Lung field-based severity score (LFSS): a feasible tool to identify COVID-19 patients at high risk of progressing to critical disease

Xin'ang Jiang[1#], Jun Hu[2#], Qinling Jiang[1#], Taohu Zhou[1,3], Fei Yao[1,4], Yi Sun[2], Qingyang Liu[1], Chao Zhou[2], Kang Shi[2], Xiaoqing Lin[1], Jie Li[1], Yueze Li[1], Qianxi Jin[1], Wenting Tu[1], Xiuxiu Zhou[1], Yun Wang[1], Xiaoyan Xin[2], Shiyuan Liu[1], Li Fan[1^]

[1]Department of Radiology, Second Affiliated Hospital of Naval Medical University, Shanghai, China; [2]Department of Radiology, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing, China; [3]School of Medical Imaging, Weifang Medical University, Weifang, China; [4]School of Medicine, Shanghai University, Shanghai, China

Contributions: (I) Conception and design: X Jiang, J Hu, L Fan, S Liu; (II) Administrative support: L Fan, S Liu, X Xin; (III) Provision of study materials or patients: Q Jiang, Q Liu; (IV) Collection and assembly of data: C Zhou, K Shi, X Lin, J Li, Y Li, Q Jin; (V) Data analysis and interpretation: T Zhou, F Yao, W Tu, X Zhou, Y Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work as co-first authors.

Correspondence to: Li Fan, MD. Department of Radiology, Second Affiliated Hospital of Naval Medical University, No. 415 Fengyang Road, Shanghai 200003, China. Email: fanli0930@163.com.

Background: Coronavirus disease 2019 (COVID-19) still poses a threat to people's physical and mental health. We proposed a new semi-quantitative visual classification method for COVID-19, and this study aimed to evaluate the clinical usefulness and feasibility of lung field-based severity score (LFSS).

Methods: This retrospective study included 794 COVID-19 patients from two hospitals in China between December 2022 and January 2023. Six lung fields on the axial computed tomography (CT) were defined. LFSS and eighteen clinical characteristics were evaluated. LFSS was based on summing up the parenchymal opacification involving each lung field, which was scored as 0 (0%), 1 (1–24%), 2 (25–49%), 3 (50–74%), or 4 (75–100%), respectively (range of LFSS from 0 to 24). Total pneumonia burden (TPB) was calculated using the U-net model. The correlation between LFSS and TPB was analyzed. After performing logistic regression analysis, an LFSS-based model, clinical-based model and combined model were developed. Receiver operating characteristic curves were used to evaluate and compare the performance of three models.

Results: LFSS, age, chronic liver disease, chronic kidney disease, white blood cell, neutrophils, lymphocytes and C-reactive protein differed significantly between the non-critical and critical group (all P<0.05). There was a strong positive correlation of LFSS and TPB (Pearson correlation coefficient =0.767, P<0.001). The area under curves of LFSS-based model, clinical-based model and combined model were 0.799 [95% confidence interval (CI): 0.770–0.827], 0.758 (95% CI: 0.727–0.788), and 0.848 (95% CI: 0.821–0.872), respectively.

Conclusions: The LFSS derived from chest CT may be a potential new tool to help identify COVID-19 patients at high risk of progressing to critical disease.

Keywords: Coronavirus disease 2019 (COVID-19); computed tomography (CT); prediction

---

^ ORCID: 0000-0003-4722-3933.

Introduction

In May 2023, the World Health Organization declared that coronavirus disease 2019 (COVID-19) pandemic no longer constitutes a public health emergency of international concern (PHEIC). Despite this declaration, the pandemic is not over yet and COVID-19 still poses a threat to people's physical and mental health, with a wide range of clinical manifestations ranging from asymptomatic or mild respiratory symptoms to severe pneumonia and acute respiratory distress syndrome (ARDS) (1,2). Chest computed tomography (CT) is widely used for the diagnosis of COVID-19 and to predict the prognosis of the disease because of its intuitive presentation and fast scanning time (3-6).

The severity of COVID-19 should be stratified to prioritize medical resources in hospitals, particularly when resources and medical staff are limited (7). Several scoring systems using chest CT have been proposed to quantify lung involvement in COVID-19 and further estimate the diagnosis of this high-risk disease (8-14), such as the chest CT score (CCTS) (10), the CT severity score (CTSS) (11) and so on. Among all the severity scoring systems, CCTS has attracted great attention of scholars (15,16). In CCTS, each lobe is assigned a CT score from 0 to 5, depending on the percentage of the involved lobe: score 0, 0% involvement; score 1, <5% involvement; score 2, 5–25% involvement; score 3, 26–49% involvement; score 4, 50–75% involvement; and score 5, >75% involvement. Although the application of these

scores has made significant progress in assessing the severity of COVID-19, they have some limitations. First, they require a lot of time due to their complexity. According to Inoue *et al.* (15), the mean interpretation times for CTSS, total CT score (TSS), and CCTS were 48.9–80.0 s, 25.7–41.7 s, and 27.7–39.5 s, respectively. They are time-consuming. Some scores range from 20 to 40 regions, which increases the difficulty of evaluation. Second, COVID-19 patients who have undergone partial lung resection surgery cannot be evaluated using these scoring systems due to the lack of a certain lung lobe or segment, and finally resulting in inaccurate scoring. Third, the sizes of both lungs are different, with the right larger than the left, including their corresponding lobes and segments themselves. Even quantitative methods for accurately calculating lung lesion volume require specialized software (17). An effective and practical scoring system must be capable of quantifying lung changes while being simple to use. In stressful medical environments, experienced and inexperienced radiologists and physicians have equal chance of using the scoring system, an effective and practical scoring system should have simple grading characteristics and criteria that produce high repeatability, strong interobserver consistency, and high diagnostic accuracy.

To alleviate this dilemma, we proposed a new semi-quantitative visual classification method, the lung field-based severity score (LFSS). On chest CT images, the entire lung field was divided into three areas: the upper (level of the aortic arch), middle (level of the carina) and lower (level of the upper end of the diaphragm), with a total of six areas for both sides. LFSS was based on summing up the parenchymal opacification involving each lung field, which was scored as 0 (0%), 1 (1–24%), 2 (25–49%), 3 (50–74%), or 4 (75–100%), respectively.

