord with recent studies[26]. Monocytes are innate immune system cells that participate in several immune function events, including phagocytosis, antigen presentation, and inflammatory responses[27]; circulating monocytes extravasate into peripheral tissues during sterile and non-sterile inflammation and undergo differentiation into macrophages or dendritic cells. A previous review article discussed the buildup of monocyte/macrophage cells in the lungs. These cells are likely sources of the proinflammatory cytokines and chemokines linked to deadly diseases brought on by human coronavirus infections, such as COVID-19[28], suggesting that the migration of monocytes into lung tissue may be the cause of the monocyte reduction in peripheral blood.

In previous related studies, additional factors such as cardiovascular disease, chronic respiratory disease, diabetes, obesity, hypertension, and high serum ferritin levels, were found to be associated with the progression

Posted on 14 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167879588.86752560/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on 14 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167879588.86752560/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

of COVID-19[29-31]. Since our study was retrospective, it is limited by missing information, and some of the valuable indicators reported by related studies were not included in this study. In addition, some of the indicators were not included in our study because they were derived from patients' complaints rather than standard medical diagnoses and had low credibility.

From the standpoint of model promotion, the more streamlined model predictions are less expensive, easier to use, and more suited to wider use, but they also result in a decline in model prediction performance.

This is a matter of balance, depending on the application scenario of the model being constructed: whether it should be applied primarily for primary screening of high-risk cases or whether it prefers higher predictive accuracy.

In our study, the pneumonia risk prediction model we constructed was mainly applied to the primary screening of people at high risk of pneumonia in SARS-CoV-2 infected individuals, so we chose a more streamlined modeling strategy.

One unexpected finding was that the model performed better in the validation cohort than in the training cohort. This result may be explained by the relatively small sample size of the validation cohort and a certain degree of homology with the training cohort.

Limitations of this study

Our study has several limitations.

First, despite applying the inclusion criteria strictly, we could not completely rule out cases with potential lesions in body parts other than the lungs at study entry from influencing the predictors. This created some confusion during the model's development and some difficulty in evaluating its predictive performance.

Second, even though external validation was carried out, the cohort for it came from just one center, and the sample size was somewhat tiny.

In later research, a larger sample size would be required to calibrate and validate the model in a multicenter population.

AUTHOR CONTRIBUTIONS

Xi Yi and Jirong Li conceived and designed the study, had full access to all the data in the study and took responsibility for the data's integrity and the data analysis's accuracy. Daiyan Fu and Lile Wang contributed to the modification and revision of the manuscript; Guiliang Wang evaluated the quality of the literature. Xi Yi wrote the manuscript. All listed authors reviewed and approved the final manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All data relevant to the study are included in the article.

ORCID

Xi Yi*https://orcid.org/0000-0001-8222-0557*

Jirong Li *https://orcid.org/0009-0003-2991-3540*

Figure legends:

FIGURE 1. Instructions for enrolling in the training cohort and validation cohort cases.

FIGURE 2. A: Lasso regression coefficient path diagram; B: Lasso regression cross-validation curve. Three predictors with non-zero coefficients were obtained by Lasso regression (screening lambda by 10-fold cross-validation, based on lambda.1se, i.e., the maximum lambda corresponding to an error mean within

one standard deviation of the minimum): age, CRP natural log-transformed value (InCRP), Monocytes percentage (%Mon).

FIGURE 3. A: Nomogram of the pneumonia risk model; B-G: ROC curves (Bootstrap=500 times), calibration curves, and DCA curves of the pneumonia risk model in the training and validation cohorts. The ROC curves show good discrimination of the pneumonia risk model in both the training and validation cohorts. The calibration curves showed that the pneumonia risk model has good calibration accuracy. The decision curve analysis showed that the pneumonia risk model has high clinical value in predicting the probability of pneumonia in SARS-CoV-2 infected patients.

FIGURE 4. Comparison of the models in the whole study cohort.

A. Receiver operator characteristic curves of the models are presented to compare their discriminatory accuracy for predicting pneumonia risk. P values show the AUC for the pneumonia risk model versus the AUCs for predictors incorporated in the model alone; the predictive ability of the predictors in the model individually and the overall predictive power of the pneumonia risk model are contrasted via IDI.

B. Decision curve analyses comparing the net benefit of the nomogram of the pneumonia risk model versus the other variables incorporated in the nomogram alone are shown. AUC: area under the curve; CI: confidence interval; IDI: Integrated discrimination improvement.

FIGURE 5. Correlation between the predicted probability of pneumonia risk and pneumonia grade(A: Training cohort; B: Validation cohort). A positive linear correlation was found between the predicted pneumonia probability of the pneumonia risk model and pneumonia grade using GAM.

FIGURE 6.