This study aimed to evaluate the clinical usefulness and feasibility of LFSS in predicting individualized prognosis of COVID-19 patients. We present this article in accordance with the TRIPOD reporting checklist (available at https://jtd.amegroups.com/article/view/10.21037/jtd-24-544/rc).

Highlight box

Key findings
• Compared with chest computed tomography (CT) score, lung field-based severity score (LFSS) improved reading efficiency by 28%.
• LFSS-based model achieved an area under the curve of 0.799 to identify high-risk coronavirus disease 2019 (COVID-19) patients.

What is known and what is new?
• The existing scoring system has limitations, for example, it is not applicable to COVID-19 patients who have undergone partial lung resection surgery.
• LFSS is based on lung field, has high repeatability, strong interobserver consistency, and high diagnostic accuracy.

What is the implication, and what should change now?
• LFSS can alleviate related clinical dilemma, by leveraging LFSS, clinicians can predict COVID-19 progression more accurately, further optimize resource allocation and improve outcomes for patients affected by COVID-19.

Methods

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the institutional review board of the Second Affiliated Hospital of Naval Medical University and Nanjing Drum Tower Hospital (ethics approval No.

2020SL006), and individual consent for this retrospective analysis was waived.

Participants

Between December 17, 2022 and January 31, 2023, consecutive patients with confirmed COVID-19 were assessed to include into this study from two hospitals retrospectively: Second Affiliated Hospital of Naval Medical University (Hospital 1) and Nanjing Drum Tower Hospital (Hospital 2). All of the patients had COVID-19 confirmed by positive reverse transcriptase polymerase chain reaction (RT-PCR) tests for severe acute respiratory syndrome coronavirus-2 (SARS-COV-2) from throat-swab specimens (18). The following criteria were applied for the inclusion of study participants: (I) complete thin-slice (1.0–1.5 mm) chest CT images; (II) comprehensive clinical records and laboratory test data. Exclusion criteria included severe trauma or advanced tumors. Patients diagnosed with critical COVID-19 at admission were also excluded because this study was to predict the progression of COVID-19 from the non-critical to critical disease. According to the "Diagnosis and treatment protocol for COVID-19 patients (Trial Version 9)" recommended by China's National Health Commission, Participants meeting any of the following criteria were regarded as the critical group: (I) respiratory failure occurred and mechanical ventilation required; (II) shock; (III) other organ failure needing intensive care unit (ICU) monitoring treatment (19). Furthermore, patients with a fatal outcome were also included in the critical group. Finally, patients were classified into two cohorts: the critical group and the non-critical group. The follow-up time was set to be 3 weeks after the patient were diagnosed with COVID-19, the patients who were still in the hospital use the electronic medical record system to record information, and the patients who have been discharged were followed up by telephone.

Finally, a total of 794 patients (351 from Hospital 1 and 443 from Hospital 2) were enrolled in our study.

CT examinations and clinical data collection

All non-enhanced chest CT images were acquired using multi-slice CT systems from four different manufacturers, including United-imaging, Philips, SIEMENS and GE (detailed scan and reconstruction parameters are shown in Table S1).

The clinical characteristics including demographics, 6

underlying diseases, clinical symptoms and routine blood tests were extracted from electronic medical records (Table 1). In each case, the first CT scan and laboratory test data after being diagnosed as COVID-19 were collected.

Image semi-quantitative analysis

Two independent radiologists (with 7 and 10 years of experience, respectively) blinded to clinical data reviewed CT images of all the patients according to the LFSS: On chest CT images, the whole lung was divided into three fields: the upper (level of the aortic arch), middle (level of the carina) and lower (level of the upper end of the diaphragm), with a total of six fields for both lungs. LFSS was defined by summing up individual scores from each lung field, scores of 0, 1, 2, 3 and 4 were respectively assigned for each lung field if parenchymal opacification involved 0%, <25%, ≥25% and <50%, ≥50% and <75%, and ≥75% of each region. CCTS, which was mentioned in the introduction, has been proven to be a relatively appropriate CT scoring system for clinical practice in some COVID-19 related studies (15,16,20). To verify the rationality and accuracy of LFSS, we randomly selected 20% of patients and scored them using CCTS and LFSS, respectively. The corresponding reading time was recorded.

Deep learning algorithm COVID-19 detection

The deep learning algorithm analysis was performed using artificial intelligence (AI) software (InferRead CT Pneumonia, V1.1.3.0, Infersion, Beijing, China), an AI solution specifically developed for diagnosis and management support of COVID-19 pneumonia (21). The model was first created by training the first batch of patients infected with COVID-19 in China. It was then refined by training a bigger population. In particular, the patient features (n=2,191 adult patients) are mixed for the trained AI model, encompassing all disease stages and clinical presentations (symptoms could be mild, moderate, or severe). The core algorithm is based on a novel deep convolutional neural network structure and uses the U-net network structure as the core segmentation network (21,22). The algorithm module includes automated segmentation of the core features of COVID-19 lung lesions and the segmentation of the lung lobes, followed by calculation of total lesion volume (sum volume of GGO, consolidation and nodular opacities) and corresponding total lung volume. Total pneumonia burden (TPB) was calculated using the

Table 1 Characteristics of patients for clinical model construction

Variables	Total (n=794)	Non-critical (n=686)	Critical (n=108)	P
Age (years)	69.44±14.74	68.76±14.65	73.77±14.64	0.001
Gender				0.29
Male	508 (63.98)	434 (63.27)	74 (68.52)	
Female	286 (36.02)	252 (36.73)	34 (31.48)	
Smoke				0.29
No	707 (89.04)	614 (89.50)	93 (86.11)	
Yes	87 (10.96)	72 (10.50)	15 (13.89)	
Hypertension				0.54
No	352 (44.33)	307 (44.75)	45 (41.67)	
Yes	442 (55.67)	379 (55.25)	63 (58.33)	
Coronary heart disease				0.02
No	643 (80.98)	564 (82.22)	79 (73.15)	
Yes	151 (19.02)	122 (17.78)	29 (26.85)	
Chronic lung disease				0.03
No	715 (90.05)	624 (90.96)	91 (84.26)	
Yes	79 (9.95)	62 (9.04)	17 (15.74)	
Diabetes				0.03
No	548 (69.02)	483 (70.41)	65 (60.19)	
Yes	246 (30.98)	203 (29.59)	43 (39.81)	
Chronic liver disease				0.10
No	742 (93.45)	645 (94.02)	97 (89.81)	
Yes	52 (6.55)	41 (5.98)	11 (10.19)	
Chronic kidney disease				<0.001
No	655 (82.49)	582 (84.84)	73 (67.59)	
Yes	139 (17.51)	104 (15.16)	35 (32.41)	
Cough				0.39
No	251 (31.61)	213 (31.05)	38 (35.19)	
Yes	543 (68.39)	473 (68.95)	70 (64.81)	
Fever				0.59
No	170 (21.41)	149 (21.72)	21 (19.44)	
Yes	624 (78.59)	537 (78.28)	87 (80.56)	
Sore throat				0.60
No	709 (89.29)	611 (89.07)	98 (90.74)	
Yes	85 (10.71)	75 (10.93)	10 (9.26)	
Muscle soreness				0.86
No	717 (90.30)	619 (90.23)	98 (90.74)	
Yes	77 (9.70)	67 (9.77)	10 (9.26)	
WBC (×10^9/L)	7.31±13.31	6.75±12.85	10.87±15.53	0.01
Neutrophils (×10^9/L)	5.01±3.78	4.61±3.31	7.49±5.36	<0.001
Lymphocytes (×10^9/L)	1.61±12.12	1.56±12.24	1.93±11.42	0.77
CRP (mg/L)	52.53±54.14	48.24±49.97	79.82±69.76	<0.001

Data are presented as mean ± standard deviation or number (percentage). WBC, white blood cell; CRP, C-reactive protein.

following formula: total lesion volume/total lung volume × 100%. Figure S1 shows screenshots of the AI viewer after the assessment of a patient with a confirmed diagnosis of COVID-19.

Statistical analysis

R software (version 4.3.1) and MedCalc Software (version 20.022) were used for statistical analysis.

Bland-Altman analysis was not only used to evaluate interobserver agreement of LFSS or CCTS, but also to analyze the consistency of results between LFSS and CCTS. Spearman rank correlation analysis was used to assess associations between LFSS and TPB.

Mean ± standard deviation (SD) and proportions were used to express continuous and categorical variables, respectively. The independent-sample t-test was used to assess normally distributed data. The Mann-Whitney U test was used to assess nonnormally distributed data. For categorical variables, the chi-squared and Fisher exact tests were performed. P value <0.05 was considered statistically significant. Receiver operating characteristic (ROC) curve was used to analyze and evaluate the prediction performance of the model. To determine whether the efficiency disparity between the models was statistically significant, the DeLong test was applied.

Results

In total, 794 patients (286 females and 508 males; age 69.4±14.7 years) were enrolled, of which 689 (86.4%) were non-critical cases (age 68.7±14.6 years) and 108 (13.6%) were critical cases (age 73.7±14.6 years). The description of the demographic and clinical features of the study population is summarized in Table 1. Age was found significantly different between the two groups (P=0.001). The critical group was more likely to have coronary heart disease, chronic lung disease, diabetes, and chronic kidney disease (CKD) than the non-critical group (P all <0.05). Compared with the non-critical group, patients in the critical group had worse laboratory test results, including white blood cell (WBC) [(10.87±15.53)×10⁹/L], neutrophils [(7.49±5.36)×10⁹/L], and C-reactive protein (CRP) (79.82±69.76 mg/L) (all P<0.05).

Reproducibility of lung field-based severity scoring system

A Blant-Atman analysis compared the LFSS or CCTS scores between two readers for a random 20% of all patients.

The mean difference in LFSS between the two readers was –0.1, with a consistency range of –2.3 to 2.1, where most data points fell within this range (Figure 1A). When in CCTS, the mean difference was +0.2, with a consistency range of –2.2 to 2.5 (Figure 1B). Figure 1C indicates consistency between the two scoring systems, with biases not significantly different and limits of agreement within 5% across the two systems.

Validation of lung field-based severity scoring system

Table 2 presents the LFSS scores for non-critical and critical groups. The scores in critical group were significantly higher than those in non-critical groups (P<0.001) not only at whole lung level but also at lung field level.

Compared to CCTS, LFSS exhibited similar diagnostic efficacy [area under curve (AUC), 0.769 vs. 0.776] (Table 3). The DeLong test indicated no significant differences between the AUCs of these two scoring systems (P=0.88, DeLong test). In identifying the risk of COVID-19 patients progressing to critical illness, LFSS demonstrated higher sensitivity, true positive rate, and accuracy. Notably, LFSS required less time compared to CCTS (21.78±6.19 vs. 30.33±5.88 s).

Deep learning COVID-19 detection result (Table 2) revealed that TPB, RPB, and LPB for all patients were 11.18%±13.76%, 12.25%±15.22%, and 10.11%±13.37%, respectively. At the whole lung level, right lung level, and left lung level, the pneumonia burden in critical group was all significantly higher than non-critical group (P<0.001). LFSS were strongly positive correlated with the pneumonia burden (Figure 2). Spearman's correlation coefficients at the whole lung level, left lung level, and right lung level were 0.767, 0.727 and 0.738, respectively.

Model construction and comparison

Based on the multiple logistic regression (Table 4), the clinical characteristics of age, neutrophils, CRP, CKD, and coronary heart disease were included in the clinical-based model. LFSS-based model was consisted of the total score of semi-quantitative visual assessment only.