A-B: A 32-year-old male presented with a 1-day history of fever with a maximum temperature of 39.2°C. At the time of presentation, he was confirmed positive by nucleic acid testing for SARS-CoV-2. His routine blood test showed a CPR of 14.08 (InCRP=2.70) and %Mon of 26.50. Combined with his age of 32 years, the patient had total points of 75 according to our pneumonia risk prediction model, with a pneumonia risk prediction probability of <0.1. The patient underwent a CT chest scan, which showed no abnormal findings.

C-D: Male, 17 years old, presented 4 days ago with a fever with a maximum temperature of 39.0°C. On presentation, he was confirmed positive by nucleic acid testing for SRRS-CoV-2. His routine blood test showed a CPR of 82.45 (InCRP=4.41) with a %Mon of 8.30. Combined with his age of 17 years, the patient had total points of 152 according to our pneumonia risk prediction model, with a pneumonia risk prediction probability of 0.68. The patient underwent a chest CT, which showed multiple lamellar ground-glass opacities in the lower lobe of the left lung, with a peripheral distribution and thickened blood vessels within the lesion.

E-F: A 63-year-old male with a 1-week history of malaise was confirmed to be nucleic acid test positive for SARS-CoV-2 on presentation. His routine blood test showed a CRP of 259.68 (InCRP=5.56) with a %Mon of 5.00. Combined with his age of 63 years, this patient had total points of 192 according to our pneumonia risk prediction model, with a pneumonia risk prediction probability of >0.9. The patient underwent a chest CT, which showed multiple lamellar hyperintensities in multiple lobes of both lungs with solid lesion density, bronchial air sign within, and halo sign at the edges of some lesions.

References:

1. Rabiul Islam M, Nasreen W, Anjum R, et al. Characteristics of the SARS-CoV-2 Omicron (B.1.1.529) Variant and Emerging Impact on Global Public Health. Clin Pathol. 2022;15:2632010X221124908

2. Miller IF, Becker AD, Grenfell BT, Metcalf C. Disease and healthcare burden of COVID-19 in the United States. Nat Med. 2020;26(8):1212-1217

3. Suzuki R, Yamasoba D, Kimura I, et al. Attenuated fusogenicity and pathogenicity of SARS-CoV-2 Omicron variant. Nature. 2022;603(7902):700-705

Posted on 14 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167879588.86752560/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on 14 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167879588.86752560/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

4. Yuan S, Ye ZW, Liang R, et al. Pathogenicity, transmissibility, and fitness of SARS-CoV-2 Omicron in Syrian hamsters. Science. 2022;377(6604):428-433

5. Bálint G, Vörös-Horváth B, Széchenyi A. Omicron: increased transmissibility and decreased pathogenicity. Signal Transduct Target Ther. 2022;7(1):151

6. Grabowski F, Kochańczyk M, Lipniacki T. The Spread of SARS-CoV-2 Variant Omicron with a Doubling Time of 2.0-3.3 Days Can Be Explained by Immune Evasion. Viruses. 2022;14(2)

7. Vitiello A, Ferrara F, Auti AM, Di Domenico M, Boccellino M. Advances in the Omicron variant development. J Intern Med. 2022;292(1):81-90

8. SARS-CoV-2 B.1.1.529 (Omicron) Variant - United States, December 1-8, 2021. MMWR Morb Mortal Wkly Rep. 2021;70(50):1731-1734

9. Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern (WHO, 2021) .

10. Zhang J, Rao X, Li Y, et al. Pilot trial of high-dose vitamin C in critically ill COVID-19 patients. Ann Intensive Care. 2021;11(1):5

11. Sahin G, Akbal-Dagistan O, Culha M, et al. Antivirals and the Potential Benefits of Orally Inhaled Drug Administration in COVID-19 Treatment. J Pharm Sci. 2022;111(10):2652-2661

12. Chu D, Pan Y, Cheng S, et al. Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia. Clin Chem. 2020;66(4):549-555

13. Grant RA, Morales-Nebreda L, Markov NS, et al. Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia. Nature. 2021;590(7847):635-641

14. Zu ZY, Jiang MD, Xu PP, et al. Coronavirus Disease 2019 (COVID-19): A Perspective from China. Radiology. 2020;296(2):E15-E25

15. Shuai H, Chan JF, Hu B, et al. Attenuated replication and pathogenicity of SARS-CoV-2 B.1.1.529 Omicron. Nature. 2022;603(7902):693-699

16. Lee JE, Hwang M, Kim YH, et al. SARS-CoV-2 Variants Infection in Relationship to Imaging-based Pneumonia and Clinical Outcomes. Radiology. 2023;306(3):e221795

17. Bai HX, Hsieh B, Xiong Z, et al. Performance of Radiologists in Differentiating COVID-19 from Non-COVID-19 Viral Pneumonia at Chest CT. Radiology. 2020;296(2):E46-E54