The combined model was developed with total LFSS score, age, neutrophils, CRP, CKD, and coronary heart disease. Multivariate logistic regression analyses showed that total LFSS score, neutrophils, and CKD were independent predictors in combined model (P<0.05). The corresponding adjusted odds ratios (ORs) were 1.28 [95% confidence interval (CI): 1.21–1.35; P<0.001], 1.12 (95%
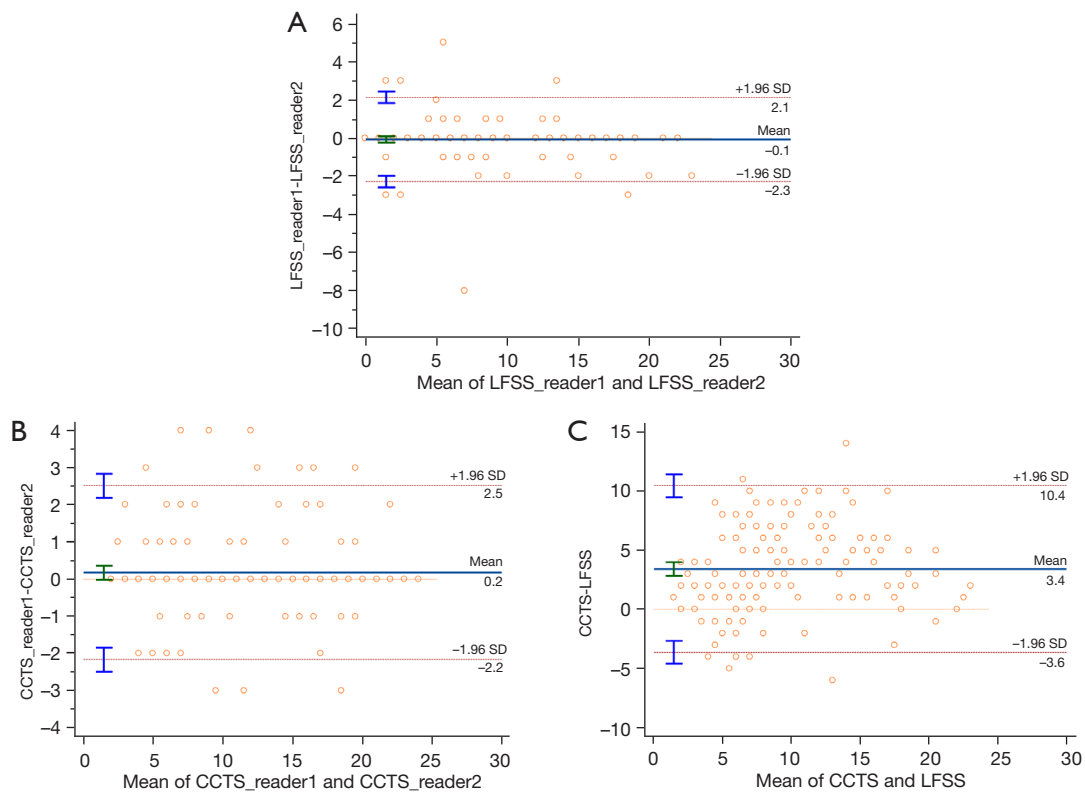
Figure 1 Bland-Altman plots for consistency assessment. SD, standard deviation; CCTS, chest computed tomography score; LFSS, lung field-based severity score.

Table 2 Results of lung field- based severity scoring system and deep learning COVID-19 detection

Parameters	Total (n=794)	Non-critical (n=686)	Critical (n=108)	P
Total score	8.02±4.53	7.23±3.79	13.06±5.55	<0.001
Right score	4.06±2.44	3.71±2.15	6.28±2.97	<0.001
Right upper	1.14±0.91	1.01±0.80	1.94±1.13	<0.001
Right middle	1.27±0.90	1.15±0.82	2.01±1.07	<0.001
Right lower	1.65±0.96	1.55±0.89	2.32±1.07	<0.001
Left score	3.96±2.37	3.58±2.04	6.39±2.88	<0.001
Left upper	1.01±0.83	0.89±0.73	1.79±1.01	<0.001
Left middle	1.25±0.87	1.13±0.75	2.05±1.12	<0.001
Left lower	1.70±1.01	1.57±0.93	2.56±1.09	<0.001
Pneumonia burden				
TPB (%)	11.18±13.76	8.66±10.56	27.16±19.79	<0.001
RPB (%)	12.25±15.22	9.63±12.17	28.90±21.09	<0.001
LPB (%)	10.11±13.37	7.72±10.15	25.29±19.91	<0.001

Data are presented as mean ± standard deviation. Pneumonia burden = lesion volume/lung volume × 100%. COVID-19, coronavirus disease 2019; TPB, total pneumonia burden; RPB, right pneumonia burden; LPB, left pneumonia burden.

Table 3 Comparison of LFSS and CCTS among 158 randomly selected patients

Parameters	LFSS	CCTS
Evaluated objects	3 lung fields in the right lung; 3 lung fields in the left lung	3 lobes in the right lung; 2 lobes in the left lung
Score for each region	0, 0%; 1, 1–24%; 2, 25–49%; 3, 50–74%; 4, 75–100%	0, 0%; 1, <5%; 2, 5–25%; 3, 26–49%; 4, 50–75%; 5, >75%
Range of total score	0–24	0–25
Total score, mean ± SD	7.75±4.89	11.16±5.48
Reading time (s), mean ± SD	21.78±6.19	30.33±5.88
AUC (95% CI)	0.769 (0.696–0.832)	0.776 (0.703–0.838)
Best cut-off value	12.5	8.5
Sensitivity (%)	79.17	66.67
Specificity (%)	71.64	73.13
True positive rate (%)	79.17	66.67
False positive rate (%)	28.36	26.87
True negative rate (%)	71.64	73.13
False negative rate (%)	20.83	33.33
Accuracy (%)	72.78	72.15

Number of two groups among 158 patients: non-critical (n=134), critical (n=24). LFSS, lung field-based severity score; CCTS, chest CT score; CT, computed tomography; SD, standard deviation; AUC, area under the curve; CI, confidence interval.
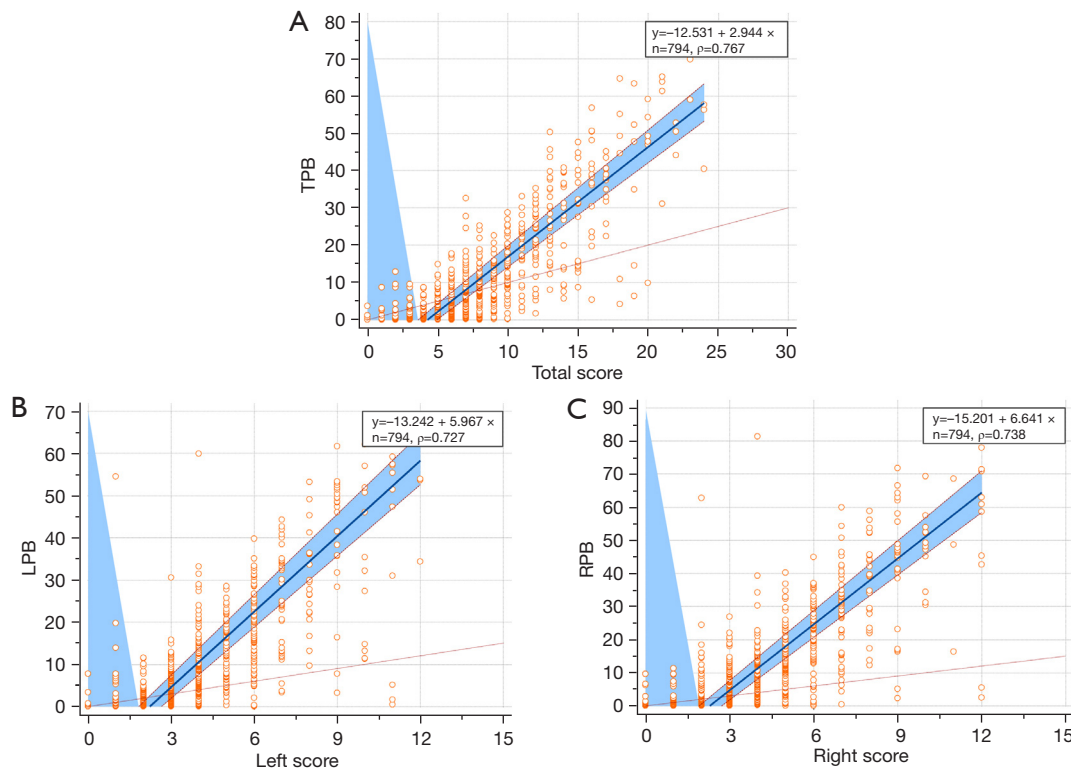
Figure 2 Scatter plot and regression line between pneumonia burden and the corresponding lung field-based severity score. (A) Correlation between TPB and total score. (B) Correlation between LPB and left score. (C) Correlation between RPB and right score. TPB, total pneumonia burden; LPB, left pneumonia burden; RPB, right pneumonia burden.

Table 4 The results of multivariate logistic regression analysis

Model	Variables	β	Adjusted OR (95% CI)	P
LFSS-based model	Total score	0.26	1.30 (1.23–1.36)	<0.001
Clinical-based model	Age	0.02	1.02 (1.01–1.04)	0.03
	Neutrophils	0.14	1.15 (1.09–1.21)	<0.001
	CRP	0.01	1.01 (1.01–1.01)	<0.001
	Chronic kidney disease	1.00	2.73 (1.67–4.44)	<0.001
	Coronary heart disease	0.39	1.48 (0.88–2.48)	0.13
Combined model	Total score	0.25	1.28 (1.21–1.35)	<0.001
	Age	0.02	1.02 (1.00–1.03)	0.08
	Neutrophils	0.11	1.12 (1.06–1.18)	<0.001
	CRP	0.01	1.00 (1.00–1.01)	0.10
	Chronic kidney disease	1.17	3.22 (1.86–5.59)	<0.001
	Coronary heart disease	0.45	1.56 (0.89–2.73)	0.11

OR, odds ratio; LFSS, lung field-based severity score; CRP, C-reactive protein; CI, confidence interval.

Table 5 Predictive performance of three different models

Model types	AUC (95% CI)	Accuracy (%)	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)
LFSS-based model	0.799 (0.770–0.827)	80.23	63.89	82.80	93.57	36.90
Clinical-based model	0.758 (0.727–0.788)	70.40	69.44	70.55	93.62	27.08
Combined model	0.848 (0.821–0.872)	83.88	71.30	85.86	95.00	44.25

LFSS, lung field-based severity score; AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value; CI, confidence interval.
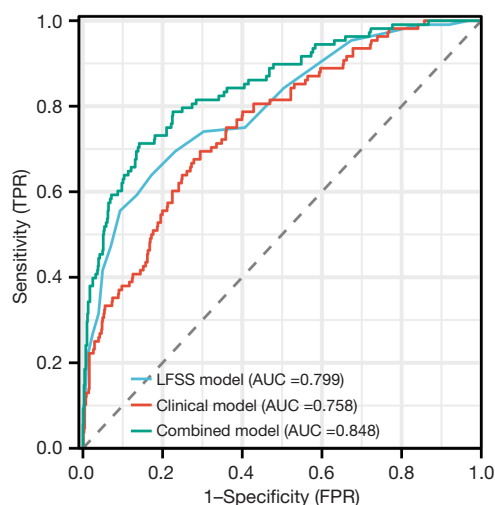

Figure 3 Receiver operating characteristic curves for three models. TPR, true positive rate; FPR, false positive rate; AUC, area under the curve; LFSS, lung field-based severity score.

CI: 1.06–1.18; P<0.001), and 3.22 (95% CI: 1.86–5.59; P<0.001), respectively.

The LFSS-based model was established by univariate logistic regression with the total score as the only independent variable. The corresponding adjusted OR was 1.30 (95% CI: 1.23–1.36; P<0.001).