18. Fields B, Demirjian NL, Dadgar H, Gholamrezanezhad A. Imaging of COVID-19: CT, MRI, and PET. Semin Nucl Med. 2021;51(4):312-320

19. Long C, Xu H, Shen Q, et al. Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT. Eur J Radiol. 2020;126:108961

20. Yang J, Chen X, Deng X, et al. Disease burden and clinical severity of the first pandemic wave of COVID-19 in Wuhan, China. Nat Commun. 2020;11(1):5411

21. Wang X, Wang S, Sun L, Qin G. Prevalence of diabetes mellitus in 2019 novel coronavirus: A meta-analysis. Diabetes Res Clin Pract. 2020;164:108200

22. Hou W, Zhang W, Jin R, Liang L, Xu B, Hu Z. Risk factors for disease progression in hospitalized patients with COVID-19: a retrospective cohort study. Infect Dis (Lond). 2020;52(7):498-505

23. Okuyan HM, Dogan S, Bal T, Çabalak M. Beclin-1, an autophagy-related protein, is associated with the disease severity of COVID-19. Life Sci. 2021;278:119596

24. Chen W, Zheng KI, Liu S, Yan Z, Xu C, Qiao Z. Plasma CRP level is positively associated with the severity of COVID-19. Ann Clin Microbiol Antimicrob. 2020;19(1):18

Posted on 14 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167879588.86752560/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

25. Hachim MY, Hachim IY, Naeem KB, Hannawi H, Salmi IA, Hannawi S. D-dimer, Troponin, and Urea Level at Presentation With COVID-19 can Predict ICU Admission: A Single Centered Study. Front Med (Lausanne). 2020;7:585003

26. Zingaropoli MA, Nijhawan P, Carraro A, et al. Increased sCD163 and sCD14 Plasmatic Levels and Depletion of Peripheral Blood Pro-Inflammatory Monocytes, Myeloid and Plasmacytoid Dendritic Cells in Patients With Severe COVID-19 Pneumonia. Front Immunol. 2021;12:627548

27. Jakubzick CV, Randolph GJ, Henson PM. Monocyte differentiation and antigen-presenting functions. Nat Rev Immunol. 2017;17(6):349-362

28. Channappanavar R, Perlman S. Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology. Semin Immunopathol. 2017;39(5):529-539

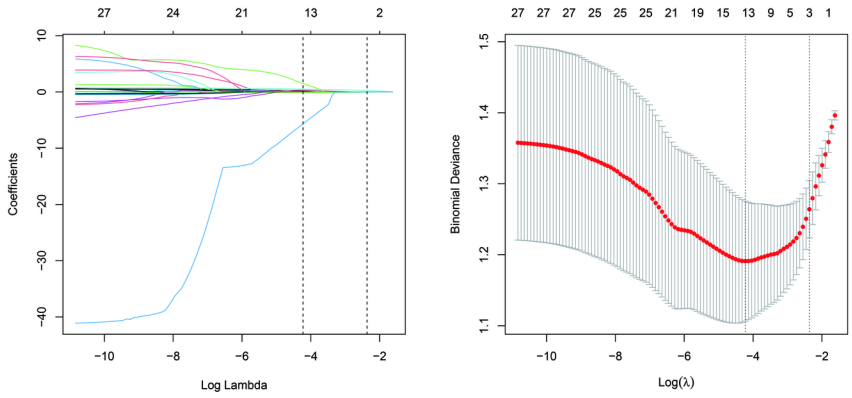
29. Chang MC, Park YK, Kim BO, Park D. Risk factors for disease progression in COVID-19 patients. BMC Infect Dis. 2020;20(1):445

30. Terada M, Ohtsu H, Saito S, et al. Risk factors for severity on admission and the disease progression during hospitalisation in a large cohort of patients with COVID-19 in Japan. BMJ Open. 2021;11(6):e047007

31. Ninomiya T, Otsubo K, Hoshino T, et al. Risk factors for disease progression in Japanese patients with COVID-19 with no or mild symptoms on admission. BMC Infect Dis. 2021;21(1):850

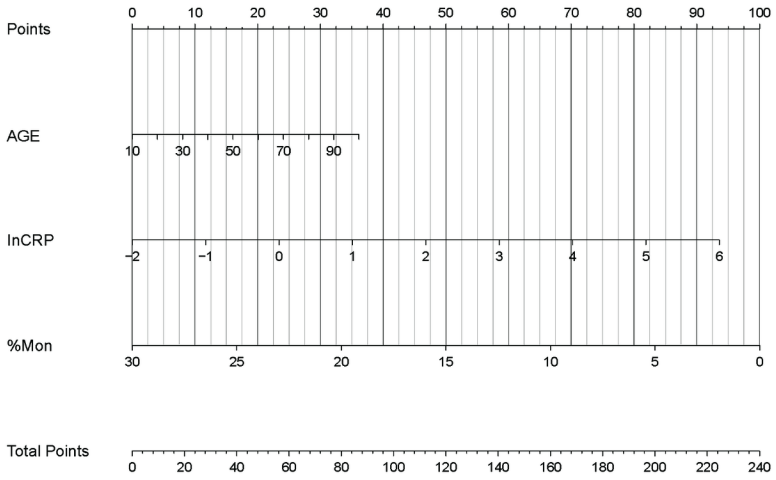
Hosted file

`Fig. 1_.docx` available at https://authorea.com/users/595835/articles/629627-development-and-validation-of-a-model-for-the-prediction-of-the-risk-of-pneumonia-in-patients-with-sars-cov-2-infection

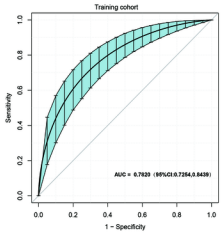


A B

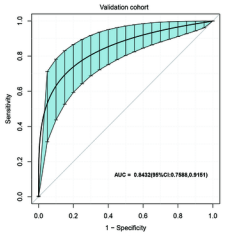
Posted on 14 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167879588.86752560/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.



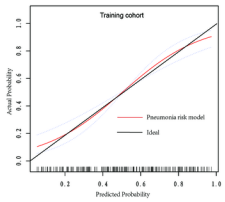
A



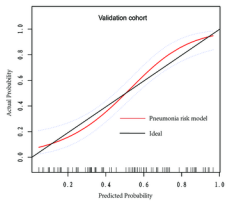
B



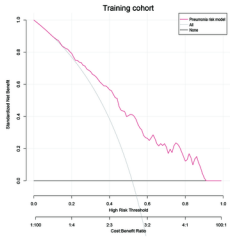
C



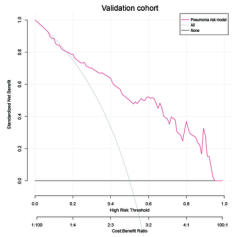
D



E

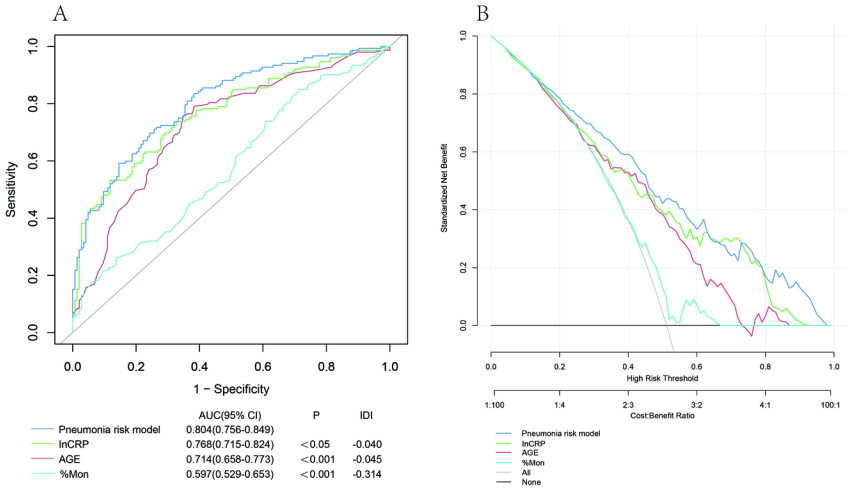


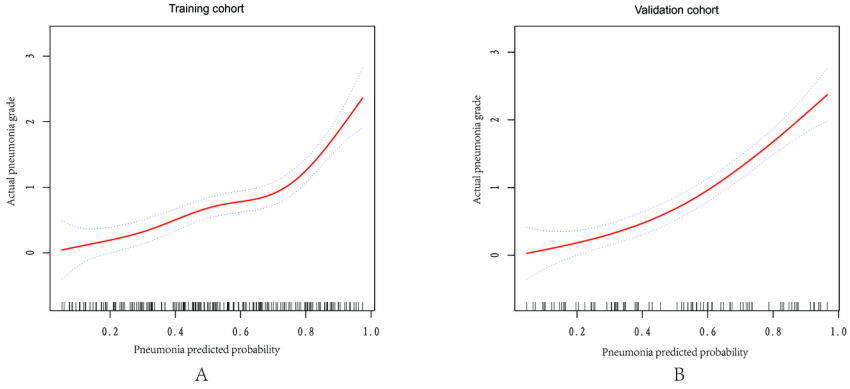
F



G

Posted on 14 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167879588.86752560/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.





Posted on 14 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167879588.86752560/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.