According to ROC curve analysis of three models (*Table 5* and *Figure 3*), at the optimal threshold, the LFSS-based model showed good performance for identifying COVID-19 patients at high risk of progressing to critical disease (AUC, 0.799; 95% CI: 0.770–0.827; sensitivity, 63.89%; specificity, 82.80%; accuracy, 80.23%). Clinical-based model performed slightly worse than LFSS-based model, with AUC, sensitivity, specificity, and accuracy values of 0.758 (95% CI: 0.727–0.788), 69.44%, 70.55%, and 70.40%, respectively. When total LFSS score was combined with clinical features, the diagnostic performance improved (AUC, 0.848; 95% CI: 0.821–0.872; sensitivity,
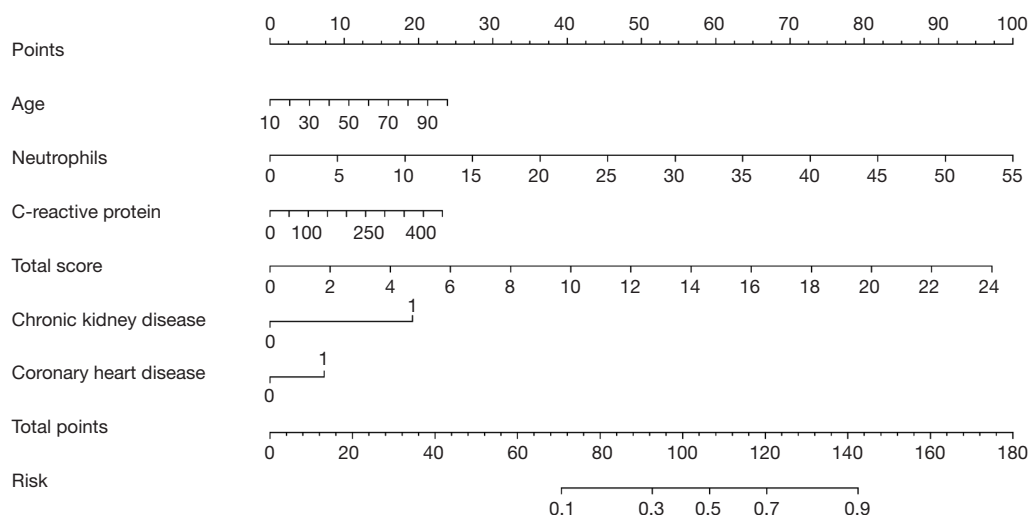
Figure 4 Corresponding nomogram for the combined model. The nomogram is constructed by combining age, neutrophils, C-reactive protein, total score, chronic kidney disease and coronary heart disease. On the point scale axis, each variable was assigned a score. The overall score was calculated by adding each score. We were able to determine the probability of critical disease using the whole-point scale.

71.30%; specificity, 85.86%; accuracy, 83.88%). To make the results more intuitive, the combined model was displayed as combined nomogram, as shown in *Figure 4*. Two examples of applying dynamic nomogram are shown in *Figure 5*. The DeLong test revealed that the combined model had enhanced predictive performance than LFSS-based model or clinical-based model (P≤0.001, DeLong test). However, there was no significant difference between clinical-based model and LFSS-based model (P=0.18, DeLong test) (Table S2).

Discussion

The COVID-19 pandemic is not over yet and poses a continuing and substantive threat to people's physical and mental health, especially the elderly individuals (23,24). Early identification of COVID-19 patients at high risk of progressing to critical disease is crucial for prompt clinical intervention and resource allocation, thereby improving overall patient outcomes and healthcare efficiency. Several different semi-quantitative scoring systems have been proposed to assess the degree of acute COVID-19 lung involvement (8-14). Given the current complexities, time constraints, and limited applicability of existing scoring systems in clinical practice, we propose a novel visual scoring method based on lung fields rather than lung lobes or segments. In this study, we validated the effectiveness and reliability of LFSS in quantifying the severity of COVID-19

lung involvement by analyzing its correlation with the TPB detected through a deep learning model. Furthermore, we assessed the feasibility of LFSS in predicting individualized prognosis among COVID-19 patients. The result showed that the LFSS-based model had good performance for identifying high-risk patients, when LFSS score and clinical features were combined, diagnostic performance was obviously improved.

Among all the severity scoring systems, CCTS has attracted great attention of scholars (15,16). Francone *et al.* (9) reported that CCTS was significantly correlated with CRP (P<0.0001, r=0.6204) and D-dimer (P<0.0001, r=0.6625) levels and got the conclusion that a CCTS score of ≥18 was associated with an increased COVID-19 mortality risk. Inoue *et al.* (15) found CCTS had the shortest interpretation time (27.7–39.5 s) among three different semiquantitative severity scoring systems, CCTS appeared to be the most appropriate CT scoring system for clinical practice. In this study, we randomly selected 20% of patients, scored them using CCTS and LFSS respectively and compared the validity of them. In terms of diagnostic effectiveness, we observed no statistically significant difference between the two systems (AUC, 0.769 *vs.* 0.776). Notably, LFSS required less time compared to CCTS (21.78±6.19 *vs.* 30.33±5.88 s), and reading efficiency improved by approximately 28%. LFSS offers radiologists a lighter interpretive burden and shorter evaluation time as it only requires assessment of involvement on three CT slices

5600

Jiang et al. LFSS to identify high-risk COVID-19 patients
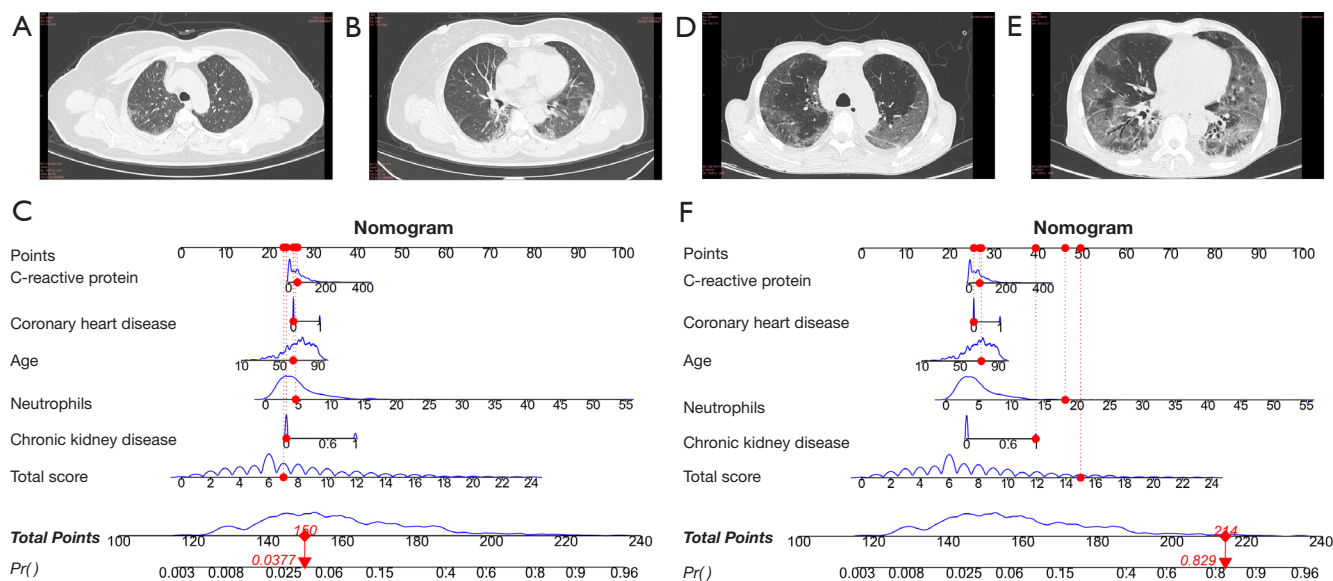
Figure 5 The dynamic nomogram was applied to two examples. (A-C) A 64-year-old non-critical female patient, the dynamic nomogram shows the total points were 150, and the corresponding prediction probability of progressing to critical disease was 0.04. (D-F) A 72-year-old critical male patient, the dynamic nomogram shows the total points were 214, and the corresponding prediction probability of progressing to critical disease was 0.83.

representing three different lung fields.

In addition, we observed that the score of the right lung middle lobe in the CCTS is often abnormally high, which may be due to the small volume and difficulty in defining the range of the right lung middle lobe. We consider that the score of the right lung middle lobe cannot match the severity of pneumonia. In other words, for instance, the effectiveness of 5 points in the right lung middle lobe is not equivalent to 5 points in the right lung upper lobe, thereby interfering with the accuracy of the total CCTS score. Unlike CCTS, LFSS assesses pneumonia involvement at the level of lung fields, effectively alleviating this dilemma. Moreover, LFSS demonstrates applicability in patients who have undergone partial lung resection surgery, a population for which existing scoring systems fail to provide accurate evaluations.

To validate the effectiveness and reliability of LFSS in quantifying the severity of COVID-19 lung involvement, we employed a widely validated deep learning model for automatic detection of COVID-19 pneumonia. The primary advantage of pneumonia burden lies in its ability to provide accurate and reproducible reference values correlated with the severity of pneumonia, thus substantiating the reliability of LFSS rather than relying on human interpretation. In this study, we got great results,

LFSS were strong positive correlated with the pneumonia burden and Spearman's correlation coefficients at the whole lung level, left lung level, and right lung level were 0.767, 0.727 and 0.738, respectively.

The predictive value of the LFSS was further demonstrated by its ability to identify patients at high risk of developing critical disease. We established three logistic regression models to validate the clinical utility of LFSS. Age, gender, smoking status, 6 underlying diseases, 4 disease symptoms, and 4 laboratory parameters were analyzed. Finally, LFSS score, age, neutrophils, CRP, coronary heart disease and chronic kidney disease were used to build different models. Neutrophils, as pivotal players in the innate immune response, have emerged as critical actors in the realm of COVID-19 research (25,26). Increased circulating neutrophil counts and neutrophil migration to the lungs have been documented and associated with severity (27,28). In addition, CRP, as an acute-phase reactive protein, reflects a hyperimmune inflammatory state (29,30). In our study, the LFSS-based model showed better performance than clinical-based model (AUC, 0.799 *vs.* 0.758). When total score was combined with clinical features, the prediction efficiency further improved, the combined model achieved an AUC of 0.848 and an accuracy of 0.839 to identify COVID-19 patients at high risk of

progressing to critical disease. It has been reported that radiomics model could assess the severity of COVID-19 and predict the progress of COVID-19 (31-33). Zhang *et al.* (34) predicted the severity of patients with COVID-19 using CT radiomics features, achieving an AUC of 0.83 in training set. Zysman *et al.* (32) developed a prediction model for the transition from mild to moderate or severe form of COVID-19 based on CT radiomics, achieving an AUC of 0.76 in training set. Our LFSS model showed superior or similar predictive performance than the reported radiomics model. Moreover, our method has better universality, especially for grassroots units that are not equipped with AI, which is convenient, easy to implement and has strong generalisability. All these results indicated that using LFSS to predict individualized prognosis of COVID-19 patients was feasible.

Although LFSS shows promise as a predictive tool, some limitations should be acknowledged. Firstly, the retrospective nature of this study introduces inherent biases and limits the generalizability of our findings. Prospective studies are needed to confirm this observation and longitudinal follow-up of patients should be conducted to assess the long-term predictive value of LFSS. Second, the most common lung abnormalities in COVID-19 included lung consolidations, ground glass opacities and reticular opacities (35,36). Our semi-quantitative visual classification method does not stratify these lung opacities, this may affect the accuracy of diagnosis. Third, when patients are admitted, they are at different stages of COVID-19, but we only included the first laboratory examinations and CT scans after admission. Further research will stratify different time points post-infection. Fourth, our study includes patients from two hospitals, which may not provide a sufficiently diverse sample to generalise the findings. In addition, the exclusion criteria include patients with severe trauma or advanced tumours, which may limit the understanding of the utility of the LFSS in a broader spectrum of COVID-19 patients. More diverse demographics are needed to improve the applicability of the LFSS.

Conclusions

In conclusion, this study proposed a new semi-quantitative visual classification method for identifying COVID-19 patients at high risk of progressing to critical disease. We recommend the use of LFSS in assessing COVID-19 lung involvement since it is efficient and has a wider range of

applications. By leveraging LFSS, clinicians can predict COVID-19 progression more accurately, further optimize resource allocation and improve outcomes for patients affected by COVID-19.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://jtd.amegroups.com/article/view/10.21037/jtd-24-544/rc

Data Sharing Statement: Available at https://jtd.amegroups.com/article/view/10.21037/jtd-24-544/dss

Peer Review File: Available at https://jtd.amegroups.com/article/view/10.21037/jtd-24-544/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://jtd.amegroups.com/article/view/10.21037/jtd-24-544/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the institutional review board of the Second Affiliated Hospital of Naval Medical University and Nanjing Drum Tower Hospital (ewthics approval No. 2020SL006), and individual consent for this retrospective analysis was waived.

5602

Jiang et al. LFSS to identify high-risk COVID-19 patients

References

1. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020;395:507-13.
2. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020;323:1239-42.
3. Zhang N, Xu X, Zhou LY, et al. Clinical characteristics and chest CT imaging features of critically ill COVID-19 patients. Eur Radiol 2020;30:6151-60.
4. Chung M, Bernheim A, Mei X, et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). Radiology 2020;295:202-7.
5. Song F, Shi N, Shan F, et al. Emerging 2019 Novel Coronavirus (2019-nCoV) Pneumonia. Radiology 2020;295:210-7.
6. Zhou Z, Guo D, Li C, et al. Coronavirus disease 2019: initial chest CT findings. Eur Radiol 2020;30:4398-406.
7. Middleton J, Lopes H, Michelson K, et al. Planning for a second wave pandemic of COVID-19 and planning for winter : A statement from the Association of Schools of Public Health in the European Region. Int J Public Health 2020;65:1525-7.
8. Li K, Fang Y, Li W, et al. CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). Eur Radiol 2020;30:4407-16.
9. Francone M, Iafrate F, Masci GM, et al. Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis. Eur Radiol 2020;30:6808-17.
10. Li K, Wu J, Wu F, et al. The Clinical and Chest CT Features Associated With Severe and Critical COVID-19 Pneumonia. Invest Radiol 2020;55:327-31.
11. Yang R, Li X, Liu H, et al. Chest CT Severity Score: An Imaging Tool for Assessing Severe COVID-19. Radiol Cardiothorac Imaging 2020;2:e200047.
12. Pan F, Ye T, Sun P, et al. Time Course of Lung Changes at Chest CT during Recovery from Coronavirus Disease 2019 (COVID-19). Radiology 2020;295:715-21.
13. Salaffi F, Carotti M, Tardella M, et al. The role of a chest computed tomography severity score in coronavirus disease 2019 pneumonia. Medicine (Baltimore) 2020;99:e22433.
14. Xie X, Zhong Z, Zhao W, et al. Chest CT for Typical Coronavirus Disease 2019 (COVID-19) Pneumonia: Relationship to Negative RT-PCR Testing. Radiology 2020;296:E41-5.
15. Inoue A, Takahashi H, Ibe T, et al. Comparison of semiquantitative chest CT scoring systems to estimate severity in coronavirus disease 2019 (COVID-19) pneumonia. Eur Radiol 2022;32:3513-24.
16. Almasi Nokiani A, Shahnazari R, Abbasi MA et al. CT severity score in COVID-19 patients, assessment of performance in triage and outcome prediction: a comparative study of different methods. Egypt J Radiol Nucl Med 2022;53:116.
17. Yin X, Min X, Nan Y, et al. Assessment of the Severity of Coronavirus Disease: Quantitative Computed Tomography Parameters versus Semiquantitative Visual Score. Korean J Radiol 2020;21:998-1006.
18. World Health Organization. Laboratory testing of 2019 novel coronavirus (2019-nCoV) in suspected human cases: interim guidance. 17 January 2020.
19. National Health Commission of the People's Republic of China. Diagnosis and Treatment Protocol for COVID-19 (Trial Version 9). March 15 2022.
20. Brumini I, Dodig D, Žuža I, et al. Validation of Diagnostic Accuracy and Disease Severity Correlation of Chest Computed Tomography Severity Scores in Patients with COVID-19 Pneumonia. Diagnostics (Basel) 2024;14:148.
21. Wang M, Xia C, Huang L, et al. Deep learning-based triage and analysis of lesion burden for COVID-19: a retrospective study with external validation. Lancet Digit Health 2020;2:e506-15.
22. Mallio CA, Napolitano A, Castiello G, et al. Deep Learning Algorithm Trained with COVID-19 Pneumonia Also Identifies Immune Checkpoint Inhibitor Therapy-Related Pneumonitis. Cancers (Basel) 2021;13:652.
23. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020;584:430-6.
24. Clift AK, Coupland CAC, Keogh RH, et al. Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. BMJ 2020;371:m3731.
25. Terpos E, Ntanasis-Stathopoulos I, Elalamy I, et al.

Hematological findings and complications of COVID-19. Am J Hematol 2020;95:834-47.

26. Qin C, Zhou L, Hu Z, et al. Dysregulation of Immune Response in Patients With Coronavirus 2019 (COVID-19) in Wuhan, China. Clin Infect Dis 2020;71:762-8.

27. Lebourgeois S, David A, Chenane HR, et al. Differential activation of human neutrophils by SARS-CoV-2 variants of concern. Front Immunol 2022;13:1010140.

28. Peyneau M, Granger V, Wicky PH, et al. Innate immune deficiencies are associated with severity and poor prognosis in patients with COVID-19. Sci Rep 2022;12:638.

29. Xu Z, Shi L, Wang Y, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. Lancet Respir Med 2020;8:420-2.

30. Jamal M, Bangash HI, Habiba M, et al. Immune dysregulation and system pathology in COVID-19. Virulence 2021;12:918-36.

31. Moradi Khaniabadi P, Bouchareb Y, Al-Dhuhli H, et al. Two-step machine learning to diagnose and predict involvement of lungs in COVID-19 and pneumonia using CT radiomics. Comput Biol Med 2022;150:106165.

32. Zysman M, Asselineau J, Saut O, et al. Development and external validation of a prediction model for the transition from mild to moderate or severe form of COVID-19. Eur Radiol 2023;33:9262-74.

33. Li L, Wang L, Zeng F, et al. Development and multicenter validation of a CT-based radiomics signature for predicting severe COVID-19 pneumonia. Eur Radiol 2021;31:7901-12.

34. Zhang M, Zeng X, Huang C, et al. An AI-based radiomics nomogram for disease prognosis in patients with COVID-19 pneumonia using initial CT images and clinical indicators. Int J Med Inform 2021;154:104545.

35. Vancheri SG, Savietto G, Ballati F, et al. Radiographic findings in 240 patients with COVID-19 pneumonia: time-dependence after the onset of symptoms. Eur Radiol 2020;30:6161-9.

36. Wong HYF, Lam HYS, Fong AH, et al. Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19. Radiology 2020;296:E72-8.